التحليل الإحصائي

أ.د. عبدالله النجار

المحاضرة 1 (الجزء 1)

المجموعات / تعريف المجموعات وأنواعها والعمليات المرتبطة بها

تعريف المجموعة:

المجموعة ببساطة هي تجمع من الأشياء أو العناصر المحددة تماماً. وقد تكون هذه الأشياء أعداداً أو أشخاصاً أو أحداثاً أو أي شئ آخر.

ترمز للمجموعات بواسطة حروف كبيرة مثل: $A, B, C, \cdots$

الأشياء التي تتكون منها المجموعة تسمى عناصر المجموعة و ترمز للعناصر بواسطة حروف صغيرة مثل: $a, b, c, \cdots$

يستخدم الرمز ϵ "ينتمي إلى" ليبين عناصر المجموعة ، فمثلاً إذا كان العنصر a من ضمن عناصر المجموعة A فإننا نقول أن a ينتمي إلى المجموعة A ويكتب بالصورة $a \in A$

أما إذا كان a ليس عنصرا من عناصر المجموعة A فإننا نقول أن a لا ينتمي إلى المجموعة A ويكتب بالصورة $a \notin A$

ملاحظة : تعد دراسة المجموعات ذات أهمية كبيرة في دراسة العلاقات والدوال.

أمثلة على المجموعات:

مثال: A = {a, b, c, d}

أي أن المجموعة A تتكون من العناصر a و b و c و d

$b \in A$ أي أن العنصر b ينتمي إلى المجموعة A

$f \notin A$ أي أن العنصر f لا ينتمي إلى المجموعة A

طرق كتابة المجموعات:

1- طريقة العد (سرد العناصر): يتم فيها وضع جميع عناصر المجموعة، أو جزء منها ، بين قوسي المجموعة { } بحيث يفصل بين كل عنصرين بعلامة فاصلة "," مثل:

- A = {1, 3, 5, 7}
- B = {a, b, c, d}
- C = {1, 2, 3, ...}

بحيث لا يتم تكرار العناصر

2- طريقة القاعدة (الصفة المميزة): يتم فيها وصف المجموعة بذكر صفة يمكن بواسطتها تحديد عناصرها، أي الصفة التي تحدد ارتباط عناصر المجموعة ، فمثلاً:

- A={x: عدد طبيعي زوجي x}
- B={x: كلية بجامعة الملك فيصل x}
- C={x: طالب مسجل بالمقرر الحالي x}
- D={x: 0≤x ≤12 ،عدد صحيح x}

مثال على طرق كتابة المجموعات:

فمن خلال رمي حجر نرد مرتين نستطيع أن نعبر عن الحادثة (الحصول على مجموع يساوي 7) من خلال التالي:

- طريقة سرد جميع العناصر وبينهما فاصلة كالتالي: A={(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)}
- ويمكن أن نعبر عن الحادثة نفسها بطريقة الصفة المميزة و هي كتابة مميزات العناصر بين القوسين {} عوضا عن كتابة العناصر نفسها كالتالي: A={ (x,y) : x + y = 7 }

إذا المجموعة بشكل عام يمكن أن تكتب بميزة عناصرها بأشكال مختلفة طالما كانت الميزة كافيه لتحديد العناصر بشكل دقيق.

أنواع المجموعات:

1- المجموعة الخالية:

وهي مجموعة الأعداد الصحيحة التي بين العددين 0,1 مجموعة خالية، أيضا مجموعة أسماء الأسماك التي تتحدث اللغة العربية مجموعة خالية بالتأكيد. ويرمز للمجموعة الخالية بالحرف اليوناني **Ø** "فاي" أو بقوسين {}.

- A = {x: x عدد طبيعي زوجي وفردي}
- B = {x: x دولة عربية تقع في اوروبا}

2- المجموعة المنتهية:

المجموعة التي تكون عناصرها محدودة، مثال: المجموعات التالية مجموعات منتهية

- A = { 2, 4, 6, 8}
- B = {1, 2, 3, ……, 100}
- C = {x, y, z, w, u}

3- المجموعة غير المنتهية:

المجموعة التي تكون عناصرها غير محدودة، مثال: المجموعات التالية مجموعات غير منتهية

- A = {x: x عدد طبيعي فردي}
- B = {10, 20, 30,……..}

4- المجموعة الكلية:

هي المجموعة التي تدرس جميع المجموعات باعتبارها مجموعات جزئية منها، ويرمز لها بالرمز U.

5- المجموعة الجزئية:

فنقول عن مجموعة A أنها مجموعة جزئية subset من مجموعة B إذا كان كل عنصر ينتمي إلى A ينتمي إلى B ونعبر عن هذا بكتابة

فإذا كانت $A \subset B$ وكانت $A \neq B$ قلنا أن A جزئية فعلية proper subset من B أو A محتواه في B أو المجموعة B تحتوي A

أما إذا كانت A = B فإن كل عنصر ينتمي إلى أحدهما ينتمي للأخرى وبالتالي $A \subset B$ و $B \subset A$

<u>أمثلة:</u>

- إذا كانت A = {2, 4, 6} و B = {1, 2, 3, 4, 5, 6, 7, 8} فان $A \subset B$

- مجموعة جميع طلاب التعليم الالكتروني بجامعة الملك فيصل مجموعة جزئية من مجموعة طلاب هذه الجامعة.

6- تساوي المجموعات:

تكون المجموعتان B ،A متساويتان إذا كانت $A \subseteq B, B \subseteq A \Rightarrow A = B$

مثال:

- $\{-1, +1\} = \{x: x^2 = 1\}$
- {م، ل، س} ≠ {x : x حرف من كلمة سلام }

أما المجموعتان المتكافئتان فهما المجموعتان اللتان تتساويان في عدد عناصرهما وتكتب على الصورة $A \equiv B$

مثال: أي المجموعات التالية متكافئة وأيها متساوية؟

١. $A = \{1,3,5,7\}$, $B = \{3,1,5,7\}$
٢. $A = \{0,1,2\}$, $B = \{a,b,c\}$

الحل:

١. $A = B$
٢. $A \equiv B$

العمليات على المجموعات:

1- الاتحاد

اتحاد المجموعتين B ، A $(A \cup B)$ هو مجموعة كل العناصر الموجودة في A أو في B أو في كليهما.

مثال: A = {1 , 2, -6 , -7} ، B = {-6, -7, -11}

$(A \cup B)$ = {1, 2, -6, -7, -11}

2- التقاطع

تقاطع المجموعتين B ، A $(A \cap B)$ هو مجموعة كل العناصر الموجودة في A و في B معاً. أي العناصر المشتركة بين A و B .

مثال: A = {1 , 2, -6 , -7} ، B = {-6, -7, -11}

$(A \cap B)$ = {-6, -7}

3- المكملة أو المتممة:

يقال أن $\bar{A}$ مكملة المجموعة A إذا كانت تحتوي على جميع عناصر المجموعة الكلية U باستثناء عناصر A. أي أن

مثال:

S = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20}

A = {1, 3, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19}

B = {1, 2, 3, 6, 8, 11, 12, 14, 16}

$\overline{A}$ = {2, 4, 6, 20}

$\overline{B}$ = {4, 5, 7, 9, 10, 13, 15, 17, 18, 19, 20}

4- الفرق

إذا كانت مجموعتان A، B فان A-B يسمى بالفرق وهو مجموعة كل العناصر الموجودة A وليست في B.أي أن

إذا كانت A = {1, 2, 3, x, y} و B = {3, 4, 5, x, w}

فإن A – B = {1, 2, y}

مثال: إذا كانت A = {1, 2, 3, x, y} و B = {3, 4, 5, x, w} وكانت المجموعة الكلية U = {1, 2, 3, 4, 5, w, x, y, z}

فأوجد:

1) $A \cup B$
2) $A \cap B$
3) $A - B$
4) $\overline{A}$
5) $\overline{B}$

الحل:

1) $A \cup B = \{1,2,3,4,5, x, y, w\}$
2) $A \cap B = \{3, x\}$
3) $A - B = \{1,2, y\}$
4) $\overline{A} = \{4,5, w, z\}$
5) $\overline{B} = \{1,2, y, z\}$

تدريبات

1- نفترض أن $A = \{3,4,5, x, y\}$ و $B = \{4, x, y, z\}$ ضع الرمز $\in$ أو $\notin$ في المكان الفارغ لتكون الجملة صحيحة

(i) 3 ________ A

(ii) 3 ________ B

(iii) x ________ A

(iv) x ________ B

(v) z ________ A

(vi) z ________ B

(vii) 1 ________ A

$(viii)$ 1 ________ B

2- اسرد عناصر كل مجموعة من المجموعات التالية . يمكن استخدام النقط للتعبير عن استمرار سرد عناصر المجموعة عندما يكون بها عدد لانهائي من العناصر

i. A = {x: عدد طبيعي اصغر من ٧ x}
ii. B={x: عدد طبيعي زوجي يقبل القسمة على٢ x}
iii. C={y: حرف من حروف الهجاء المحصور بين c و h y}
iv. D={x: عدد طبيعي فردي اصغر من ١٧ x}

المحاضرة 1 (الجزء 2)

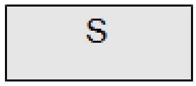
المجموعات / تمثيل المجموعات من خلال استعمال الأشكال الهندسية

أشكال ِفنْ VIN Figures

يمكن تمثيل المجموعات والعمليات المختلفة عليها من خلال استعمال اشكال هندسية تسمى أشكال ِفنْ وذلك وفق ما يلي:

1- المجموعة الكلية:

تمثل المجموعات الكلية بمستطيل ويرمز لها بالرمز S

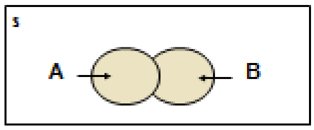


S

2- إتحاد الحوادث Events Union :

لأي حادثتين A و B فإن الحادثة التي تتضمن كافة العناصر التي تنتمي إلى A أو إلى B أو إلى كليهما معا يطلق عليها إتحاد حادثتين ويرمز لها $(A \cup B)$ أو (A أو B) والشكل التالي يوضح ذلك:

$\cup$



شكل فن لتمثيل إتحاد حادثتين A و B

$(A \cup B)$

وبشكل عام لأي n حادثة A1 , A2, A3, ... An فإن إتحاد هده الحوادث هو: $\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup A_3 ...... \cup_n$

ويمكن القول أن $\bigcup_{i=1}^{n} A_i$ هو حدث يقع إذا وقع أحد هذه الحوادث Ai على الأقل وهو ما يطلق عليه جمع الأحداث

فالإتحاد $\cup$ يعني اتحاد المجموعتين A و B وهو مجموع العناصر الموجودة في كلتا المجموعتين دون تكرار العناصر.

مثال:

A={1 , 2, -6 , -7} ، B={-6, -7, -11}

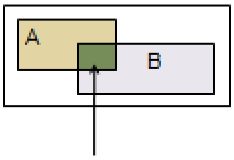
$(A \cup B)$ = {1, 2, -6, -7, -11}

خواص العمليات الجبرية لإتحاد الحوادث:

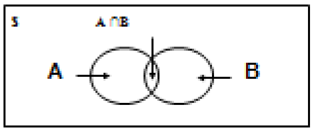
- إذا كانت A و B و C ثلاث حوادث فإن: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ ويعني ذلك توزيع الإتحاد على التقاطع.
- وكذلك هناك خاصية التبديل والتي تعني أن $(A \cup B) = (B \cup A)$

3- تقاطع الحوادث Events Intersection :

لأي حادثتين A و B فإن الحادثة التي تتضمن كافة نقاط العينة التي تنتمي إلى A و B أو إلى كليهما معا في نفس الوقت يطلق عليها تقاطع حادثتين ويرمز له $(A \cap B)$ أو (A و B) وباستخدام أشكال ِفنْ يكون الجزء المحدد بـ A and B هو الذي يمثل تقاطع الحادثتين :



$(A \cap B)$



شكل فِن لتمثيل تقاطع حادثتين A و B

وبشكل عام لأي n حادثة An .... ,A1 , A2, A3 فإن تقاطع هده الحوادث هو: $\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap A_3 ......\cap_n$

ويمكن القول أن $\bigcap_{i=1}^{n} A_i$ هو حدث يقع إذا وفقط وقعت كل الحوادث Ai على الأقل وهو ما يطلق عليه ضرب الحوادث

تقاطع الحوادث Events Intersection :

فالتقاطع $\cap$ إذا هو مجموعة العناصر المشتركة بين مجموعتين أو أكثر.

مثال:

A={1 , 2, -6 , -7} ، B={-6, -7, -11}

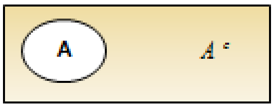
$(A \cap B)$ = {-6, -7}

خواص العمليات الجبرية لتقاطع الحوادث:

- إذا كانت A و B و C ثلاث حوادث فإن: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ ويعني ذلك توزيع التقاطع على الإتحاد.
- وكذلك هناك خاصية التبديل والتي تعني أن $(A \cap B) = (B \cap A)$

4- الحادثة المتممة Complementary Event :

لأي حادثة A فإن متممتها هي الحادثة التي تتضمن كافة العناصر التي لا تنتمي إلى A ويرمز لها بالرمز A^c أو $\bar{A}$ وهو حدث يتألف من جميع عناصر Ω غير المنتمية إلى A وباستخدام أشكال ِفنْ فإن الجزء المظلل يمثل الحادثة المتمة :



$\bar{A}$ A^c

شكل فِن لتمثيل مكملة الحادثة A

مثال:

S={1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20}

A={1, 3, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19}

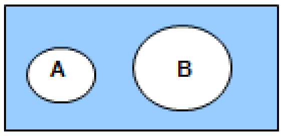
B={1, 2, 3, 6, 8, 11, 12, 14, 16}

$\overline{A}$ ={2, 4, 6, 20}

$\overline{B}$ ={4, 5, 7, 9, 10, 13, 15, 17, 18, 19, 20}

5- الحوادث المتنافية Mutually Execlusive Events :

الحادثتان A و B متنافيتان أو منفصلتان إذا كان تقاطعهما خاليا أي أن $A \cap B = \phi$ ويمكن القول أيضا أن $A \cap A^c = \phi$ ، وباستخدام أشكال ٍفنْ فإن الحادثتان المنفصلتان يتمثلان بالشكل التالي:



$A \cap B = \phi$

$A \cap A^c = \phi$

B و A شكل فن لتمثيل حادثتان متنافيتان

بعض العلاقات المهمة

$\overline{B \cup A} = \overline{B} \cap \overline{A}$

$\overline{A \cup B} = \overline{A} \cap \overline{B}$

إذا كانت $A \subset B$ فإن:

$A = A \cap B$

$B = A \cup B$

$\overline{B} \subset \overline{A}$

$A \cup A^c = S$

$A \cap A^c = \phi$

$\overline{S} = \phi$

$\overline{\phi} = S$

$A \cup S = S$

$A \cap S = A$

$A \cap \phi = \phi$

أمثلة وتمارين

يراد شراء ثلاثة أنواع من اللحوم من جزار معين، فإذا رمزنا للحم الدجاج بـ A ولحم الضأن بـ B ، ولحم العجل بـ C فإن :

- توفر أنواع اللحوم الثلاثة يعني توفر لحم A و B و C أي بمعنى: $A \cap B \cap C$
- عدم توفر أي نوع من اللحوم يعني عدم توفر A و B و C أو كلها أي بمعنى: $\overline{A} \cap \overline{B} \cap \overline{C}$
- توفر نوع واحد من اللحوم على الأقل هو توفر A أو B أو C أو كلها أي بمعنى: $A \cup B \cup C$
- توفر نوع A فقط يعني: $A \cap \overline{B} \cap \overline{C}$
- توفر نوع واحد من اللحم يعني إما توفر A وعدم توفر النوعين الآخرين أو توفر B وعدم توفر النوعين الآخرين ، أو توفر C وعدم توفر النوعين الآخرين أي بمعنى: $(A \cap \overline{B} \cap \overline{C}) \cup (\overline{A} \cap B \cap \overline{C}) \cup (\overline{A} \cap \overline{B} \cap C)$

قذفت قطعة نقود معدنية ثلاث مرات، أوجد فراغ العينة Ω وعدد عناصرها واكتب الحوادث التالية وعدد عناصر كل منها:

١. الحادثة A ظهور صورة في الرمية الأولى.
٢. الحادثة B ظهور صورة واحدة على الأقل.
٣. الحادثة C ظهور كتابة في الرمية الأولى وصورة في الرمية الثانية.
٤. الحادثة $(A \cap B)$
٥. الحادثة $(A \cup C)$
٦. الحادثة $(\overline{A} \cup \overline{B})$
٧. الحادثة $(A \cap \overline{B})$
٨. الحادثة $(\overline{A \cap B})$

الحل :

فراغ العينة Ω يمكن إيجاده من خلال حساب ظهور كل رمية مباشرة على النحو التالي:

Ω= {(HHH), (HHT), (HTH), (HTT), (THH), (THT), (TTH), (TTT)}

١. الحادثة A ظهور صورة في الرمية الأولى: {(HHH), (HHT), (HTH), (HTT)} =A
٢. الحادثة B ظهور صورة واحدة على الأقل.{(HHH), (HHT), (HTH), (HTT), (THH), (THT), (TTH)} =B
٣. الحادثة C ظهور كتابة في الرمية الأولى وصورة في الرمية الثانية.{(THH), (THT)} = C
٤. $(A \cap B)$ = {(HHH), (HHT), (HTH), (HTT)}
٥. $(A \cup C)$ = {(HHH), (HHT), (HTH), (HTT), (THH), (THT)}
٦. $(\overline{A} \cup \overline{B})$ = {(THH), (THT), (TTH), (TTT)}
٧. $(A \cap \overline{B}) = \phi$
٨. $(\overline{A \cap B})$ = {(THH), (THT), (TTH), (TTT)}

المحاضرة 2 (الجزء 1)

نظرية الاحتمالات

تعريف الاحتمالات:

يمكن تعريف الاحتمالات بطرق عديدة غير أن أبسطها "هو مقياس لامكانية وقوع حدث (Event) معين"

وتستعمل كلمة احتمالات بشكل دائم في حياتنا اليومية **مثل:**

- احتمال أن ينجح الطالب في مقرر دراسي.
- احتمال نزول المطر اليوم

وفي أحيان أخرى تستخدم كلمة احتمالات كلمة مرادفة لبعض الكلمات الأخرى مثل: ممكن، غالبا، مؤكد، مستحيل

وقد ارتبط المفهوم التقليدي للاحتمال بألعاب الحظ لمدة طويلة، وتختلف درجة إمكانية تحقق أي حادثة من شخص إلى آخر حسب خبرته والمعلومات المتوفرة لديه عن الحادثة

وقد تطور علم الاحتمالات تطورا كبيرا وسريعا وأصبح أساسا لعم الإحصاء وبحوث العمليات وغيرها

1- التجربة العشوائية Random Experiment :

هي تلك التجربة التي تكون جميع نتائجها معلومة مسبقا ولكن لا يمكن التنبؤ بحدوث أي من هده النتائج بصفة مؤكدة **مثلا:**

- رمي حجر نرد مرة واحدة يعتبر تجربة عشوائية، حيث نعلم جميع قيم نتائج التجربة وهي إما 1 أو 2 أو 3 أو 4 أو 5 أو 6 ولكن لا يمكن تحديد أي الأرقام يظهر إلى الأعلى بصورة مؤكدة قبل إجراء التجربة.
- رمي عملة معدنية مرة واحدة أو عدد من المرات يعتبر تجربة عشوائية معروف جميع نتائجها قبل أن تبدأ التجربة، ولكن لا يمكن الجزم بظهور أي منها في رمية معينة.
- المشاركة في سباق الخيل لحصان معين يعتبر تجربة عشوائية فهو إما أن يفوز أو يخسر أو يتعادل.

نستنتج من ذلك أنه في مثل هذه التجارب تكون النتائج التي نحصل عليها من تكرار للتجربة تتذبذب عشوائيا ومهما حاولنا التحكم بظروف التجربة فإن النتائج المتعاقبة ستتغير بشكل غير منتظم.

2- الحادث والفراغ العيني:

فراغ العينة: هو المجموعة الشاملة التي تحتوي على جميع النتائج الممكنة للتجربة العشوائية ويرمز له بالرمز Ω ويطلق علية الحالات الممكنة Possible Cases

افترض أننا نقوم بإجراء تجربة ما كرمي زهرة النرد مثلاً ونلاحظ كل النتائج الممكنة وهي ظهور أحد الأوجه الستة 1 أو 2 أو 3 أو 4 أو 5 أو 6 ونفترض أننا مهتمون بظهور رقم فردي أي 1 أو 3 أو 5 من التجربة . وهكذا فإن عملية رمي الزهرة تسمى **تجربة** (Experiment) وظهور رقم فردي هو محل اهتمامنا يسمى **حادثاً** (Event) ومجموعة جميع الحالات الممكنة الظهور تسمى **بالفراغ العيني** (Sample Space) ويلاحظ أن الحادث قد يكون حالة او اكثر من الفراغ العيني .

الحادثة: هي مجموعة جزئية من فراغ العينة وتمثل مجموعة النتائج التي تحقق الحدث وتسمى أيضا الحالات المواتية Favorable Cases ، فمثلا الحصول على رقم زوجي في تجربة رمي زهرة النرد مرة واحدة تكون الحادثة هي {2 ، 4 ، 6} ، ويمكن أن تحتوي الحادثة على عنصر واحد أو أكثر.

أمثلة وتمارين

أوجد فراغ العينة في كل من التجارب العشوائية التالية:

١. رمي عملة معدنية واحدة.
٢. رمي عملة معدنية مرتين.
٣. رمي حجر نرد مرتين.

الحل:

1- عند رمي عملة معدنية مرة واحدة جميع النتائج الممكن الحصول عليها إما صورة H أو كتابة T، فيكون بالتالي فراغ العينة:
Ω= {H , T}

2- عند رمي عملة معدنية مرتين تكون جميع النتائج الممكن الحصول عليها إما صورة في الرمية الأولى وصورة في الرمية الثانية أو صورة في الرمية الأولى وكتابة في الرمية الثانية أو كتابة في الرمية الأولى وصورة في الثانية أو كتابة في الأولى وكتابة في الثانية فيكون بالتالي فراغ العينة في هذه التجربة: Ω={(H,H), (H,T), (T,H), (T,T)}

3- عند رمي حجر نرد مرتين (حجر النرد هو مكعب صغير كتب أو رسم على أوجهه الستة الأرقام من 1 إلى 6) فتكون جميع النتائج الممكن الحصول عليها إما ظهور رقم 1 في الرمية الأولى ورقم 1 في الرمية الثانية أو رقم 1 في الرمية الأولى ورقم 2 في الرمية الثانية وهكذا، فيكون بالتالي فراغ العينة في هده التجربة :

Ω = {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}

كما يمكننا كتابة فراغ العينة في تجربة رمي حجر النرد مرتين على شكل جدول كالتالي:

x,y	1	2	3	4	5	6
1	1,1	1,2	1,3	1,4	1,5	1,6
2	2,1	2,2	2,3	2,4	2,5	2,6
3	3,1	3,2	3,3	3,4	3,5	3,6
4	4,1	4,2	4,3	4,4	4,5	4,6
5	5,1	5,2	5,3	5,4	5,5	5,6
6	6,1	6,2	6,3	6,4	6,5	6,6

مثال:

في تجربة رمي عملة معدنية ثلاث مرات، اكتب فضاء العينة لهذه التجربة وعبر عن الحوادث التالية:

١. الحصول على H (صورة) مرة واحدة
٢. الحصول على H (صورة) مرتين
٣. الحصول على H (صورة) ثلاث مرات
٤. عدم الحصول على H (صورة)

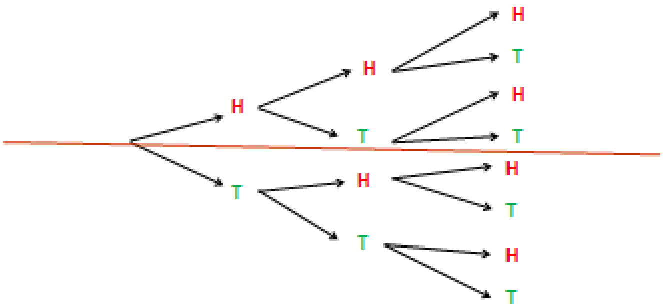
الحل:

عند رمي عملة معدنية ثلاث مرات فيكون بالتالي فراغ العينة :

Ω= {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

١. ويمكن الحصول على الحادثة H (صورة) لمرة واحدة ونرمز لها بالرمز A1 كالتالي: A1 ={HTT, THT, TTH}
٢. ويمكن الحصول على الحادثة H (صورة) لمرتين ونرمز لها بالرمز A2 كالتالي: A2 ={HHT, HTH, THH}
٣. ويمكن الحصول على الحادثة H (صورة) ثلاث مرات ونرمز لها بالرمز A3 كالتالي: A3 ={HHH}
٤. ويمكن عدم الحصول على الحادثة H (صورة) ونرمز لها بالرمز A4 كالتالي: A4 ={TTT}

ويمكن من خلال استخدام **الرسم الشجري** معرفة فراغ العينة للمثال السابق (في تجربة رمي عملة معدنية ثلاث مرات) كالتالي:



مثال:

في طريقك إلى الجامعة توجد إشارتا مرور، ما هو فضاء العينة لتجربة ذهابك إلى الجامعة؟

الحل:

نفترض أنه عندما تكون الإشارة خضراء نرمز لها بالرمز G وعندما تكون حمراء نرمز لها بالرمز R فيكون بالتالي فضاء العينة كالتالي: Ω={GG, GR, RG, RR}

مثال:

في تجربة رمي حجر نرد مرتين عبر عن الحوادث التالية بدلالة نقاط العينة والصفة المميزة؟

١. الحصول على مجموع يساوي 7
٢. الفرق بين العددين الناتجين يساوي القيمة المطلقة 1
٣. الحصول على مجموع يساوي 9 على الأقل
٤. الحصول على 1 في الرمية الأولى
٥. الحصول على حاصل ضرب يساوي 6 على الأكثر
٦. الحصول على مجموع أقل من أو يساوي 2

الحل:

يمكننا كتابة فراغ العينة في تجربة رمي حجر النرد مرتين على شكل جدول كالتالي:

x,y	1	2	3	4	5	6
1	1,1	1,2	1,3	1,4	1,5	1,6
2	2,1	2,2	2,3	2,4	2,5	2,6
3	3,1	3,2	3,3	3,4	3,5	3,6
4	4,1	4,2	4,3	4,4	4,5	4,6
5	5,1	5,2	5,3	5,4	5,5	5,6
6	6,1	6,2	6,3	6,4	6,5	6,6

وإذا رمزنا للرمية الأولى بـ x والرمية الثانية بـ y فإنه يمكننا كتابة الحوادث المطلوبة في السؤال على النحو التالي :

1- الحصول على مجموع يساوي 7

- بطريقة سرد جميع العناصر بينهما فاصلة (طريقة نقاط العينة): A={(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)}
- بطريقة الصفة المميزة: A={ (x,y) : x + y = 7 }

2- الفرق بين العددين الناتجين يساوي القيمة المطلقة 1

- بطريقة سرد جميع العناصر بينهما فاصلة (طريقة نقاط العينة):
B={(1,2), (2,1), (2,3), (3,2), (3,4), (4,3), (4,5), (5,4), (5,6), (6,5)}
- بطريقة الصفة المميزة: B={ (x,y) : x – y = | 1 |}

3- الحصول على مجموع يساوي 9 على الأقل

- بطريقة سرد جميع العناصر بينهما فاصلة (طريقة نقاط العينة):
C={(4,5), (4,6), (5,4), (5,5), (5,6), (6,3), (3,6). (6,4), (6,5), (6,6)}
- بطريقة الصفة المميزة: C={ (x,y) : x + y ≥ 9 }

4- الحصول على الرقم 1 في الرمية الأولى

- بطريقة سرد جميع العناصر بينهما فاصلة (طريقة نقاط العينة): D={(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)}
- بطريقة الصفة المميزة: D={ (x,y) : x = 1 }

5- الحصول على حاصل ضرب يساوي 6 على الأكثر

- بطريقة سرد جميع العناصر بينهما فاصلة (طريقة نقاط العينة):
E={(1,1), (1,2), (1,3), (1,4), (1,5), (2,1), (2,2), (2,3), (3,1), (3,2), (4,1), (1,6), (6,1), (5,1)}
- بطريقة الصفة المميزة: E={ (x,y) : x * y ≤ 6 }

6- الحصول على مجموع أقل من أو يساوي 2

- بطريقة سرد جميع العناصر بينهما فاصلة (طريقة نقاط العينة): F={(1,1)}
- بطريقة الصفة المميزة: F={ (x,y) : x + y ≤ 2 }

مثال:

عبر بالكلمات عن كل الحوادث الممثلة بالمجموعات الجزئية التالية من نقاط العينة:

- G={(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)}
- H={(1,1), (2,1), (1,2), (3,1), (2,2), (1,3)}
- I={(5,1), (1,5), (6,2), (2,6)}
- J={(1,4), (2,4), (3,4), (4,4), (5,4), (6,4)}
- K={(2,2), (2,4), (2,6), (4,2), (4,4), (4,6), (6,2), (6,4), (6,6)}

الحل:

الحادثة	التعبير بالكلمات عن الحوادث
الحادثة G	تعني الحصول على نفس العدد في الرمية الأولى والرمية الثانية
الحادثة H	تعني الحصول على مجموع رميتين أقل من (5)
الحادثة I	تعني الحصول على فرق بين الرميتين يساوي (4)
الحادثة J	تعني الحصول على (4) في الرمية الثانية
الحادثة K	تعني الحصول على عدد زوجي في كلا الرميتين

3- الحالات الممكنة (Possible Cases)

هي الحالات أو النتائج المختلفة التي يمكن أن تظهر نتيجة لإجراء تجربة معينة، فمثلاً عند رمي قطعة عملة تكون نتيجتها صورة أو كتابة، وعند رمي زهرة نرد تكون نتيجتها 1 أو 2 أو 3 أو 4 أو 5 أو 6 فيقال أن عدد الحالات الممكنة 2 في حالة رمي قطعة العملة و 6 في حالة رمي زهرة النرد.

4- الحالات المواتية (Favorable Cases)

هي النتائج او الحالات التي تؤدي إلى تحقيق الحادث الذي هو موضع اهتمامنا، فإذا كان الحادث هو الحصول على رقم فردي في حالة رمي زهرة النرد فإن الحالات التي تحقق هذا الحادث هي الحصول على 1 أو 3 أو 5 ، هذه الحالات الثلاثة تسمى الحالات المواتية.

5- الحالات المتماثلة (Equally Likely Cases)

إذا كان لدينا عدة كرات معدنية مصنوعة من مادة واحدة متجانسة في الكثافة ولها نفس الوزن والحجم وضعناها في كيس وسحبنا كرة منها بعد خلطها جيداً فإن هذه الكرات تكون حالات متماثلة أي يكون لكل منها نفس النصيب في السحب.

6- الحوادث المتنافية (Mutually Exclusive Events)

يقال عن الحادثين A و B أنهما متنافيان إذا استحال حدوثهما معا. فمثلاً عند رمي عملة معدنية لا يمكن الحصول على وجهين في وقت واحد.

7- الحوادث المستقلة (Independent Events)

يعتبر الحادثين A أو B حادثين مستقلين إذا كان وقوع إحداهما أو عدم وقوعه لا يؤثر في وقوع الآخر. فمثلاً عند رمي قطعة عملة واحدة مرتين متتاليتين فإن نتيجة الرمية الثانية لا تتأثر بنتيجة الأولى.

8- الحوادث الشاملة (Exhaustive Events)

تسمى الحوادث A ، B ، C ... حوادث شاملة في تجربة ما إذا كان لابد من حدوث إحداها عند إجراء التجربة.

فمثلاً عند اختيار طالب من الجامعة لمعرفة حالته ما إذا كان مدخنا أو غير مدخن تعتبر هذه الحالات حوادث شاملة لأنه لابد للفرد أن يكون له صفة واحدة من هذه الصفات. كذلك فإن الحصول على العدد 1 أو 2 أو 3 أو 4 أو 5 أو 6 عند رمي حجر النرد تعتبر حوادث شاملة لأنه لابد من حدوث إحداها.

المحاضرة 2 (الجزء 2)

نظرية الاحتمالات

ترتبط كلمة احتمال دائما بذكر حدثا ما، فاحتمال وقوع حدث معين هو نسبة وقوع هذا الحدث في الأجل الطويل، فعندما نقول إن احتمال الحصول على وحدة معيبة من إنتاج إحدى الآلات هو 0.08 ، فإننا نعني بذلك أن 8 في المئة من إنتاج هذه الآلة في الأجل الطويل سيكون معيبا، ونحن لا نستطيع ضمان وجود نسبة معينة من الوحدات المعيبة بالضبط في أي 100 وحدة من إنتاج هذه الآلة، ولكننا نتوقع أن نجد نسبة معينة في المئة من إنتاج هذه الآلة معيبا إذا فحصنا عددا كبيرا وكافيا من إنتاجها.

وللإحتمالات تعريفات عدة سنتعرض لها فيما يلي:

تعريف الاحتمالات

التعريف النسبي للاحتمالات Rational Probability Definition :

عند إجراء تجربة عدد N من المرات وكان A حدثا عشوائيا متعلقا بهده التجربة فإن التكرار النسبي يعرف على أنه حاصل قسمة عدد مرات حدوث الحادثة A مقسوما على عدد مرات حدوث التجربة أو بمعنى : $f_N(A)=\frac{N_A}{N_\Omega}$

أي أن التكرار النسبي لـ A يساوي تكرار A مقسوما على التكرار الكلي.

$0\le f_N(A)\le 1$ التكرار النسبي لـ A أكبر من أو يساوي (صفر) وأصغر من أو يساوي (1)

$f_N(A)=1$ إذا وفقط إذا وقع الحدث A في N مرة أجريت فيها التجربة.

$f_N(A)=0$ إذا وفقط إذا لم يقع الحدث A في N مرة أجريت فيها التجربة.

إذا كان A و B حادثتان متنافيتان فإن $f_N(A\cup B)=f_N(A)+f_N(B)$

إن التكرار النسبي $f_N(A)$ لحادثة A يأخذ قيمة ثابتة إدا زاد عدد محاولات التجربة عن عدد معين ، ويكون في العادة العدد كبير، وهذا ما نسميه إحتمال وقوع الحادثة A .

مثال:

إذا أخذنا التجربة العشوائية رمي قطعة نقود عدة مرات وتسجيل عدد مرات ظهور الحادثة الحصول على صورة H وكررنا التجربة لعدد من المرات سجلت النتائج في الجدول التالي:

عدد الرميات الكلي N	عدد مرات ظهور الصورة H	التكرار النسبي
30	12	12 / 30
50	20	20 / 50
80	38	0.475
100	49	0.49
300	150	0.5
500	250	0.5
1000	500	0.5
1500	750	0.5

نلاحظ أنه كلما زاد عدد الرميات N فإن التكرار النسبي يختلف اختلافا بسيطا حتى تستقر عند قيمة معينة 0.5 وهذا يوضح لنا تعريف الاحتمال وفق مفهوم التكرار النسبي.

التعريف التقليدي للاحتمالات Classical Probability Definition :

لأي حادثة A فإن احتمال حدوثها يمثل نسبة عدد حالات ظهورها إلى عدد حالات ظهور فراغ العينة الكلي أي بمعنى:

$$\frac{\text{عدد الحالات المواتية}}{\text{عدد الحالات الممكنة}}$$

أي أن

$$P(A)=\frac{N_A}{N_\Omega}$$

وللتعريف التقليدي للاحتمال عدد من المسلمات وهي :

قيمة أي احتمال أكبر من أو يساوي صفر، بمعنى أنه لأي حادثة A فإن $P(A)\geq 0$

قيمة احتمال فراغ العينة يساوي واحد صحيح $P(\Omega)=1$

إذا كانت A1 و A2 حادثين متنافيين أو منفصلين بمعنى $A_1\cap A_2=\phi$

فإن $P(A_1\cup A_2)=P(A_1)+P(A_2)$

ويمكن القول بشكل عام لأي n حادثة منفصلة:

$$P(A_1\cup A_2\cup A_3\cup.....\cup A_n)=P(A_1)+P(A_2)+P(A_3)...+P(A_n)$$

مثال:

رمي حجر نرد مرد واحدة ، أحسب التالي:

١. احتمال الحصول على رقم 5
٢. احتمال الحصول على رقم زوجي
٣. احتمال الحصول على رقم أكبر من 2
٤. احتمال الحصول على رقم أقل من 7
٥. احتمال الحصول على رقم 7

الحل:

فراغ العينة لهذه التجربة هو : Ω= {1, 2, 3, 4, 5, 6} فيكون بالتالي الحل كما يلي:

١. P(A=5)=1/6
٢. P(A=2,4, 6)=3/6
٣. P(A>2)=4/6
٤. P(A<7)=6/6=1
٥. P(A=7)=0/6=0

ويستنتج من ذلك أن أقل قيمة للاحتمال تساوي الصفر ويقال أن الحدث في هده الحالة (حدث مستحيل) بينما تساوي واحد إذا كان الحدث مؤكد

مثال:

الجدول التالي يمثل توزيع عمال أحد المصانع حسب الحالة الاجتماعية للعامل والقسم الدي يعمل به:

الحالة الاجتماعية	أعزب	متزوج	المجموع
القسم الأول	5	7	12
القسم الثاني	8	14	22
القسم الثالث	10	6	16
المجموع	23	27	50

اختر عامل من الجدول السابق بطريقة عشوائية، ثم احسب الاحتمالات التالية:

١. أن يكون أعزبا
٢. أن يكون متزوجا
٣. أن يكون من القسم الأول
٤. أن يكون من القسم الأول أو الثاني
٥. أن يكون من القسم الأول وأعزب

الحل:

1- نفرض أن الحادثة A أن يكون العامل أعزب أي A = {أن يكون العامل أعزب} فيكون الإحتمال المطلوب:

$$P(A)= \frac{\text{عددالعمال العزاب}}{\text{عدد العمال الكلي}} = 23/50= 0.46$$

2- نفرض أن الحادثة B أن يكون العامل متزوج أي أن B = {أن يكون العامل متزوج} فيكون الإحتمال المطلوب:

$$P(B)= \frac{\text{عددالعمال المتزوجين}}{\text{عدد العمال الكلي}} = 27/50= 0.54$$

3- نفرض أن الحادثة C أن يكون العامل من القسم الأول أي أن C = {أن يكون العامل من القسم الأول} فيكون الإحتمال المطلوب:

$$P(C)= \frac{\text{عدد عمال القسم الأول}}{\text{عدد العمال الكلي}} = 12/50= 0.24$$

4- نفرض أن الحادثة D أن يكون العامل من القسم الأول أو الثاني أي أن D = {أن يكون العامل من القسم الأول أو الثاني} فيكون الإحتمال المطلوب:

$$P(D)= \frac{\text{عدد عمال القسم الأول أو الثاني}}{\text{عدد العمال الكلي}} = (12+22)/50=34/50= 0.68$$

5- نفرض أن الحادثة E أن يكون العامل من القسم الأول و أعزب أي أن E = {أن يكون العامل من القسم الأول وأعزب} فيكون الإحتمال المطلوب:

$$P(D)= \frac{\text{عدد عمال القسم الأول وعزاب}}{\text{عدد العمال الكلي}} = 5/50= 0.1$$

جمع الاحتمالات

أ- في حالة كون الحوادث متنافية

إذا كانت الحوادث A1 , A2 , A3 .... حوادث متنافية بمعنى أن حدوث أحدها يـؤدي إلى استحالة حدوث أي من الحوادث الأخرى وبالتالي فإن احتمال حدوث هذه الحوادث معاً يكون معدوماً فإن احتمال وقوع أي حادث من الحوادث المتنافية يساوي مجموع احتمالات وقوع هذه الحوادث

فإذا كان A ، B حادثين متنافيين كما في الشكل (1)

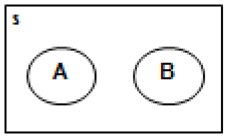
فإن P (A أو B) = P(A) + P(B)

يرمز أيضاً لاحتمال وقوع أحد الحادثين بالرمز

حوادث متنافية

P (AUB) = P (A) + P (B)

الشكل (١)



حوادث متنافية

مثال:

رمي حجر نرد مرة واحدة ، احسب:

١. احتمال الحصول على رقم 5 أو 6
٢. احتمال الحصول على رقم زوجي

الحل:

1- حيث أن الحصول على رقم 5 أو 6 حادثتان متنافيتان ، أي أن:

A1={الحصول على الرقم 5} ، و A2= {الحصول على الرقم 6} فإن:

P(A1 ᴜ A2)=(1/6)+(1/6)=1/3

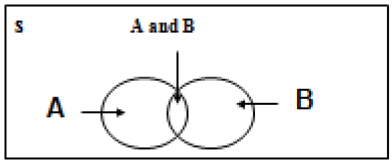
2- وحيث أن الحصول على رقم زوجي يعني الحصول على رقم 2 أو رقم 4 أو رقم 6 وكلها حوادث متنافي، أي أن:

A1={الحصول على الرقم 2} ، و A2= {الحصول على الرقم 4} ، و A3= {الحصول على الرقم 6} فإن:

P(A1 ᴜ A2 ᴜ A3)=(1/6)+(1/6) +(1/6) = 1/2

ب- في حالة كون الحوادث غير متنافية

عند عدم اشتراط تنافي الحادثين A و B يكون المقصود بالحادث (A أو B) وقوع A على انفراد أو وقوع B على انفراد أو وقوع الحادثين A و B معا في وقت واحد كما يتضح من الشكل التالي



حوادث غير متنافية

الآن P (A) + P (B) تمثل مجموع الحالات المواتية للحادث A مضافاً إليها مجموع الحالات المواتية للحادث B ولكن يجب ملاحظة أن كل من الحالات المواتية للحادث A وتلك المواتية للحادث B تتضمن الحالات المواتية لوقوع A و B معاً ، وبهذا فإنه في حالة جمع P (A) و P (B) فإننا نجمع P (A و B) مرتين ، لهذا لابد من طرح P (A و B) مرة واحدة لنحصل على الاحتمال P (A أو B) وهذا هو :

P (A او B)= P (A) + P (B) – P (A و B)

أو P (AUB) = P (A) + P(B) – P (A ∩ B)

مثال:

الجدول التالي يمثل توزيع عمال أحد المصانع حسب الحالة الاجتماعية للعامل والقسم الدي يعمل به:

الحالة الاجتماعية	أعزب	متزوج	المجموع
القسم الأول	5	7	12
القسم الثاني	8	14	22
القسم الثالث	10	6	16
المجموع	23	27	50

اختير عامل من الجدول السابق بطريقة عشوائية، احسب الاحتمالات التالية:

١. احتمال أن يكون العامل من القسم الأول أو الثاني
٢. احتمال أن يكون العامل متزوجا أو من القسم الأول
٣. احتمال أن يكون العامل من القسم الثالث أو أعزب

الحل:

1- نفرض أن الحادثة A1 أن يكون العامل من القسم الأول أي أن A1 = {أن يكون العامل من القسم الأول}

نفرض أن الحادثة A2 أن يكون العامل من القسم الثاني أي أن A2 = {أن يكون العامل من القسم الثاني}

فيكون الإحتمال المطلوب:

$P(A1)=12/50$

$P(A2)=22/50$

$P(A1 \cup A2) = P(A1)+P(A2)= (12/50)+(22/50) = 34/50= 0.68$

2- نفرض أن الحادثة A1 أن يكون العامل متزوجا أي أن A1 = {أن يكون العامل متزوج}

نفرض أن الحادثة A2 أن يكون العامل من القسم الأول أي أن A2 = {أن يكون العامل من القسم الأول}

فيكون الإحتمال المطلوب:

P(A1)=27/50

P(A2)=12/50

$P(A1 \cup A2) = P(A1)+P(A2)-P(A1 \cap A2) = (27/50)+(12/50)-(7/50) = 32/50= 0.64$

3- نفرض أن الحادثة A1 أن يكون العامل من القسم الثالث أي أن A1 = {أن يكون العامل من القسم الثالث}

نفرض أن الحادثة A2 أن يكون العامل أعزب أي أن A2 = {أن يكون العامل أعزب}

فيكون الإحتمال المطلوب:

P(A1)=16/50

P(A2)=23/50

$P(A1 \cup A2) = P(A1)+P(A2)-P(A1 \cap A2) = (16/50)+(23/50)-(10/50) = 29/50= 0.58$

الاحتمال الشرطي Conditional Probability

إذا كان لدينا الحادثين A1 , A2 وكان P: (A2) لا يساوي الصفر فأن الاحتمال الشرطي للحادث A1 بشرط وقوع الحادث A2 يعطي بالمعادلة التالية: $P(A1 | A2) = \frac{P(A_1 \cap A_2)}{P(A_2)}$

أي أن الاحتمال الشرطي للحادث A1 بشرط وقوع الحادث A2 يساوي حاصل قسمة الاحتمال المركب لـ A1 , A2 على احتمال الحادث A2

مثال:

إذا كان احتمال نجاح طالب في مقرر الرياضيات 0.64 واحتمال نجاحه في مقرر الإحصاء ومقرر الرياضيات معا 0.32 فما هو احتمال نجاحه في مقرر الإحصاء؟ علما بأنه نجح في مقرر الرياضيات.

الحل:

نفرض أن A1={نجاح الطالب في مقرر الإحصاء}

A2={نجاح الطالب في مقرر الرياضيات}

وبذلك يكون: P(A2)=0.64

$P(A1 \cap A2)=0.32$

ويكون المطلوب في هذه المسألة هو حساب $P(A1 | A2)$ وبتطبيق العلاقة :

$$P(A1 | A2) = \frac{P(A_1 \cap A_2)}{P(A_2)} = \frac{0.32}{0.64} = 0.5$$

إذا احتمال نجاح الطالب في مقرر الإحصاء علما بأنه نجح في مقرر الرياضيات هو 0.5

مثال:

الجدول التالي يمثل توزيع عمال أحد المصانع حسب الحالة الاجتماعية للعامل والقسم الدي يعمل به:

اختير عامل من الجدول السابق بطريقة عشوائية، احسب الاحتمالات التالية:

- احسب احتمال أن يكون العامل من القسم الأول بشرط أنه متزوج؟
- احتمال أن يكون العامل أعزب بشرط أنه من القسم الثالث؟

الحالة الاجتماعية	أعزب	متزوج	المجموع
القسم الأول	5	7	12
القسم الثاني	8	14	22
القسم الثالث	10	6	16
المجموع	23	27	50

الحل:

نفرض أن A1={أن يكون العامل من القسم الأول}

A2={أن يكون العامل متزوج}

B3={أن يكون العامل من القسم الثالث}

B4={أن يكون العامل أعزب}

فيكون بالتالي:

1- احتمال أن يكون العامل من القسم الأول بشرط أنه متزوج هو:

$$\frac{\text{احتمال أن يكون من القسم الأول ومتزوج}}{\text{احتمال أن يكون متزوج}}$$

$$P(A1 | A2) = \frac{P(A_1 \cap A_2)}{P(A_2)} = \frac{\frac{7}{50}}{\frac{27}{50}} = \frac{7}{27}$$

إذا احتمال أن يكون العامل من القسم الأول بشرط أنه متزوج هو 0.259

2- احتمال أن يكون العامل أعزب بشرط أنه من القسم الثالث هو:

$$\frac{\text{احتمال أن يكون من القسم الثالث وأعزب}}{\text{احتمال أن يكون من القسم الثالث}}$$

$$P(B1 | B2) = \frac{P(B_1 \cap B_2)}{P(B_2)} = \frac{\frac{10}{50}}{\frac{16}{50}} = \frac{10}{16}$$

إذا احتمال أن يكون العامل من القسم الأول بشرط أنه متزوج هو 0.625

ضرب الاحتمالات

إن احتمال حدوث حادثين مستقلين أو أكثر معاً يساوي حاصل ضرب احتمال حدوث كل واحد من هذه الحوادث ببعضها بعضاً

فمثلا إذا كان لدينا صندوق به 10 كرات متماثلة منها 6 بيضاء و 4 سوداء وسحبنا كرة من الصندوق فإن احتمال أن تكون بيضاء 6/10 واحتمال أن تكون سوداء 4/10، فإذا أعدنا الكرة إلى الصندوق (ليصبح العدد مكتملا كما كان) وسحبنا الكرة مرة أخرى فإن احتمال أن تكون بيضاء 6/10 واحتمال أن تكون سوداء 4/10 ، فتكرار العملية يؤدي إلى نفس الاحتمال.

ومن هذا نرى أن نتيجة السحب الأول لا تؤثر على نتيجة السحب الثاني وهدا ما يسمى سحب بإرجاع أو إحلال أو إعادة.

فإذا كان لدينا الحادثين المستقلين A1 و A2 فإن احتمال حدوثهما معا هو:

$$P(A1 \cap A2) = P(A1)\,P(A2)$$

بمعنى أن احتمال وقوع حدثين مستقلين معا يساوي حاصل ضرب احتمال وقوع أي منهما بمفرده في احتمال وقوع الحدث الآخر بمفرده وفي حالة التعميم لـ n فإن :

$$P(A1 \cap A2 \cap A3 \cap \ldots\ldots An) = P(A1)\,P(A2)\,P(A3)\ldots\ldots P(An)$$

مثال:

إذا رمينا قطعة نقود مرة واحدة ، احسب الاحتمالات التالية:

- أن تكون الأولى صورة والثانية كتابة.
- أن تكون كلتاهما صورة.

الحل:

نفرض أن: A1={ظهور صورة في الرمية الأولى}

A2={ظهور كتابة في الرمية الثانية}

فيكون: $P(A_1) = \dfrac{1}{2} \quad , P(A_2) = \dfrac{1}{2}$

حيث أن الحادثتان A1 و A2 مستقلتين فإن احتمال أن تكون الرمية الأولى صورة والثانية كتابة هو :

$$P(A_1 \cap A_2) = P(A_1) \times (P(A_2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

نفرض أن: B1={ظهور صورة في الرمية الأولى}

B2={ظهور صورة في الرمية الثانية}

فيكون: $P(B_1) = \dfrac{1}{2} \quad , P(B_2) = \dfrac{1}{2}$

وحيث أن الحادثتان B1 و B2 مستقلتين فإن احتمال أن تكون الرمية الأولى صورة والثانية صورة هو :

$$P(B_1 \cap B_2) = P(B_1) \times (P(B_2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

مما سبق يمكن القول أن الحوادث المعطاة تكون مستقلة عندما تبقى الاحتمالات ثابتة مثل لحوادث:

١. رمي قطع نقود (أو قطعة واحدة عدة مرات)
٢. رمي أحجار نرد (أو حجر نرد عدة مرات)
٣. السحب مع الإرجاع (أو الإعادة)

مثال:

الجدول التالي يمثل توزيع عمال أحد المصانع حسب الحالة الاجتماعية للعامل والقسم الدي يعمل به فإذا سحب عاملان من المصنع مع الإرجاع (أي إرجاع العامل الأول قبل سحب العامل الثاني) احسب :

الحالة الاجتماعية	أعزب	متزوج	المجموع
القسم الأول	5	7	12
القسم الثاني	8	14	22
القسم الثالث	10	6	16
المجموع	23	27	50

اختير عاملان من الجدول السابق بطريقة عشوائية، احسب الاحتمالات التالية:

١. احتمال أن يكون العاملان من القسم الأول؟
٢. احتمال أن يكون العاملان متزوجان؟
٣. احتمال أن يكون للعاملان نفس الحالة الاجتماعية؟
٤. احتمال أن يكون العاملان من القسم نفسه؟

الحل:

1- احتمال أن يكون العاملان من القسم الأول يعني أن يكون العامل الأول من القسم الأول (الحادثة A1) والعامل الثاني من القسم الأول (الحادثة A2) وحيث أنهما مستقتان (لأن السحب مع الإرجاع) فإن:

$$P(A_1 \cap A_2) = P(A_1) \times (P(A_2) = \frac{12}{50} \times \frac{12}{50} = \frac{144}{2500} = 0.0576$$

2- احتمال أن يكون العاملان متزوجان ، يعني أن يكون العامل الأول متزوج (الحادثة B1) والعامل الثاني متزوج (الحادثة B2) وحيث أنهما مستقتان (لأن السحب مع الإرجاع) فإن:

$$P(B_1 \cap B_2) = P(B_1) \times (P(B_2) = \frac{27}{50} \times \frac{27}{50} = \frac{729}{2500} = 0.2916$$

3- احتمال أن يكون للعاملان نفس الحالة الاجتماعية يعني أن يكون العاملان كلاهما متزوجين (الحادثة A) أو أن يكون كلاهما أعزبين (الحادثة B) فإن:

$$P(A \cup B) = P(A) + (P(B)$$
$$= P(A_1A_2) + P(B_1B_2)$$
$$= P(A_1 \times A_2) + P(B_1 \times B_2)$$
$$= \left[\frac{27}{50} \times \frac{27}{50}\right] + \left[\frac{23}{50} \times \frac{23}{50}\right] = 0.5032$$

4- احتمال أن يكون العاملان من القسم نفسه يعني أن يكون العاملان كلاهما من القسم الأول (الحادثة A) أو أن يكون كلاهما من القسم الثاني (الحادثة B) أو أن يكون كلاهما من القسم الثالث (الحادثة C) فإن:

$$P(A \cup B \cup C) = P(A) + (P(B) + (P(C)$$
$$= P(A_1A_2) + P(B_1B_2) + P(C_1C_2)$$
$$= P(A_1 \times A_2) + P(B_1 \times B_2) + P(C_1 \times C_2)$$
$$= \left[\frac{12}{50} \times \frac{12}{50}\right] + \left[\frac{22}{50} \times \frac{22}{50}\right] + \left[\frac{16}{50} \times \frac{16}{50}\right] = 0.3536$$

التباديل:

تشير التبديلة إلى الطريقة التي ترتب بها المفردات، فإذا كان لدينا مجموعة مكونة من n من المفردات، فإن عدد التبديلات الممكنة يعتمد على عدد المفردات (r) المأخوذة لهذه المجموعة لترتيبها، وتستخدم التباديل عندما تكون n أكبر من r (n>r) أو n تساوي r (n=r) ، وبالطبع فإن من المستحيل في التباديل أن تكون n أصغر من r (n<r) حيث لا يمكن أخذ r من المفردات من عدد أقل منها.

قاعدة: عندما تكون n>r فإن عدد التبديلات الممكنة لعدد r من المفردات مأخوذة من n من المفردات هو:

n!=n(n-1)(n-2)(n-3)

مثال :

6!=6(5)(4)(3)(2)(1)=720

وممكن كتابة ذلك بالصيغة التالية: 6!=6(5)(4)(3!)=720

حيث نجد أن: 6(5)(5)=120 و 3!=3(2)(1)=6

120(6)=720

قاعدة: لاحظ أن $(n-K)! \neq n! - K!$

فمثلا

$$(7-4)! \neq 7! - 4!$$

لأن $(7-4)! = 3! = 3(2)(1) = 6$

بينما $7! = 7(6)(5)(4)(3)(2)(1) = 5040$

$$4! = 4(3)(2)(1) = 24$$

لذا $7!-4!=5040-24=5016$

وهذا بالطبع لا يساوي $3!=6$

قاعدة: $P_r^n=\frac{n!}{(n-r)!}$

مثال:

رشح عشرة أشخاص لشغل الوظائف الأربع التالية: رئيس مجلس إدارة، نائب رئيس مجلس إدارة، مدير عام، ومدير إدارة موارد بشرية، بين عدد الطرق الممكنة لشغل هذه الوظائف الأربع من بين هؤلاء المرشحين.

الحل:

لاحظ أن n=10 و r=4 ، وبالتالي يوجد عشر طرق لشغل الوظيفة الأولى، وبمجرد شغل الوظيفة يتبقى تسعة مرشحين، أي يمكن شغل الوظيفة الثانية بتسع طرق، وبالمثل يوجد ثمان طرق لشغل الوظيفة الثالثة، وسبع طرق لشغل الوظيفة الرابعة، وبالتالي فإن عدد الطرق الممكنة أي عدد التبديلات الممكنة لشغل هذه الوظائف هو:

$$P_4^{10}=(10)(9)(8)(7)=5040$$

ويمكن حساب ذلك من خلال المعادلة السابقة:

$$P_4^{10}=\frac{10!}{(10-4)!}$$

$$P_4^{10}=\frac{10!}{(6)!}=\frac{10(9)(8)(7)(6!)}{(6!)}=10(9)(8)(7)=5040$$

قاعدة: إذا كانت n=r فإن عدد التراتيب الممكنة لعدد n من المفردات مأخوذة مرة واحدة هو: $P_n^n=n!$

مثال: حدد عدد الطرق الممكنة لمنح خمس جوائز لأحسن خمسة من الطلبة المتفوقين.

الحل:

يظهر لنا أنه يمكن منح الجائزة الأولى لأي طالب من هؤلاء الطلبة، أي توجد خمس طرق لمنح الجائزة الأولى، وبعد منحها فإنه يمكن منح الجائزة الثانية لأي من الطلبة الأربعة المتبقين، وبعد منح هذه الجائزة يمكن منح الجائزة الثالثة لأي طالب من الطلبة الثلاثة المتبقين وهكذا.

وبالتالي يمكن منح الجوائز الخمس بطرق عددها: 5!=(5)(4)(3)(2)(1)=120

مثال: ترتيب الحروف: **a, b, c**

abc, acb, bac, bca, cab, cba

$$P_n^n=n!=(3)(2)(1)=6$$

التوافيق:

تهتم التوافيق بعدد الطرق الممكنة لاختيار r من المفردات من n من المفردات دون أخذ الترتيب في الإعتبار، ويسمى ذلك بعدد التوفيقات الممكنة لأخذ r من المفردات من n من المفردات.

قاعدة: إذا رمزنا لعدد من التوفيقات الممكنة لعدد r من المفردات مأخوذة من n من المفردات بالرمز C_r^n ، وإذا لاحظنا أن كل توفيقة لعدد r من المفردات يناظرها !r من التبديلات، فإنه يمكن وضع المعادلة التالية للحصول على عدد التوفيقات الممكنة لاختيار r من المفردات من n من المفردات كما يلي: $C_r^n = \frac{n!}{r!}$

مثال:

حدد عدد الطرق الممكنة لاختيار ثلاث أشخاص من بين عشرة أشخاص تم ترشيحهم لعضوية مجلس إدارة إحدى الشركات.

الحل:

هنا نجد أن r=3 و n=10 وبالتالي: أي أنه توجد 120 طريقة لاختيار ثلاثة أفراد من بين عشرة مرشحين لعضوية مجلس الإدارة، لاحظ أن ترتيب اختيار المرشحين الثلاثة ليست له أهمية تذكر.

$$C_r^n = \frac{n!}{r!} = C_3^{10} = \frac{(10)(9)(8)}{3!} = \frac{720}{(3)(2)(1)} = \frac{720}{6} = 120$$

الفرق بين التباديل والتوافيق:

تختلف التوافيق عن التباديل في أن التباديل تأخذ ترتيب المفردات في الإعتبار بينما لا تهتم التوافيق بترتيبها.

وحتى نتمكن من معرفة متى يمكن استخدام التباديل لعد الأشياء ومتى يمكن استخدام التوافيق نطرح الأمثلة التالية:

مثال:

حدد عدد الطرق الممكنة لاختيار ثلاث طلبة من بين عشرين طالبا في أحد الفصول الدراسية.

الحل:

وحيث أن ترتيب اختيار الطلبة ليست له أهمية، فإننا سنستخدم التوافيق لتحديد عدد التوفيقات الممكنة (عدد الطرق الممكنة لاختيار الطلبة وذلك كما يلي: $C_r^n = \frac{n!}{r!} = C_3^{20} = \frac{(20)(19)(18)}{(3)(2)(1)} = \frac{6480}{6} = 1140$

مثال:

حدد عدد الطرق الممكنة لاختيار ثلاث طلبة من بين عشرين طالبا لمنحهم ثلاث جوائز مختلفة.

الحل:

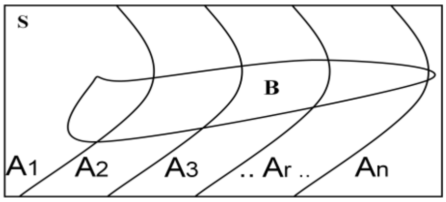
وحيث أن الجوائز مختلفة لذا فإن ترتيب الاختيار ضروري، وبالتالي تستخدم التباديل لتحديد عدد الطرق الممكنة، لاحظ وجود 20 طريقة لمنح الجائزة الأولى، و19 طريقة لمنح الجائزة الثانية، و 18 طريقة لمنح الجائزة الثالثة، أي أن الطرق الكلية (عدد التبديلات الكلية) لمنح الجوائز الثلاث لثلاث طلاب من عشرين طالبا هو: $P_r^n = n! = P_3^{20} = 20(19)(18) = 6840$

نظرية بايز (Bayes' Theorem)

إذا كانت A1, A2, ... An مجموعة أحداث متنافية وكانت احتمالات حدوثها

P(A1), P(A2), ... P(An) وإذا كان هناك حدث B يحدث إذا حدث أي من الأحداث المتنافية أنظر الشكل بالأسفل ، فإن احتمال حدوث الحدث Ar بشرط حدوث B هو:

$$P(A_r|B)=\frac{P(A_r)P(B|A_r)}{\sum_{i=1}^{n}P(A_i)P(B|A_i)} \qquad 1\le r\le n$$



مثال:

مصنع يقوم بإنتاج سلعة معينة به ثلاث آلات، تنتج الآلة الأولى 20% من إجمالي إنتاج السلعة وتنتج الآلة الثانية نسبة 35% والثالثة بنسبة 45% ، فإذا كانت نسبة الإنتاج المعيب في الثلاث آلات على الترتيب هو 2% و 2.5% و 3% ، سحبت وحدة عشوائيا من إنتاج المصنع فوجد أنها معيبة، احسب الاحتمالات التالية:

١. أن تكون القطعة المعيبة من إنتاج الآلة الأولى؟
٢. أن تكون القطعة المعيبة من إنتاج الآلة الثانية؟

الحل:

نفرض أن

A1={إنتاج الآلة الأولى} P(A1)=0.2

A2={إنتاج الآلة الثانية} P(A2)=0.35

A3={إنتاج الآلة الثالثة} P(A3)=0.45

B= {إنتاج سلعة معينة}

فيكون بالتالي:

$P(B|A1)=0.02$

$P(B|A2)=0.025$

$P(B|A3)=0.03$

إذا أن تكون السلعة من إنتاج الآلة الأولى إذا علم - بشرط - أنها معيبة هو:

$$P(A_1|B)=\frac{P(A_1)P(B|A_1)}{\sum_{i=1}^{n}P(A_i)P(B|A_i)}=\frac{0.2\times0.02}{(0.2\times0.02)+(0.35\times0.025)+(0.45\times0.03)}=0.152$$

واحتمال أن تكون السلعة من إنتاج الآلة الثانية إذا علم - بشرط - أنها معيبة هو:

$$P(A_2|B)=\frac{P(A_2)P(B|A_2)}{\sum_{i=1}^{n}P(A_i)P(B|A_i)}=\frac{0.35\times0.025}{(0.2\times0.02)+(0.35\times0.025)+(0.45\times0.03)}=0.333$$

مثال:

مستشفى به أربعة أقسام، نسب عمال النظافة في هده الأقسام هي 30% ، 40% ، 20% ، 10% على التوالي، إذا كانت نسب العمال المدخنين بهده الأقسام هي 15% ، 18% ، 12% ، 9% على التوالي، أختير عامل عشوائيا فوجد أنه مدخن ، احسب الاحتمالات التالية:

١. أن يكون العامل من القسم الأول؟
٢. أن يكون العامل من القسم الثاني؟
٣. أن لا يكون العامل من القسم الأول؟

الحل:

نفرض أن

A1={أن يكون العامل من القسم الأول} P(B│A1)=0.15 P(A1)=0.3

A2={أن يكون العامل من القسم الثاني} P(B│A2)=0.18 P(A2)=0.4

A3={أن يكون العامل من القسم الثالث} P(B│A3)=0.12 P(A3)=0.2

A4={أن يكون العامل من القسم الرابع} P(B│A4)=0.09 P(A4)=0.1

1- إذا احتمال أن يكون العامل من القسم الأول إذا علم – بشرط – أنه مدخن :

$$P(A_1|B)=\frac{P(A_1)P(B|A_1)}{\sum_{i=1}^{4}P(A_i)P(B|A_i)}=\frac{0.3\times0.15}{(0.3\times0.15)+(0.4\times0.18)+(0.2\times0.12)+(0.1\times0.09)}=0.3$$

2- واحتمال أن يكون العامل من القسم الثاني إذا علم – بشرط – أنه مدخن :

$$P(A_2|B)=\frac{P(A_2)P(B|A_2)}{\sum_{i=1}^{4}P(A_i)P(B|A_i)}=\frac{0.4\times0.18}{(0.3\times0.15)+(0.4\times0.18)+(0.2\times0.12)+(0.1\times0.09)}=0.48$$

3- واحتمال أن لا يكون العامل من القسم الأول إذا علم – بشرط – أنه مدخن :

$$P(A_1^c|B)=1-0.3=0.7$$

المحاضرة 3

المتغيرات العشوائية

المتغير العشوائي Random Variable:

المتغير العشوائي هو الذي يأخذ قيما حقيقية مختلفة تعبر عن نتائج فراغ العينة، ومن ثم مجال هذا المتغير، يشمل كل القيم الممكنة له، ويكون لكل قيمة من القيم التي يأخذها المتغير احتمال معين

وينقسم المتغير العشوائي إلى قسمين هما:

١. **المتغيرات العشوائية المنفصلة** Discrete Random Variables
٢. **المتغيرات العشوائية المتصلة (المستمرة)** Continuous Random Variables

1- المتغيرات العشوائية المتقطعة (المنفصلة):

المتغير العشوائي المنفصل هو الذي يأخذ قيم بينية، ومتباعدة، ويرمز للمتغير العشوائي بشكل عام بحرف من الحروف الأبجدية الكبيرة *X, Y, Z,...* ويرمز للقيم التي يأخذها المتغير بالحروف الأبجدية الصغيرة، *x, y, z, ...*

فالمتغير العشوائي المنفصل هو كل قيمة من قيم المتغير العشوائي كنتيجة لعد الأشياء، **ومن أمثلة هذه المتغيرات:**

- عدد الأولاد الذكور في الأسرة المكونة من أربع أولاد *X*، *X:{x=0,1,2,3,4}*.
- عدد العملاء الذين يتم إنهاء خدمتهم البنكية كل *10* دقائق *Y*، *Y:{y=0,1,2,3,....}*.
- عدد مرات استخدام نوع معين من الأسمدة خلال الدورة الزراعية.
- عدد الوحدات التالفة من إنتاج مزرعة معينة تنتج *200* وحدة كل موسم.
- عدد الوحدات التي تستهلكها الأسرة من سلعة معينة خلال الشهر.
- وهكذا..... الأمثلة كثيرة

التوزيع الاحتمالي للمتغير العشوائي المنفصل:

التوزيع الاحتمالي، هو الذي يبين احتمالات حدوث القيم التي يمكن يأخذها المتغير، والتي ترتبط باحتمالات النتائج الممكنة في فراغ العينة، وبمعنى آخر هو التكراري النسبي للقيم التي يمكن أن يأخذها المتغير.

فإذا كان المتغير العشوائي المنفصل *X* يأخذ القيم $X:\{x=x_1,x_2,...,x_n\}$، وكان $P(X=x_i)=f(x_i)$ هو احتمال أن المتغير العشوائي يأخذ القيمة x_i، فإنه، يمكن تكوين جدول التوزيع الاحتمالي للمتغير العشوائي *X*، وهو جدول مكون من عمودين، الأول به القيم الممكنة للمتغير $X:\{x=x_1,x_2,...,x_n\}$، والثاني به القيم الاحتمالية لهذا المتغير $P(X=x_i)=f(x_i)$، أي أن:

جدول التوزيع الاحتمالي للمتغير العشوائي المنفصل

x_i	$f(x_i)$
x_1	$f(x_1)$
x_2	$f(x_2)$
$\vdots$	$\vdots$
x_n	$f(x_n)$
$\sum$	1

وتسمى الدالة $f(x_i)$ بدالة الاحتمال

مثال:

إذا كانت نسبة مبيعات أحد المراكز التجارية من التفاح الأمريكي 0.60، بينما يكون نسبة مبيعاته من الأنواع الأخرى للتفاح 0.40، اشترى أحد العملاء عبوتين،

المطلوب: كون فراغ العينة.

إذا عرف المتغير العشوائي بأنه عدد العبوات المشتراة من التفاح الأمريكي، فأوجد الآتي:

- التوزيع الاحتمالي للمتغير العشوائي
- ارسم دالة الاحتمال لهذا المتغير

الحل:

تكوين فراغ العينة:

التجربة هنا هو شراء وحدتين من عبوات التفاح، ومن ثم فراغ العينة يتكون من أربع نتائج، هي:

التوزيع الاحتمالي لعدد العبوات المشتراة من التفاح الأمريكي

من المعلوم أن العميل اشترى عبوتين، وأن المتغير العشوائي هو عدد العبوات المشتراة من التفاح الأمريكي، لذا تكون القيم الممكنة للمتغير العشوائي هي:

x=0 إذا كانت العبوتين من النوع الآخر، أى إذا كانت نتيجة التجربة (آخر، آخر)

x=1 إذا كان أحد العبوتين من النوع الأمريكي، أي إذا كانت نتيجة التجربة (آخر، أمريكي) أو (أمريكي،آخر)

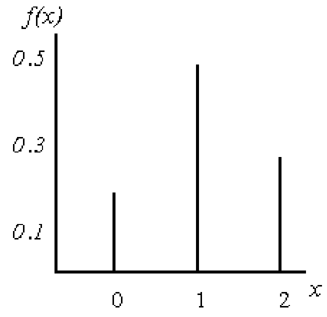
x=2 إذا كان العبوتين من النوع الأمريكي، أي إذا كانت نتيجة التجربة (أمريكي ، أمريكي)

ومن ثم يأخذ المتغير القيم: *X:{x=0,1,2}* ، ويرتبط احتمالات هذه القيم باحتمالات نتائج التجربة المناظرة لها كما هو مبين أعلاه، ومن ثم يكون التوزيع الاحتمالي للمتغير العشوائي *X* هو:

<u>جدول التوزيع الاحتمالي لعدد العبوات المشتراة من التفاح الأمريكي:</u>

x_i	$f(x_i)$
0	0.4*0.4 = 0.16
1	(0.4*0.6) + (0.4*0.6) = 0.48
2	0.6*0.6 = 0.36
$\sum$	1

<u>رسم دالة الاحتمال *f(x)*:</u>



الوسط الحسابي والتباين للمتغير العشوائي المنفصل:

يرمز **للوسط الحسابي** للمتغير العشوائي بالرمز (ميو)، ويحسب بتطبيق المعادلة التالية: $\mu = \sum x_i f(x_i)$

وأما **التباين** ويرمز له بالرمز σ^2 (سيجما)، فيحسب بتطبيق المعادلة التالية:

$$\sigma^2 = \sum (x_i - \mu)^2 f(x_i) \\ = \sum x_i^2 f(x_i) - \mu^2$$

مثال:

في المثال السابق احسب ما يلي:

١. الوسط الحسابي لعدد العبوات المشتراة من النوع الأمريكي:
٢. احسب الانحراف المعياري لعدد العبوات المشتراة من النوع الأمريكي.
٣. أوجد معامل الاختلاف النسبي

الحل:

1- الوسط الحسابي لعدد العبوات من النوع الأمريكي:

لحساب الوسط الحسابي والانحراف المعياري يتم استخدام المعادلة الخاصة بذلك وهذا يتطلب تكوين جدول يشمل المجاميع التالية: $\sum x_i \; f(x_i)$, $\sum x_i^2 \; f(x_i)$ ، وذلك كما يلي:

x_i	$f(x_i)$	$x_i f(x_i)$	$x_i^2 f(x_i)$
0	0.16	0	0
1	0.48	0.48	0.48
2	0.36	0.72	1.44
Σ	1	1.20	1.92

إذا الوسط الحسابي هو: $\mu = \sum x_i \; f(x_i) = 1.20$

2- ولحساب الانحراف المعياري يجب أولا حساب التباين وهو: $\sigma^2 = \sum x_i^2 \; f(x_i) - \mu^2 = 1.92 - (1.20)^2 = 0.48$

إذا الانحراف المعياري قيمته هي: $\sigma = \sqrt{\sigma^2} = \sqrt{0.48} = 0.693$

3- معامل الاختلاف النسبي هو: $C.V = \dfrac{\sigma}{\mu} \times 100 = \dfrac{0.693}{1.2} \times 100 = 57.7$

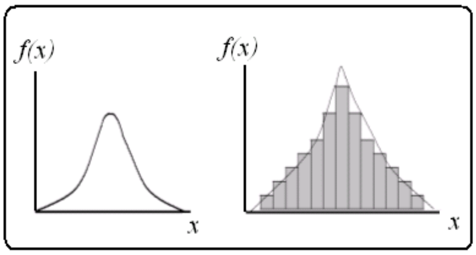
2- المتغيرات العشوائية المستمرة Continuous Random Variables

المتغير العشوائي المستمر هو الذي يأخذ قيما متصلة، ويأخذ عدد لانهائي من القيم الممكنة له داخل مجاله، فإذا كان متغير عشوائي مستمر، ويقع في المدى *(a,b)*، أي أن: $\{X = x : a < x < b\}$ ، فإن للمتغير *X* عدد لانهائي من القيم تقع بين الحدين الأدنى والأعلى *(a,b)*، ومن الأمثلة على المتغيرات الكمية المستمرة ما يلي:

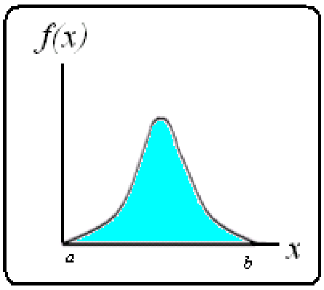
- كمية الألبان التي تنتجها البقرة في اليوم باللتر: $\{X = x : 10 < x < 40\}$
- المساحة المنزرعة بالأعلاف في المملكة بالألف هكتار $\{X = x : 1000 < x < 15000\}$
- فترة صلاحية حفظ الدجاج المبرد بالأيام، $\{X = x : 1 < x < 5\}$
- وزن الجسم بالكيلوجرام للأعمار من *(40-30)*، $\{X = x : 55 < x < 80\}$
- وهكذا الأمثلة على المتغير الكمي المستمر كثيرة.

التوزيع الاحتمالي للمتغير المستمر Continuous Probability

عند تمثيل بيانات المتغير الكمي المستمر في شكل مدرج تكراري النسبي، نجد أن شكل هذا المدرج هو أقرب وصف لمنحنى التوزيع الاحتمالي للمتغير المستمر، وكلما ضاقت الفترات بين مراكز الفئات، يمكن الحصول على رسم دقيق للمنحنى الخاص بدالة احتمال المتغير المستمر، كما هو مبين بالشكل التالي:



والمساحة أسفل المنحنى تعبر عن مجموع الاحتمالات الكلية، ولذا تساوي هذه المساحة الواحد الصحيح، وتسمى الدالة *f(x)* بدالة كثافة الاحتمال *(p.d.f)* Probability Distribution Function، وبفرض المتغير العشوائي المستمر يقع في المدى: $X = \{x : a \prec x \prec b\}$ وأن منحنى هذه الدالة يأخذ الصورة التالية:



المتوسط والتباين في التوزيع الاحتمالي المستمر:

إذا كانت $f(x)$ هي دالة كثافة الاحتمال للمتغير العشوائي x ، $a < x < b$ فإن التوقع الرياضي للدالة $h(x)$ تأخذ الصورة التالية: $E(h(x)) = \int_a^b h(x)dx$

ومن ثم يمكن كتابة معادلة الوسط والتباين كما يلي:

$$\mu = E(x) = \int_a^b xf(x)dx$$

$$\sigma^2 = E(x^2) - u^2, E(x^2) = \int_a^b x^2 f(x)dx$$

المحاضرة 4

التوزيعات الإحتمالية (الجزء الأول)

التوزيعات الإحتمالية للمتغيرات المنفصلة:

وكما أوضحنا في المحاضرة السابقة أن المتغير العشوائي المنفصل (مقارنة بالمتصل) هو المتغير الذى يمكن أن يأخذ قيما محدودة ومتميزة، وتسمى مجموعة كل القيم الممكنة للمتغير العشوائي واحتمالاتها المناظرة بالتوزيع الاحتمالي، ويكون مجموع الاحتمالات 1

أ- توزيع ذي الحدين:

يستخدم هذا التوزيع في الحالات التي يكون للظاهرة محل الدراسة نتيجتان فقط متنافيتان، النتيجة محل الاهتمام وتسمى بحالة النجاح، والأخرى تسمى بحالة الفشل، **ومن أمثلة ذلك:**

- عند إعطاء مريض نوع معين من الأدوية، لها نتيجتان: (استجابة للدواء، أو عدم استجابة)
- عند فحص عبوة بداخلها نوع معين من الفاكهة، لها نتيجتان (الوحدة إما أن تكون سليمة، أو تكون معيبة)
- عند إلقاء قطعة عملة، لها نتيجتان (ظهور الوجه الذي يحمل الصورة، أو الوجه الذي يحمل الكتابة)
- نتيجة الطالب في الاختبار (نجاح، رسوب)
- استخدام المزارع لبرنامج معين في الزراعة (يستخدم، أو لا يستخدم)

شكل التوزيع الاحتمالي ثنائي الحدين:

إذا كررت محاولة من المرات، بحيث أن كل محاولة لها نتيجتان فقط متنافيتان هما:

- النتيجة محل الاهتمام " حالة نجاح " وتتم باحتمال ثابت في كل محاولة هو p
- النتيجة الأخرى " حالة فشل " وتتم باحتمال ثابت أيضا هو $q = 1 - p$

وبافتراض أن هذه المحاولات مستقلة، بمعنى أن نتيجة كل محاولة ليس لها علاقة بنتيجة المحاولة الأخرى، وإذا كان المتغير العشوائي X يعبر عن عدد حالات النجاح "عدد النتائج محل الاهتمام" في الـ n محاولة، فإن مدى المتغير العشوائي X والذي يعبر عن عدد حالات النجاح هو: $X : \{x = 0, 1, 2, ..., n\}$

إذا فتوزيع ذو الحدين هو أحد التوزيعات الاحتمالية المنفصلة، ويستخدم لإيجاد احتمال وقوع حدث معين (نجاح) عدداً من المرات مقداره X من بين n من المحاولات لنفس التجربة ونرمز لهذا الاحتمال بالرمز P(X) وذلك عندما تتحقق الشروط التالية:

- هناك ناتجان ممكنان فقط ومتنافيان لكل محاولة
- المحاولات وعددها n مستقلة عن بعضها البعض
- احتمال وقوع الحدث المعين فى كل محاولة (النجاح) **P** ثابت ولا يتغير من محاولة لأخرى .

فبالتالي يمكن حساب الاحتمال من خلال المعادلة التالية: $P(X) = \dfrac{n!}{X!(n-X)!} P^{x} (1-P)^{n-X}$

حيث n ! (وتقرأ " مضرب n ") = n. (n-1). (n-2)…3.2.1 ,0! =1.

ويكون متوسط توزيع ذى الحدين $\mu = np$

وانحراف المعيارى $\sigma = \sqrt{np(1-p)}$

تحديد شكل التوزيع:

يتحدد شكل التوزيع ثنائي الحدين وفقا لقيمة احتمال النجاح كما يلي:

- إذا كان $p = 0.5$ فإن التوزيع الاحتمالي ثنائي الحدين يكون **متماثل.**
- إذا كان $p < 0.5$ فإن التوزيع الاحتمالي ثنائي الحدين يكون **موجب الالتواء.**
- إذا كان $p > 0.5$ فإن التوزيع الاحتمالي ثنائي الحدين يكون **سالب الالتواء.**

$p > 0.5$ الاحتمال أكبر من 0.5

مثال:

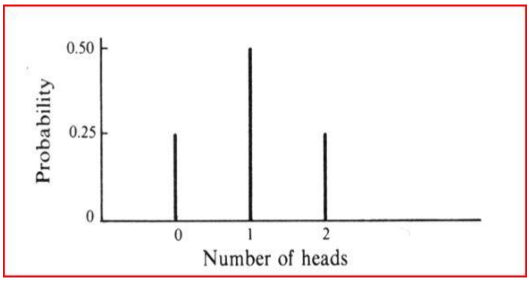
عند رمى عملة متوازنة مرتين فإن النواتج الممكنة هي TT,TH,HT,HH وإذن:

$$P(0H) = \frac{1}{4} \qquad \mathrm{P(1H)} = \frac{1}{2} \quad \text{and} \quad \mathrm{P(2H)} = \frac{1}{4}$$

وهكذا فإن عدد الصور متغير عشوائى منفصل ، وتمثل مجموعة كل النواتج الممكنة مع احتمالاتها المناظرة توزيعاً احتمالياً منفصلا، أنظر الجدول التالي:

الاحتمال	إمكانية حدوثها	عدد الصور
0.25	TT	0
0.50	TH, HT	1
0.25	HH	2

ويمكن كذلك تمثيل ذلك من خلال الرسم التالي:



مثال:

باستخدام توزيع ذي الحدين يمكننا إيجاد احتمال الحصول على 4 صور في 6 رميات لعملة متوازنة كالآتي:

$$P(4) = \frac{6!}{4!(6-4)!}(1/2)^4(1/2)^2 = \frac{6\bullet5\bullet4\bullet3\bullet2\bullet1}{4\bullet3\bullet2\bullet1\bullet2\bullet1}(1/16)(1/4) = 15(1/64) = \frac{15}{64} \cong 0.23$$

إن عدد الصور المتوقع في ست رميات هو: $\mu = np = (6)(1/2) = 3$

ويكون الانحراف المعياري للتوزيع الاحتمالي لست رميات هو:

$$\sigma = \sqrt{np(1-p)} = \sqrt{(6)(1/6)(1/2)} = \sqrt{6/4} = \sqrt{1.5} \cong 1.22 heads$$

ب- توزيع بواسون:

توزيع بواسون هو توزيع احتمالي منفصل آخر يستخدم لتحديد احتمال وقوع عدد معين من النجاحات في وحده الزمن، وذلك عندما تكون الأحداث أو "النجاحات " مستقلة عن بعضها البعض وعندما يبقى متوسط عدد النجاحات ثابتاً لوحدة الزمن . عندئذ:

$$P(x)=\frac{e^{-\mu}\mu^{x}}{x!}$$
$$,\quad x=0,1,2,\ldots$$

حيث :

- x = العدد المعين من النجاحات.
- P (x) = احتمال عدد x من النجاحات.
- e = أساس نظام اللوغاريتمات الطبيعي وتوجد في بعض الآلات الحاسبة، وقيمتها هي: $e=2.718$ تقريبا، ويمكن حساب قيمتها باستخدام الآلة الحاسبة.
- $x!$ = مضروب العدد x " ويساوي: $x!=x(x-1)(x-2)\ldots 3\times 2\times 1$

ويكثر استخدام هذا التوزيع في الحالات التي تقع فيها الأحداث وفقا لمعدلات زمنية، وكذلك في حالة الأحداث نادرة الوقوع، **ومن أمثلة ذلك:**

- عدد الوحدات التي تستهلكها الأسرة من سلعة معينة خلال الشهر $X:\{x=0,1,2,\ldots\}$
- عدد مرات ري نوع معين من المحاصيل الزراعية خلال الموسم $X:\{x=0,1,2,\ldots\}$
- عدد العملاء الذين يتم خدمتهم البنكية كل 10 دقائق $X:\{x=0,1,2,\ldots\}$
- عدد مرات زيارة المريض للطبيب كل سنة $X:\{x=0,1,2,\ldots\}$
- عدد مرات تناول الأسرة للحوم الحمراء خلال الأسبوع $X:\{x=0,1,2,\ldots\}$
- عدد أخطاء الطباعة لكل صفحة من صفحات الكتاب $X:\{x=0,1,2,\ldots\}$

شكل التوزيع الاحتمالي البواسوني

إذا كان متوسط عدد مرات وقوع حادث وفقا لمعدل زمني معين هو μ ، وكان المتغير العشوائي X يعبر عن عدد مرات وقوع الحادث وفقا لهذا المعدل، فإن مدى المتغير العشوائي X هو: $X:\{x=0,1,2,\ldots\}$ ، وهذا المدى عبارة عن فئة مفتوحة من اليمين، فإن الاحتمال $P(X=x)=f(x)$ والذي يعبر عن احتمال وقوع الحادث عدد من المرات وفقا لهذا المعدل.

مثال:

إذا كان من المعلوم أن عدد الوحدات التي تستهلكها الأسرة من سلعة معينة خلال الشهر تتبع توزيع بواسون بمتوسط 3 وحدات شهريا، إذا عرف المتغير العشوائي X بأنه عدد الوحدات التي تستهلكها الأسرة خلال الشهر من هذه السلعة.

المطلوب:

- ما نوع المتغير العشوائي؟
- اكتب شكل دالة الاحتمال لهذا المتغير.
- احسب الاحتمالات التالية:
  - احتمال أن الأسرة تستهلك وحدتين خلال الشهر؟
  - احتمال أن أسرة ما تستهلك 3 وحدات على الأكثر خلال الشهر؟
- احسب الوسط الحسابي، والانحراف المعياري لعدد الوحدات المستهلكة.
- حدد شكل التوزيع.

الحل:

عدد الوحدات التي تستهلكها الأسرة متغير كمي منفصل، ومدى هذا المتغير في هذه الحالة هو: $X:\{x=0,1,2,3,\ldots\}$

شكل دالة الاحتمال:

بما أن متوسط عدد الوحدات التي تستهلكها الأسرة خلال الشهر هو: $\mu = 3$ ، إذا دالة الاحتمال هي:

$$P(x)=\frac{e^{-\mu}\mu^{x}}{x!}$$

$$=\frac{e^{-3}3^{x}}{x!} \quad , \quad x=0,1,2,\ldots$$

حساب الاحتمالات:

حساب احتمال أن أسرة ما تستهلك وحدتين خلال الشهر، $p(2)$

$$P(2)=\frac{e^{-3}3^{2}}{2!}=\frac{0.0498(9)}{2\times1}=0.22404$$

احتمال أن أسرة ما تستهلك 3 وحدات على الأكثر خلال الشهر هو:

$$P(X\le 3)=p(3)+p(2)+p(1)+p(0)$$

$$=\left[\frac{3^{3}}{3!}+\frac{3^{2}}{2!}+\frac{3^{1}}{1!}+\frac{3^{0}}{0!}\right]\left[\frac{0.0498}{1}\right]$$

$$=[0.0498]\left(\frac{27}{6}+\frac{9}{2}+\frac{3}{1}+\frac{1}{1}\right)=0.0498(13)=0.6474$$

حساب الوسط الحسابي، والانحراف المعياري لعدد حالات الاستجابة:

الوسط الحسابي (μ) في حالة التوزيع البواسون هو معلمة معطاة هي: $\mu=3$

في هذا التوزيع، فإن التباين يساوي الوسط الحسابي: أي أن: $\sigma^{2}=\mu=3$

ومن ثم يكون الانحراف المعياري هو: $\sigma=\sqrt{\mu}=\sqrt{3}=1.732$

ويمكن حساب معامل الاختلاف النسبي، بتطبيق المعادلة التي سبق استخدامها في الفصل السابق، وهو:

$$V.C=\frac{\sigma}{\mu}\times100=\frac{1.732}{3}\times100=57.7\%$$

تحديد شكل التوزيع: دائما التوزيع البواسون موجب الالتواء

مثال:

يتلقى قسم شرطة فى المتوسط 5 مكالمات في الساعة فيكون احتمال تلقي مكالمتين في ساعة مختارة عشوائياً هو :

$$P(x) = \frac{e^{-\mu}\mu^{x}}{x!}$$

$$= \frac{e^{-5}5^{x}}{x!} = \;, \quad x = 0,1,2,\ldots$$

$$= \frac{(25)(0.00674)}{(2)(1)} = 0.08425$$

المحاضرة 5

التوزيعات الإحتمالية (الجزء الثاني)

يعني **التوزيع الإحصائي** الشكل الذي تأخذه مجموعة البيانات، وشكل البيانات مهم جدا في تحليلها ووصفها وكخطوة تسبق قرار استخدام أي اسلوب احصائي.
ويرتبط التوزيع الاحصائي عادة بنوعين من البيانات المتصلة والمنفصلة، ويناسب النوع المنفصل المقاييس الاسمية والرتبية، وهناك بعض المقياس المنفصلة ثنائية أي انه لايوجد بها الا قيميتين، وهي لا تسمي توزيعات طبيعية وانما تسمى توزيعات ثنائية، ومن أهم مقاييس التوزيعات المنفصلة مقياس ذو الحدين وذلك عائد لان الاجابة على المقياس الاسمي اما نعم أو لا ، ولذلك غالبا ما يرمز لها في الحاسب بصفر (غياب الصفة) [ذكور – لا] أو 1 (وجود الصفة) [اناث – نعم].
أما التوزيعات الاحصائية المتصلة فهي ذات أهمية كبيرة في العلوم الإحصائية وذلك لأن اغلب الاختبارات الاحصائية تتعامل مع هذا النوع من البيانات.

التوزيعات الإحتمالية للمتغيرات المتصلة:

هناك بعض التوزيعات الاحتمالية المتصلة لها دوال كثافة احتمال محددة ومنها:

- التوزيع الطبيعي
- التوزيع الطبيعي (القياسي) المعياري
- توزيع t

وسنقوم في هذه المحاضرة بتناول هذه التوزيعات بشئ من التوضيح والتفصيل:

وكما أوضحنا أن **المتغير العشوائي المتصل x** هو ذلك المتغير الذي يمكن أن يأخذ عدداً لا نهائياً من القيم المعلومة، واحتمال أن تقع **x** داخل أى فترة يمثلها مساحة التوزيع الاحتمالي (ويسمى أيضاً دالة الكثافة) داخل هذه الفترة، والمساحة الكلية تحت المنحنى (الاحتمال) تساوى **1**

أ- التوزيع الطبيعي:

هو أفضل وأكثر التوزيعات الاحتمالية المتصلة استخداماً في النواحي التطبيقية، ومنها الاستدلال الإحصائي شاملا التقدير، واختبارات الفروض، كما أن معظم التوزيعات يمكن تقريبها إلى هذا التوزيع.

والتوزيع الطبيعي هو توزيع احتمالي متصل وهو جرسي الشكل ومتماثل حول الوسط الحسابي، ويمتد إلى مالا نهاية فى الاتجاهين، ولكن معظم المساحة (الاحتمال) تتركز حول الوسط الحسابي

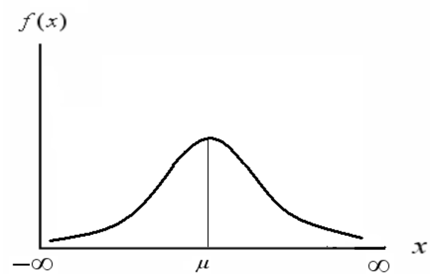
خصائص التوزيع الطبيعي:

يعتبر التوزيع الطبيعي من أهم أنواع التوزيعات الاحصائية المتصلة **ومن خصائصه** انه:

- توزيع جرسي أي يشبه الجرس.
- توزيع متصل
- توزيع متماثل حول الوسط
- الالتواء (الاطراف) والتفلطح (القمة) يساوي صفر.
- ومن أهم صفاته أن يتصف بمنوال ووسط ووسيط واحد وذات قيم متساوية بمعنى أن الجزء الذي على يمين الوسط مطابق للجزء الايسر
- الذيلين الايمن والايسر يقتربان من الخط الافقي ولكن لا تلامسه
- المساحة الكلية تحت المنحنى تساوي واحد صحيح
- منحنى دالة الاحتمال للتوزيع الطبيعي له خاصية شكل الجرس. ويتحدد شكل الجرس تماماً لأي توزيع طبيعي خاصة إذا علمنا الوسط الحسابي μ والانحراف المعياري σ لهذا التوزيع.

- تدل قيمة μ على مكان مركز الجرس، كما تدل σ على كيفية الانتشار.
- القيمة الصغيرة لـ σ تعني أنه لدينا جرس طويل مدبب، والقيمة الكبيرة لـ σ تعني أن الجرس قصير ومفرطح.

والشكل التالي يوضح ذلك:



والتوزيع الطبيعي وتطبيقاته الاحصائية ليس موضوعا جديدا بل عرف منذ القرن السابع عشر الميلادي ومن ابرز الدراسات المعروفة تلك الدراسة البريطانية التي اخذت اطوال 8585 من الافراد البريطانيين في القرن التاسع عشر وعمل هذا المنحنى وبالتالي تم اعتبار هذه العينة تمثل التوزيع الطبيعي.

معالم هذا التوزيع:

توجد معلمتين لهذا التوزيع هما :

الوسط الحسابي : $E(x)=\mu$ والتباين : $\operatorname{var}(x)=\sigma^2$

ومن ثم يعبر عن توزيع المتغير بالرموز : $x \sim N(\mu,\sigma^2)$ ويعني ذلك أن المتغير العشوائي x يتبع التوزيع الطبيعي بمتوسط μ ، وتباين σ^2

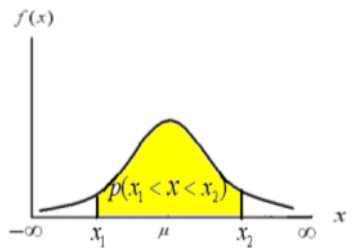
شكل دالة كثافة الاحتمال:

إذا كان لدينا توزيع طبيعي ذو وسط حسابي μ وانحراف معياري σ فإن معادلة منحنى دالة كثافة الاحتمال تكون على الصورة التالية:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},-\infty<x<\infty,\pi=22/7$$

كيفية حساب الاحتمالات:

بفرض أن الاحتمال المطلوب حسابه هو $p(x_1<X<x_2)$ وهذا الاحتمال يحدد بالمساحة التالية:



وحيث أن هذا التوزيع من التوزيعات المستمرة، فإن هذه المساحة (الاحتمال) تحسب بإيجاد التكامل التالي:

$$p(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx = \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

وهذا التكامل يصعب حسابه، ومن ثم لجأ الإحصائيين إلى عمل تحويلة رياضية Transform، يمكن استخدام توزيعها الاحتمالي في حساب مثل هذه الاحتمالات، وهذه التحويلة هي: $z = \left(\frac{x-\mu}{\sigma}\right)$

ويعرف المتغير الجديد بـ z وهو المتغير الطبيعي القياسي Standard Normal Variable ، وهذا ما سوف ندرسه في الجزء التالي من هذه المحاضرة بما يسمى بالتوزيع الطبيعي القياسي أو المعياري

ب- التوزيع الطبيعي القياسي (المعماري):

هو توزيع طبيعي وسطه الحسابي 0 وانحرافه المعياري 1 (أي أن $\sigma = 1, \mu = 0$)

ويمكن تحويل أي توزيع طبيعي (بوحدات x) إلى توزيع طبيعي قياسي (بوحدات z) . وتحت هذه الشروط ، فإن 68.26% من المساحة (الاحتمال) تحت المنحنى الطبيعي القياسى تقع بين إحداثيين رأسيين يبعدان بمقدار انحراف معيارى واحد عن الوسط الحسابى (أى داخل $\mu \pm 1\sigma$)، 95.54% تقع بين $\mu \pm 2\sigma$ ، 99.74% تقع بين $\mu \pm 3\sigma$

ولإيجاد الاحتمالات (المساحات) فى مسائل تحتوى على التوزيع الطبيعى ، فإننا نحول أولا قيم **x** إلى قيم **z** المناظرة لها ، من خلال المعادلة التالية: $Z = \frac{X-\mu}{\sigma}$

ثم نكشف عن قيمة **z** فى الجداول المخصصة لذلك، ويعطي هذا قيمة الجزء من المساحة (الاحتمال) تحت المنحنى بين قيمة الوسط الحسابى وقيمة **z**

- احتمال وقوع أية مشاهدة على بعد انحراف معياري واحد من الوسط الحسابي هو 0.6827
- احتمال وقوع أي مفردة على بعد إنحرافين معياريين من الوسط الحسابي هو 0.9545
- احتمال وقوع أية مفردة على بعد ثلاثة انحرافات معيارية من الوسط الحسابي هو 0.9973

والشكل التالي يوضح ذلك:

العلاقة بين التوزيع الطبيعي والتوزيع الطبيعي القياسي

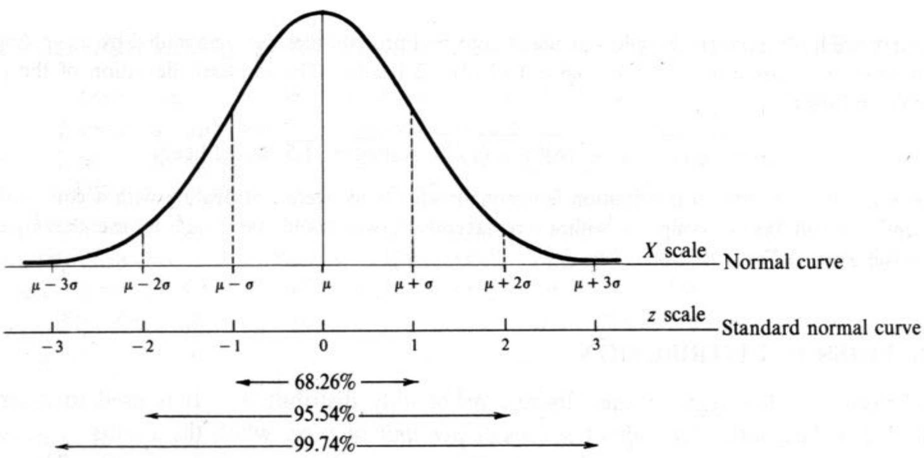


Fig. 3-4

فالمساحة (الاحتمال) تحت المنحنى الطبيعى المعيارى z=0 وz=1.96 نحصل عليها مقابلة للقيمة 1.96 في جدول التوزيع الطبيعي، ففي عمود z نبدأ بالقيمة 1.9 ونتحرك في الصف المناظر لها حتى نصل إلى العمود المعنون 06.، وتكون القيمة التي نحصل عليها 0.4750 .

ويعنى هذا أن %47.50 من المساحة الكلية (1 أو %100) تحت المنحنى تقع بين z=0, z=1.96 المساحة المظللة في الشكل فوق الجدول) . ولأن التوزيع متماثل ، فإن المساحة بين z=0 , z=-1.96 (ليس مدرجة في الجدول) هي أيضاً 0.4750 أو 47.50%

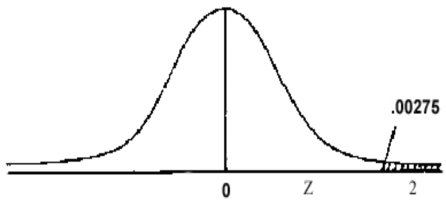
مثال

احتمال أن تكون قيمة Z أكبر من 2

الحل:

حيث أن احتمال أن تكون Z أقل من صفر =0.5000 ومن الجدول احتمال Z في (0،2) =0.47725 اذن احتمال أن تكون قيمة Z أكبر من 2 هي :

0.02275 = 0.47725 - 0.5000



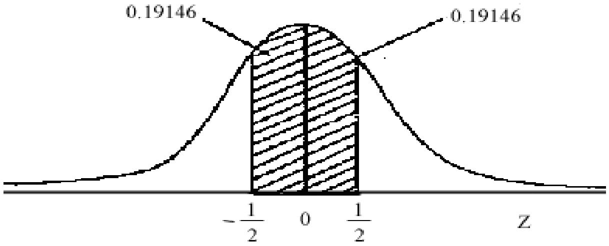
مثال:

- احتمال أن تقع Z بين صفر و 0.5.
- احتمال أن تقع Z بين 0.5 و 0.5-.

الحل:

من الرسم المساحة المظللة بين 0 و 0.5 تمثل احتمال أن تقع Z بين (0 و 0.5)، والمساحة المظللة إلى شمال 0 ويمين 0.5 - هي احتمال أن تقع Z في الفترة (0.5- ,0) . واحتمال أن تقع Z في الفترة (0, 05) = المساحة المقابلة لقيمة Z = 0.5 هي 0.19146 ، كذلك احتمال أن تقع Z في الفترة 0.19146 = (0, -0.5)

اذن احتمال أن تقع Z في الفترة (0.5 و 0.5-) = 2 × 0.19146 = 0.38292 وهي تتمثل بالمساحة المظللة في الرسم التالي



مثال

قامت إحدى الشركات بإجراء اختبار للمتقدمين لشغل بعض الوظائف الشاغرة بها، إذا علمت أن درجات هذا الاختبار تتبع توزيعا معتدلا وسطه الحسابي 500 وانحرافه لمعياري 100 درجة وأن أحد الممتحنين قد اختير عشوائيا ، ما هو احتمال أن تكون درجته أكبر من 700؟

الحل:

إذا كانت X تركز لدرجة أي ممتحن، فإن X تتبع توزيعا معتدلا وسطه الحسابي 500 درجه وإنحرافه المعياري 100 درجه، وباستخدام المعادلة الخاصة بالدرجة المعيارية $Z = \frac{X-\mu}{\sigma}$

نجد أن القيمة المعيارية Z للقيمة 700 هي: $Z = \frac{X-\mu}{\sigma} = \frac{700-500}{100} = \frac{200}{100} = +2$

وبالتالي يمكننا صياغة السؤال السابق كما يلي: إذا اختير أحد الممتحنين عشوائيا، فما هو احتمال أن تزيد درجته عن الوسط الحسابي بأكثر من انحرافين معياريين؟ للإجابة على هذا السؤال فإننا نستخدم الشكل التالي:

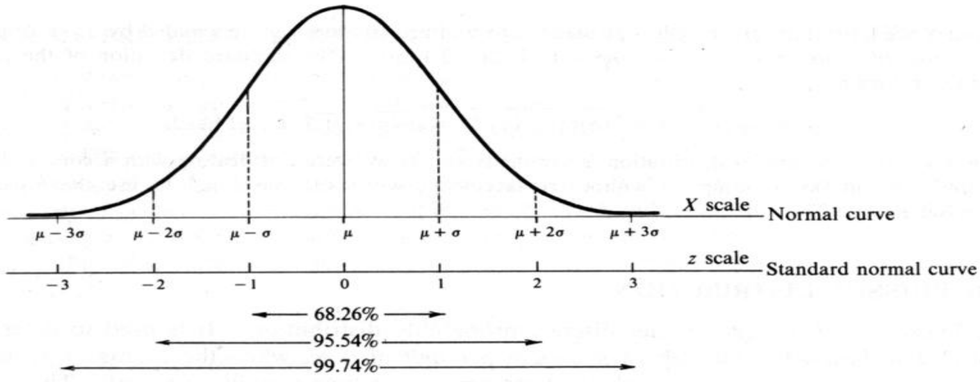


Fig. 3-4

والذي يبين أن المساحة تحت المنحنى المحصورة بين انحرافين معياريين من الوسط الحسابي 95.45% وبالتالي تكون المساحة المتبقية من المنحنى أي مساحة طرفي المنحنى هي (1-0.9545=0.0455)، ونتيجة لتماثل المنحنى حول وسطه الحسابي، فإن مساحة الطرف الأيمن للمنحنى تساوي مساحة طرفه الأيسر، أي تساوي (0.0455/2=0.02275)، لذا فإن المساحة تحت المنحنى على يمين ($\mu+2\sigma$) من الوسط الحسابي (أي على يمين $+2\sigma$) تساوي 0.02275 وهي قيمة احتمال أن تكون درجة الشخص الذي اختير عشوائيا أكبر من 700

كيفية استخدام جدول توزيع الاحتمالات المتجمعة للمتغير العشوائي Z

وبمعرفة القيمة المعيارية Z يمكننا أن نحصل على احتمالات أي متغير عشوائي معتدل، والتعبير Z <+2 يعني أن القيمة المشاهدة تقع على مسافة أقل من $+2\sigma$ على يمين الوسط الحسابي، أيضا فإن التعبير -1< Z <+3 يعني أن القيمة المشاهدة تقع بين $\mu-1\sigma$ و $\mu+3\sigma$ ومن الواضح نه لايمكن استخدام الشكل السابق لتحديد الاحتمالات المطلوبة بسهوله كافية، لذا يستخدم جدول توزيع الاحتمالات المتجمعة للمتغير العشوائي Z لإيجاد الإحتمالات المطلوبة، ويعطي العمود الأول بيسار الجدول مع الصف العلوي قيم Z المختلفة إلى رقمين عشريين فقط، والرقم الأول بالعمود الأول على يسار الجدول هو 0.0 والرقم الأول بالصف العلوي من الجدول هو 0.00 ومجموع هذين الرقمين يعطينا القيمة المعيارية Z=0.00 والاحتمال المتجمع المناظر هو 0.5000 أي أن P(Z > 0.000)=0.5000 وهذه بطبيعة الحال نتيجة منطقية لأن توزيع Z متماثل حول وسطه الحسابي وهو الصفر، وبالتالي لا يوجد أي احتمال متجمع بالجدول قيمته أقل من 0.5000

مثال:

أوجد احتمال أن Z أقل من (>) 1.64

الحل:

الجدول التالي جزء من جدول الاحتمالات المتجمعة للتوزيع المعتدل المعياري

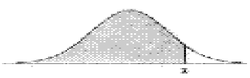
Z	.00	.01	.02	.03	.04	.05
1.6					0.9495	

ويتم الحصول على القيمة المعيارية Z بجمع القيمتين المناسبتين الموجودتين بالصف العلوي والعمود الأول بيسار الجدول، ويحوي العمود الأول من جهة اليسار على قيم تصل إلى رقم عشري واحد فقط، بينما يحوي الصف العلوي على الرقم المئوي.

فالإحتمال المتجمع المناظر للقيمة 1.64 يوجد أمام الصف 1.6 وتحت العمود 0.04 (لاحظ أن 1.6 + 0.04 = 1.64) وهي قيمة Z المطلوب إيجاد الاحتمال المتجمع عندها، وهذا الإحتمال هو 0.9495 ، أي أن P(Z<1.64)=0.9495 وهذا هو الإحتمال المتجمع للمتغير Z من (-∞) إلى 1.64

والجدول التالي يوضح ذلك:

Tables of the Normal Distribution



Probability Content from -oo to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

مثال:

أوجد أن احتمال أن Z أكبر من (>) 1.64 .

الحل:

إن مجموع الاحتمالات المتجمعة لأي متغير عشوائي يساوي (1)، وحيث أن المساحة الكلية تحت منحنى أي متغير عشوائي مستمر تمثل مجموع الاحتمالات، لذا فإن هذه المساحة تساوي (1) لذا فإن:

P(Z>1.64)=1-P(Z<1.64)=1-0.9495=0.0505

مثال:

أوجد المساحة تحت المنحنى المعتدل المعياري على يمين Z=-1.65 .

الحل:

المنحنى المعتدل كما أوضحنا منحنى متماثل حول الصفر، وبالتالي فإن المساحة تحت المنحنى على يمين 1.65- تساوي المساحة تحت المنحنى على يسار 1.65 ، أي أن P(Z>-1.65)=P(Z<1.65) وباستخدام جدول التوزيع الطبيعي نجد أن P(Z<1.65)=0.9505 أي أن الإحتمال المتجمع من 1.65- إلى ∞ +

أي أن: P(Z<-1.65)=1-P(Z<1.65)=1-0.9505=0.0495

استخدامات التوزيع الطبيعي القياسي:

يستخدم التوزيع الطبيعي القياسي في التعامل مع الكثير من المشاكل العملية وإيجاد القيم الإحتمالية لها وإليك بعض الأمثلة على ذلك:

مثال:

افترض أن إدارة المرور بالاحساء وضعت جهازا للرادار على طريق الدمام عند مدخل المبرز وذلك لضبط السيارات المسرعة في فترة معينة من اليوم، افترض أن X تمثل السرعة في الساعة للسيارات التي تمر بمدخل المبرز في فترة عمل الرادار، إذا كانت X تتوزع توزيعا معتدلا وسطه الحسابي 60 ميلا وتباينه 25 ميلا، أوجد التالي:

١. نسبة السيارات التي تقل سرعتها عن 50 ميلا في الساعة .
٢. نسبة السيارات التي تزيد سرعتها عن 65 ميلا في الساعة.
٣. نسبة السيارات التي تكون سرعتها بين 60 ميلا و 77.45 ميلا في الساعة .
٤. عدد السيارات التي تكون سرعتها بين 60 ميلا و 77.45 ميلا من بين 10000 سيارة .

الحل:

١. نسبة السيارات التي تقل سرعتها عن 50 ميلا في الساعة:

$$P(X<50)=P(\frac{x-\mu}{\sigma}<\frac{50-60}{\sqrt{25}})=P(Z<-2)=1-0.9772=0.0228$$

٢. نسبة السيارات التي تزيد سرعتها عن 65 ميلا في الساعة:

$$P(X>65)=P(\frac{x-\mu}{\sigma}>\frac{65-60}{\sqrt{25}})=P(Z>1)=1-P(Z\le 1)=1-0.8413=0.1587$$

٣. نسبة السيارات التي تكون سرعتها بين 60 ميلا و 77.45 في الساعة:

$$P(60\le X\le 77.45)=P(\frac{60-60}{\sqrt{25}}\le Z\le\frac{77.45-60}{\sqrt{25}})$$

$$=P(0\le Z\le 3.49)=P(Z\le 3.49)-P(Z\le 0)$$

$$=0.9998-0.5000=0.4998$$

٤. عدد السيارات المتوقع سرعتها بين 60 ميلا و 77.45 ميلا من بين 10000 سيارة: 1000(0.4998)=4998

ج- توزيع t ستيودنت

توجد عائلة أخرى من المتغيرات العشوائية المتصلة المستخدمة في الإحصاء الاستدلالي وهي مجموعة المتغيرات العشوائية t ويعتبر وليم جوست w.s. Gosset هو أول من درس تلك المتغيرات حيث سجل نتائجه عام 1908 تحت اسم مستعار هو student ولذلك يسمى توزيع t في بعض الأحيان بتوزيع ستيودنت.

ويرمز لهذه العائلة من التوزيعات بالرموز (t1,t2,t3 ......tdf) كما يرمز لدرجات حريتها بالرمز (V حرف إغريقي ينطق نيو) وهي تأخذ القيم (1,2,3،....،df)

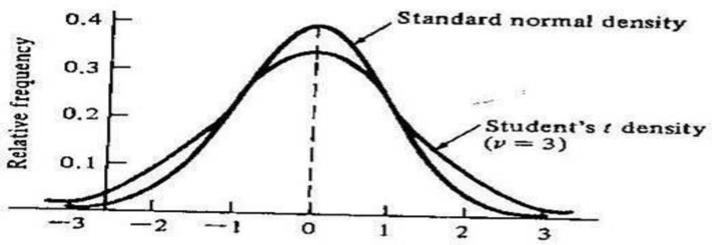
الفرق بين توزيع t والتوزيع الطبيعي:

يختلف المتغير العشوائي t عن المتغير العشوائي الاعتدالي ،حيث يتحدد المتغير العشوائي الاعتدالي بمعلمين هما الانحراف المعياري والمتوسط، بينما يتحدد المتغير العشوائي t بمعلم واحد فقط هو درجة الحرية.

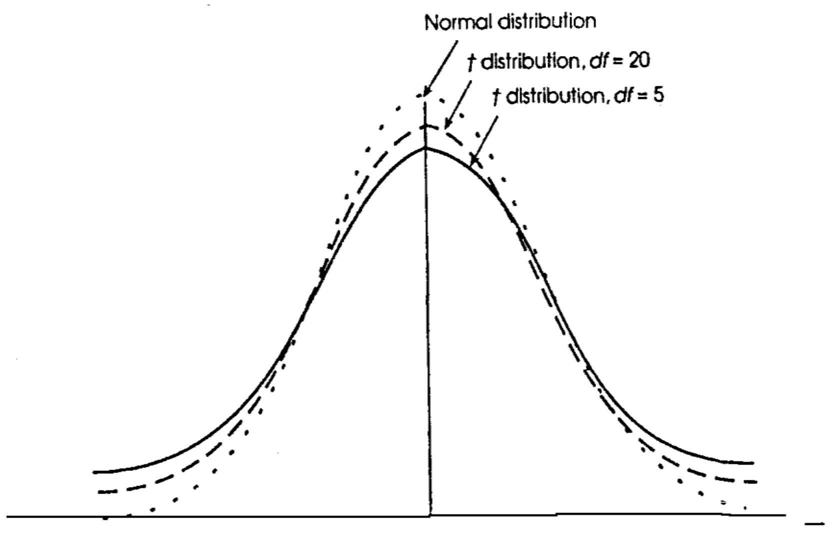
ولا شتقاق المتغير العشوائي t من المتغير العشوائي الاعتدالي، فإن ذلك يتطلب معرفة قيمة المتوسط μ للمتغير العشوائي الاعتدالي، بينما لانحتاج إلي معرفة انحرافه المعياري.

ولنفرض أن قيمة متغير العشوائي الاعتدالي التي تم ملاحظتها n من المرات (n>=z)وأن هذه الملاحظات البالغ عددها n تكون عينة متوسطها μ وانحرافها المعياري s وحسبنا قيم المتغير العشوائي t باستخدام الصيغة التالية: t= X-μ/(s/n)

وتتحدد درجة حرية المتغير العشوائي t بأنها تساوي (n-1) كذلك فإنه لكل قيم n نجد أن توزيع t له قيمة واحدة عند النقطة صفر، وهو توزيع متماثل يقل تدريجيا كلما اتجهنا ناحيتي الذيلين الأيمن والأيسر، وهذا ما يوضحه الشكل التالي:



ونلاحظ من الشكل السابق ان توزيع t يشبه توزيع z فيما عدا أنه أكثر انتشارا diffuse لأنه أكثر كثافة عند الذيلين وخاصة عندما تكون n صغيرة ، اما إذا كانت n كبيرة فإن توزيع t يكون أقل انتشارا وأكثر قربا من شكل توزيع z ، **وبزيادة درجات الحرية يقترب توزيع t من التوزيع الاعتدالي**، وهذا ما يوضحه الشكل:



خصائص توزيع t :

- متوسط المتغير العشوائي t يساوي صفر لكل درجات الحرية (n-1) . وهذا يعني أن $\mu = 0$
- الانحراف المعياري للمتغير العشوائي t درجات حرية أكبر من اثنين يساوي: $\sigma = s/s-2$

حيث df هي درجة حرية المتغير العشوائي t .

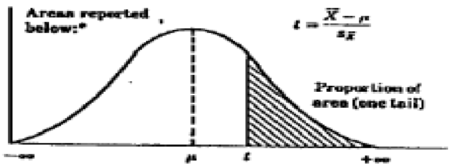
ويتبين من المعادلة السابقة للانحراف المعياري أنه كلما زادت درجات حرية المتغير العشوائي t بحيث تصل إلى 30 فأكثر، فإن الانحراف المعياري يقترب من الواحد الصحيح، وبصفة عامة فإن الإنحراف المعياريي لتوزيع t يساوي 1.035 أو أقل.

ولذلك فإن التوزيع الاحتمالي للمتغير يكون قريبا جدا من التوزيع الاحتمالي للمتغير العشوائي z وبصفة خاصة عندما تكون df>30 وفي هذه الحالة نستخدم جدول z للإجابة على الأسئلة الاحتمالية حول المتغير العشوائي t.

الجدول أدناه يعطي قيمة t

المقابلة للمساحة المظللة وقيمتها α

Proportions of Area for the t Distributions



df	0.10	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898

df	0.10	0.05	0.025	0.01	0.005
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

***t* Table**

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

مثال :

احسب القيمة الحرجة (نقطة القطع) بتوزيع t لدرجات حرية 8 ومستوى الدلالة 10 . (الاحتمال بالذيل الأيمن)

الحل :

بالبحث في الجدول توزيع tعند درجات 8 والعمود الخاص بمستوى الدلالة 10 .نجد أن القيمة عن تقاطع الصف و العمود تساوي 1.397

$P(t8 >=1.397)=.10$

$P(t8<=1.397)=.90$

ويختلف الجدول الاحصائي t عن الجدول الاحصائي لتوزيع ذي الحدين ، حيث يعطي جدول t نقاط القطع (القيم الحرجة) بينما يعطي جدول ذي الحدين الاحتمالات . مع زيادة حجم العينة ومن ثم درجات الحرية فإن: $t_{\alpha} = -t_{1-\alpha}$

مثال:

أوجد نقطة القطع العليا للمتغير العشوائي t عندما تكون درجات الحرية 20 ومستوى الدلالة 0.95

الحل :

$t_{(20,0.95)}=-t_{(20,0.05)}=1.725$

المحاضرة 6

توزيعات المعاينة Sampling Distribution's (الجزء الأول)

تهتم نظرية العينات بدراسة العلاقة بين المجتمع والعينات المسحوبة منه فيما يسمى **بالاستدلال الإحصائي** **statistical inference**

يعتبر **الاستدلال الإحصائي** من أهم الأدوات المساعدة على اتخاذ القرارات في الاقتصاد والأعمال والعلوم، ويشمل الاستدلال الإحصائي اختبار الفرضيات والتقدير.

ولكي يكون التقدير (واختبار الفروض) سليما، ينبغي أن يبنى على عينة ممثلة للمجتمع، ويمكن تحقيق ذلك **بالمعاينة العشوائية**، حيث يكون لكل مفردة فى المجتمع فرصة متكافئة للدخول في العينة.

وهناك عدة طرق لأخذ العينات من المجتمع لاستخدامها في الاستدلال الإحصائي **ومن أشهر هذه الطرق** هي العينة العشوائية وهي العينة التي تكون لكل مفرده من مفردات المجتمع نفس فرصة الاختيار في العينة.

فمثلاً نستعين بعينه مسحوبة من المجتمع لتقدير معالم هذا المجتمع مثل متوسطه أو تباينة أو غير ذلك. أو أعطاء عينه من المرضى بارتفاع الضغط، مثلاً دواء معين ثم قياس ضغطهم قبل وبعد تناولهم لهذا الدواء لمعرفة ما إذا كان هذا الدواء مفيد في خفض الضغط أم لا.

يقصد بالمجتمع أي مجموعات من المفردات تشترك في صفه أو صفات وتكون موضوع دراسة أو بحث فإن هذه المجموعة يطلق عليها إحصائيا مجتمع الدراسة أو اختصاراً **المجتمع Population**.

- والمجتمع قد يكون مجموعة ما من البشر أو أشجار أنواع معينه من الفاكهة أو الحيوانات الزراعية أو إنتاج دولة ما لسلع معينه خلال فترة زمنية محدده...الخ.
- والمجتمع قد يكون محدوداً إذا كان يمكن حصر عدد أفراده مثل سكان مدينة ما أو طلاب مرحلة تعليمية معينة
- وقد يكون المجتمع غير محدود (لانهائي) إذا كان لا يمكن حصر عدد أفراده مثل النجوم والكواكب أو الكائنات الحية بمياه المحيطات والأنهار

وعند دراسة صفة ما أو صفات معينه لمجتمع ما فإن البيانات الإحصائية عن تلك الصفة أو الصفات تجمع بأحد أسلوبين:

أولاً: أسلوب الحصر الشامل (census):

وفيه تجمع البيانات عن كل مفرده من مفردات المجتمع، وهذا الأسلوب يتطلب وفرة في الوقت والمال والمجهود الفني وتزداد هذه المتطلبات وتتضاعف كلما ازداد حجم المجتمع (عدد أفراد المجتمع). وهذا الأسلوب لا يتبع عادة إلا في حالة التعدادات التي تجريها الدول وتدعمها بإمكانيات ضخمه مثل تعدادات السكان والتعدادات الصناعية والتعدادات الزراعية.

الثاني: أسلوب المعاينة (Sampling method):

وفيه يتم جمع البيانات عن جزء من مفردات المجتمع يختار بطريقة أو بأخرى ويطلق عليه عينه (Sample) ثم بعد ذلك يتم تعميم نتائج الدراسة على المجتمع بأكمله. أي أن إسلوب العينة يقصد به دراسة خصائص المجتمع من خلال دراسة عينه مسحوبة منه، ونجاح هذا الإسلوب يعتمد على أن تحمل العينة أقصى درجة من دقة التمثيل للمجتمع المسحوبة منه.

بعض مزايا أسلوب المعاينة: يتميز أسلوب المعاينة عن أسلوب الحصر الشامل بمزايا عدديه منها:

١. يؤدي استخدام العينات العشوائية إلى خفض تكاليف الدراسات الميدانية بسبب صغر حجم العينة بالنسبة إلى حجم المجتمع وهو ما يؤدي إلى تخفيض الأعباء الإدارية والفنية التي تتطلبها أي دراسة ميدانية.

٢. يتحقق وفر واضح في الوقت الذي ينفق في دراسة ميدانية على أساس عينة بدلاً من الحصر الشامل وتتضح أهمية الوقت عندما نقوم بدراسة ظاهرة تتغير بمرور الوقت، فتكون البيانات المجموعة والنتائج وقت ظهورها غير مطابقة

لواقع المجتمع وتصبح النتائج ذات قيمه محدودة بعد أن فقدت عنصر المطابقة مع واقع الظاهرة وتوزيعها الحالي في المجتمع.

٣. في المجتمعات غير المحدودة (اللانهائية) مثل مجتمع الكائنات الحية في البحار والمحيطات لا يمكن أن تتم الدراسة على أساس الحصر الشامل ولكن لابد وأن تتم الدراسة بأسلوب المعاينة.

٤. أيضاً هناك بعض الاختبارات لابد وأن تتم بأسلوب المعاينة لأن إجراء مثل هذه الاختبارات على أساس الحصر الشامل يؤدي إلى تلف المادة المختبرة أو هلاكها.. فاختبار صلاحية شحنه من المفرقعات مثلاً لابد وأن يتم على أساس العينة وبالمثل تحليل دم المرضى يتم على أساس عينه.

أقسام العينات: تنقسم العينات عادة إلى قسمين رئيسين وهما عينات عشوائية وعينات غير عشوائية، وفيما يلي تفصيل لكل قسم منها:

1- العينات العشوائية:

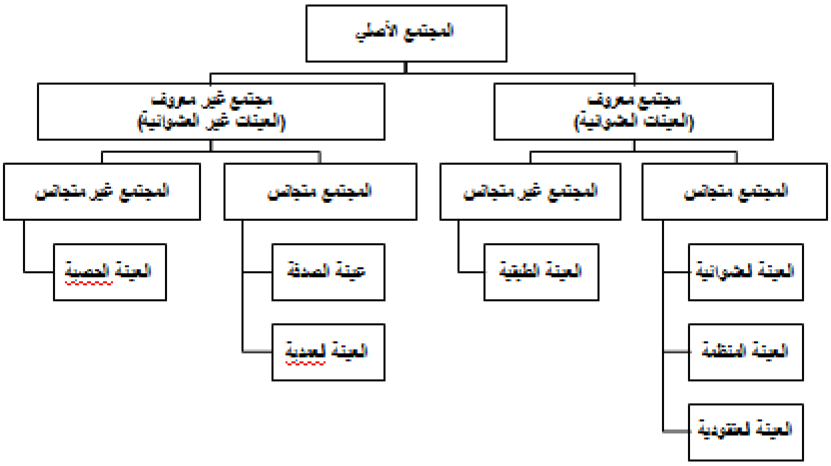
وهي تلك العينات التي يتم اختيار مفرداتها حسب خطه إحصائية لا يكون فيها للباحث أو لمفردات العينة دخل في اختيار أي مفرده فيها، حيث يتم الاختيار باستخدام أساليب معينة تلعب الصدفة خلالها الدور الأول في اختيار المفردة ولكن بشرط أن يتحقق لجميع المفردات احتمال ثابت ومحدد للاختيار. والعينات العشوائية إذا ما تم اختيارها بالطريقة العلمية السليمة والمناسبة يمكن أن تكفل درجه عالية من دقة التمثيل للمجتمعات المسحوبة منها لذلك فهي الوسيلة الأساسية في حالة البحوث العلمية الدقيقة.

2- العينات غير العشوائية:

وهي تلك العينات التي لا تكفل لجميع مفردات المجتمع احتمال ثابت ومحدد للاختيار، وغالباً يتدخل الباحث في عملية الاختيار بصورة أو بأخرى.

وسيتم فيما يلي استعراض لأهم أنواع العينات العشوائية والعينات غير العشوائية

أقسام العينات:



أ - العينات الاحتمالية:

العينة العشوائية	جميع عناصر المجتمع لها نفس الفرصة في الظهور في العينة
العينة الطبقية	يقسم المجتمع إلى طبقتين على الأقل ثم نختار العينة من كل منهما
العينة المنتظمة	نختار نقطة بداية من المجتمع ثم نختار العنصر الموجود على بعد ثابت من هذه النقطة
العينة العنقودية	يقسم المجتمع إلى مساحات أو أجزاء ثم نختار عشوائيا بعض هذه المساحات، ثم نختار جميع عناصرها بالعينة

ب - العينات غير الاحتمالية:

عينة الصدفة	يتم اختيارها عن طريق الصدفة
العينة العمدية (القصدية)	يتم اختيار أفراد العينة تحت شروط معينة لتحقيق الهدف من التجربة
العينة الحصية	يقسم المجتمع إلى أجزاء ثم نختار العينة من كل جزء من أجزاء المجتمع وفقا للنسب المحددة

أخطاء البيانات الإحصائية:

تتعرض البيانات الإحصائية التي يتم جمعها إلى نوعين من الأخطاء:

1- خطأ التميز أو التحيز: وهو ذلك الخطأ الناتج عن مصادر متعددة، منها أخطاء في تصميم البحث أو التجربة أو أخطاء فنيه أثناء جمع البيانات أو خلال العمليات الحسابية التي تتم على البيانات المتجمعة.. أخطاء التميز تزداد بازدياد الفروق بين الإمكانيات (المادية والفنية) اللازم توافرها لضمان أقصى درجة دقة ممكنه وبين الإمكانيات الفعلية المتاحة للباحث.

2- خطأ المعاينة العشوائية أو خطأ الصدفة: وهو الخطأ الناتج عن فروق الصدفة بين مفردات المجتمع التي دخلت العينة وبين تلك المفردات التي لم تشأ الصدفة أن تدخل العينة.

وفيما يلي شرح لهذين الخطأين:

1- خطأ التميز أو التحيز:

اذا سحبنا عدة عينات من مجتمع ما وحسبنا المتوسط الحسابي لكل عينة من هذه العينات ثم حسبنا المتوسط الحسابي لهذه المتوسطات فهذا المتوسط يجب أن يساوي المتوسط الحسابي الحقيقي للمجتمع المسحوبة منه هذه العينات، وفي حال وجود فرق بين المتوسطين فإن هذا الفرق يسمى **بخطأ التميز أو التحيز**

أسباب خطأ التميز أوالتحيز:

- الاختيار غير العشوائي للعينة: تعتمد بعض طرق الاختيار للعينة على خاصية معينة كالاعتماد على دليل الهاتف (عند دراسة الدخل والانفاق).
- التحيز المقصود (تعمد إدخال بعض الوحدات)
- استبدال وحدة بوحدة أخرى غير مدرجة ضمن الإطار العام للدراسة

كيفية التقليل من أخطاء التميز أوالتحيز:

- اختيار جميع وحدات العينة عشوائياً باستخدام إحدى طرق الاختيار العشوائي
- عدم استبدال أية وحدة تم اختيارها بوحدة أخرى
- تدريب الباحثين بشكل جيد على جمع البيانات والتقيد بالتعليمات

2- خطأ المعاينة العشوائية Random Sampling Error

عند اختيار العينة العشوائية هناك خطأ ينتج عن الاختلاف أو التشتت Variation بين قيم الوحدات التي تتكون منها العينة وتلك الوحدات التي لم تشأ الصدفة أن تدخلها في العينة وهذا الخطأ يسمى **بخطأ المعاينة العشوائي**

كيف نقلل من خطأ المعاينة العشوائي:

- زيادة حجم العينة
- طريقة الاختيار المناسب التي تقلل من اختلاف قيم الوحدات الإحصائية (كالأسلوب الطبقي أو العينة المنتظمة...الخ)

المحاضرة 7

توزيعات المعاينة Sampling Distribution's (الجزء الثاني)

تقويم عيِّنة الدراسة:

على الباحث كما أوضحنا في المحاضرة السابقة أن يتنبَّه إلى مواقع الخطأ في اختيار عيِّنة دراسته، والتي من أبرزها الآتـي:

١. **أخطاء التحيُّز:** وهي أخطاءٌ تحدث نتيجة للطريقة التي يختار بها الباحثُ عيِّنة دراسته من مجتمعها الأصليِّ.
٢. **أخطاء الصدفة:** وهي أخطاءٌ تنتج عن حجم العيِّنة فلا تمثِّل المجتمع الأصليَّ نتيجةً لعدم إعادة استبانات الدراسة أو عدم إكمال الملاحظة أو المقابلة لمفردات مجتمع الدراسة.
٣. **أخطاء الأداة:** وهي أخطاء تنتج من ردود فعل المبحوثين نحو أداة أو وسيلة القياس.

- ويمكن تلافي هذه العيوب بالتدرُّب الذاتي المكثِّف للباحث ليتقنَ أسلوب الدراسة بالعيِّنة وكيفيَّة اختيارها وتطبيقها بما تحقِّق تمثيلاً مناسباً لمجتمع دراسته، وأن يقوم بتدريب المتعاونين معه تدريباً يحقِّق له ذلك

المعالم والإحصاءات:

اعتاد البعض على معاملة القيم التي يحصل عليها من العينة وكأنها قيم مجتمعها، وهذا خطأ فادح. فلكي يستدل على خصائص مجتمع الدراسة تعتمد معادلات عديدة، ومتنوعة حسب نوع العينة.

فالمقاييس الإحصائية التي تحسب من بيانات مجتمع الدراسة بأكمله يطلق عليها معالم المجتمع (Parameters of population)

أما المقاييس الإحصائية التي تحسب من بيانات عينه مسحوبة من مجتمع الدراسة فيطلق عليها إحصاءات (Statistics)

وللتفرقة بين المعالم والإحصاءات يجب أن نرمز لكل منها برموز تختلف عن رموز الأخرى، على سبيل المثال يرمز للمتوسط الحسابي للمجتمع بالرمز μ بينما يرمز للمتوسط الحسابي للعينة بالرمز $\bar{x}$ ، أيضاً للانحراف المعياري للمجتمع بالرمز σ بينما يرمز للانحراف المعياري للعينة بالرمز S وهكذا.

ويعتبر كل إحصاء بمثابة تقدير أو قيمه تقديرية لمعلمة المجتمع المناظر، فيكون المتوسط الحسابي المحسوب من بيانات العينة تقدير لمعلمة المجتمع المناظرة وهي المتوسط الحسابي المسحوب منه هذه العينة وهكذا..

إن حساب قيمة المتوسط الحسابي من بيانات العينة ليس هدفاً في حد ذاته ولكن وسيلة للتعرف على المتوسط الحسابي للمجتمع موضوع الدراسة.. وهكذا الحال بالنسبة لباقي المقاييس الإحصائية التي تحسب من العينة.

الاستدلال على خصائص مجتمع الدراسة

المتوسط والانحراف المعياري:

إن الهدف من اخذ العينة هو معرفة خصائص مجتمعها، ونادرا ما يكون تمثيل العينة لمجتمعها دقيقا. فعند اخذ مائة قيمة من جداول الأرقام العشوائية وترتيبها بعشرة أعمدة ومثلها اسطر وحساب متوسط قيم كل سطر ومتوسط قيم كل عمود، وحساب قيمة الانحراف المعياري للقيم عن معدلاتها وجد أن قيم متوسطات الأسطر قد تراوحت بين (41.6) و (59.6) وبين (16.7) و (33.7) لقيم الانحراف المعياري لها. أما متوسطات قيم الأعمدة فقد تراوحت بين (43.6) و (66.1) وقيم انحراف معياري بين (19.7) و (32.4) ، في وقت كان متوسط جميع القيم (المتوسط الحقيقي) يساوي (50.61) وبانحراف معياري قدره (26.055)

وبقيام خمس طلبة بدراسة هذا المجتمع مع اختلاف في حجم العينة، حيث اختاروا (10، 20، 30، 40، و 50) على التوالي ، تبين أن المعدلات كانت (66.1، 49.1، 55.5، 47.3 و48.1) على التوالي، وكانت قيم الانحراف المعياري لها (24.8، 26.8، 26.6، 26.1 و 25.6) على التوالي، وهذا يعني إن أي منهم لم يمثل معدل مجتمع الدراسة الحقيقي، بل بقيت التقديرات تدور حوله

ويشير المثال السابق إلى أن العينة لا تعرض الواقع بل تقترب منه، وبعبارة أدق، إن كل عينة لم تؤخذ بموضوعية ولم تحسب احتمالات تمثيلها لمجتمعها بدقة لا يمكن الركون إليها لأنها تمثل نفسها دون مجتمعها. فاخذ العينات ليس القصد منه العينة لذاتها بل المجتمع الذي أخذت منه ، **فالعينة وسيلة وليست الهدف**.

ويأتي هنا دور نظرية العينات لتساعد في تقدير قيم خصائص مجتمع الدراسة وضمن مجال محدد للخطأ، وحسب تنظير الحد المركزي Central Limit Theorem، فانه عند اخذ عينات بحجم كبير من أي مجتمع فان معدلات العينات ستتوزع بصورة طبيعية Normal، وان متوسط متوسطات العينات سيقترب من متوسط مجتمع الدراسة

وتقدم العينات تقديرات لخصائص مجتمعها، وهذه التقديرات تدور حول المتوسط الحقيقي لمجتمع الدراسة. أي أن متوسط العينة هو ليس متوسط مجتمعها، وليس تقديرا له، **بل قيمة تمثل العينة ذاتها** ، وتعتمد في تقدير القيمة المحتملة لمتوسط المجتمع وفق حدود معينة للثقة.

ولهذا من الضروري حساب متوسط قيم العينة ودرجة التباين Variance فيها وفق المعادلة:

التباين = مجموع (تربيع الفرق بين قيم العينة عن متوسطها) / حجم العينة - 1

$$S^2 = \frac{\sum (x-\mu)^2}{n-1}$$

وباشتقاق الجذر التربيعي لقيم التباين نحصل على قيمة الانحراف المعياري: $S = \sqrt{\frac{\sum (x-\mu)^2}{n-1}}$

إن ارتفاع قيمة الانحراف المعياري يدلل على التباين الكبير بين قيم العينة، ولهذا أهمية خاصة عند تحديد حجم العينة وعند تحديد درجة الثقة بالتقديرات التي توفرها عن مجتمعها. فكلما كان التباين كبيرا في خصائص المجتمع كانت معدلات العينة بعيدة عن معدل مجتمعها، ولهذا فان قيمة الخطأ المعياري SE Standard Error of Sample Standard Deviation في قيمة الانحراف المعياري للعينة ستكون كبيرة.

لهذا فان اخذ العينات بعدد قليل قد لا يعكس الصورة الصحيحة للتباين في خصائص مجتمع الدراسة، وقد تتشابه قيم العينات عن طريق الصدفة، ولهذا يفضل اعتماد عدد (حجم) كبير للعينات.

الخطأ المعياري:

بمعرفة قيمة الانحراف المعياري لقيم العينة يمكن تقدير قيمة الخطأ المعياري في الانحراف المعياري للعينة باعتماد المعادلة الآتية: $SE = \left[\frac{S}{\sqrt{n}}\right]\left[\sqrt{1-(\frac{n}{N})}\right]$ حيث ان: (N) = حجم مجتمع العينة و (n) = حجم العينة

وفق هذه المعادلة تؤخذ نسبة العينة إلى مجتمعها، وكلما كبرت هذه النسبة تحسن تمثيلها لمجتمع الدراسة، أما عندما يكون حجم مجتمع العينة مجهولا، وهي حالة أكثر شيوعا في الدراسات الجغرافية، حينها تعتمد المعادلة الآتية: $SE = \frac{S}{\sqrt{n}}$

الخطأ المعياري = الانحراف المعياري لقيم العينة / جذر حجم العينة

أما عندما يكون حجم العينة أكثر من (100) فتعتمد المعادلة أدناه: $SE = \frac{S}{\sqrt{n^2}}$

الخطأ المعياري = الانحراف المعياري لقيم العينة / جذر (حجم العينة)^2

ويقدم كريكوري تعديل بيسل عندما يكون حجم العينة اقل من (30): $SE = [S]\left[\sqrt{\frac{n}{(n-1)}}\right]$

ويتمثل هذا التعديل بأخذ الجذر التربيعي لحاصل قسمة حجم العينة على (حجم العينة - 1) والهدف من ذلك الحصول على أفضل تقدير للخطأ المعياري . بعبارة أخرى ، تستخدم القيمة المشتقة عن معادلة بيسل كبديل لقيمة الخطأ المعياري في المعادلات الأخرى عندما يكون حجم العينة صغيراً

إن الانحراف المعياري للتوزيع النظري لمتوسطات العينات يقيس خطأ المعاينة ويسمى بالخطأ المعياري للمتوسط، ومن الضروري التذكر دوما أن متوسط المجتمع قيمة محددة تقع ضمن مجال محدد Certain Interval ، والباحث غير متأكد من قيمتها، ولكنه يحسب احتمالية وجودها ضمن المجال المحدد وبمستوى ثقة إحصائية معلوم

مستوى الثقة وحدودها:

إذا أخذت جميع العينات المحتملة من مجتمعها فيتوقع أن تكون متوسطات العينات موزعة بالتساوي حول متوسط مجتمع الدراسة. **بعبارة أخرى، إن متوسط متوسطات العينات يساوي متوسط مجتمعها.**

وتتوزع متوسطات العينات دائما بصورة متماثلة Normal Distribution ، والذي يمتاز رياضيا بالابتعاد بنسب ثابتة عن المتوسط مع كل درجة معيارية، وبالتالي تباينت متوسطات العينات المأخوذة منه فانه يتوقع أن يقع متوسطه وباحتمالية قدرها كالتالي:

- مستوى ثقة إحصائية قدره (68.26%) أو باحتمالية قدرها (0.6827) يقع متوسط مجتمع الدراسة بين قيمة متوسط متوسطات العينات و (+ و –) درجة واحدة من الخطأ المعياري .
- مستوى ثقة إحصائية قدرها (95%)، أو باحتمالية (0.95) يقع متوسط مجتمع الدراسة بين متوسط متوسطات العينات و (+ و –) درجتان من الخطأ المعياري تقريبا.
- مستوى ثقة إحصائية قدرها (99%) أو باحتمالية قدرها (0.99) يقع متوسط مجتمع الدراسة بين قيمة متوسط متوسطات العينات و (+ و –) ثلاث درجات من قيمة الخطأ المعياري تقريبا.

وتسمى هذه بمستويات الثقة Confidence Level و يعبر عنها بإشارة النسبة المئوية (%) بان تكون التقديرات صحيحة أو باحتمالية (0.01) أو (0.05) أن تكون خاطئة.

وتحسب حدود الثقة لاعتمادها في تقدير متوسط مجتمع الدراسة كما يأتي: $\hat{\mu} = \overline{X} \pm Z\sigma\overline{x}$

وتعتمد قيمتي الخطأ المعياري و حدود الثقة على درجة التباين في خصائص المجتمع قيد الدراسة، فقد يوفر عدد قليل من العينات تقديرات جيدة لخصائص المجتمع عندما يكون هذا متجانسا نسبيا، والعكس صحيح أيضا. أي أن تكون التقديرات غير واقعية عندما تكون التباينات كبيرة في خصائص المجتمع وحجم العينة صغيرا.

مثال:

قام أحد الباحثين في مجال الزراعة بدراسة مائة مزرعة، فوجد أن متوسط مساحة المزرعة الواحدة (53) هكتارا، وبانحراف معياري عن المتوسط بقيمة (26) هكتارا

أحسب حدود الثقة في تقدير متوسط مساحة المزرعة في منطقة الدراسة؟

الحل:

$$\hat{\mu} = \overline{X} \pm Z\sigma\overline{x}$$
$$= 53 \pm (1.96)\frac{26}{\sqrt{100}}$$
$$= 53 \pm 5.1$$

عندها يمكن القول بأنه وبثقة إحصائية قدرها (95%) إن معدل مساحة المزرعة في منطقة الدراسة يقع بين (53 + 5.1 = 58.1) هكتار و (53 – 5.1 = 47.9) هكتار . أما إذا أراد الباحث أن يكون مستوى الثقة الإحصائية (99%) حينها يستبدل (1.96) بـ (2.58) في المعادلة أعلاه ليكون المعدل المتوقع بين (46.3) و (59.7) هكتار.

حجم العينة:

إن لحجم العينة أهمية كبيرة في تحديد الثقة بالنتائج، لذا من الضروري أن يسلط الضوء عليه بشيء من التفصيل وحسب التوزيعات المعروفة للقيم، وسيتم هنا تناول نوعين من التوزيعات وهي:

١. التوزيع الطبيعي للقيم
٢. توزيع (ت) للقيم

(أ) التوزيع الطبيعي للقيم:

كلما كبر حجم العينة ازدادت دقة تمثيلها لمجتمعها واقترب توزيع القيم فيها من التوزيع الطبيعي (المتماثل الجانبين) وأصبحت عملية الاستدلال أكثر دقة. وللتوضيح نورد مثالا، إذا أريد معرفة نسبة طلبة قسم الإدارة إلى مجموع طلبة الكلية فان عينة من عشرة طلبة قد لا تفي بالغرض، ولكن عينة من مائة طالب تفي بالغرض حتما. بعبارة أخرى، إن حجم العينة أساسي لإعطاء صورة عن مجتمع الدراسة وليس النسبة المئوية للعينة قياسا بحجم مجتمعها. فكلما ازداد حجم العينة ازدادت الثقة بتقديرات خصائص المجتمع وصغرت معه حدود الثقة.

وقد لا يعرف الباحث الكثير عن مجتمع الدراسة، فهو بحاجة إلى بعض المؤشرات عن تباين خصائص المجتمع قبل أن يحدد حجم العينة. يتطلب منه هذا القيام بمسح تجريبي لاستخراج قيمة المعدل والانحراف المعياري للعينة، وبتطبيق المعادلة المذكورة في أدناه يستطيع تحديد حجم العينة المناسبة لدراسته: $n = \frac{Z^2\sigma^2}{e^2}$

- Z = هو معامل الثقة (أو الدرجة المعيارية) المقابل لدرجة الثقة المطلوبة، ونحصل عليا من جدول التوزيع الطبيعي المعياري.
- $\{\sigma^2\}$ = هو تباين المجتمع (أو هو مربع الانحراف المعياري).
- e = هو أقصى خطأ مسموح به في تقدير الوسط، وهو عادة ما يحدده الباحث، وتتوقف قيمته على أهمية الموضوع أو الظاهرة السياسية المراد دراستها، ومدى الدقة المطلوبة في التقدير، ويسمى اختصاراً "الخطأ في تقدير الوسط".

أما إذا كان اهتمام الباحث منصبا على نسبة الخصائص وليس قيمتها، لذا لا يمكن اشتقاق قيمة الانحراف المعياري، بل يستعاض عنه بقيمة التباين في النسب، وتحسب كالتالي:

$$S^2 = \sqrt{(P(100-P))}$$

حيث تمثل (P) النسبة المئوية للخاصية موضع الدراسة

ولهذا تكون معادلة حجم العينة: $n = \left[\frac{(Z)(S^2)}{e}\right]^2$

فإذا أراد باحث دراسة ملكية السيارات في مدينة ما، واختار (2%) حدا للثقة وثقة إحصائية قدرها (95%)، ويتوقع أن يمتلك نصف السكان وسائط نقل خاصة حينها :

$$S^2 = \sqrt{(50(100-50))} = 50$$

$$n = \left[\frac{(1.96)(50)}{.02}\right]^2 = 24.01$$

أي اننا بحاجة إلى عينة بحجم 24 لضمان الدقة المرجوة في تمثيل خصائص المجتمع، وقد يتم البحث في مواضيع ذات مجتمعات صغيرة الحجم وبهذا تشكل النسبة المئوية حجما كبيرا فتكون النتائج مضللة، في بعض الأحيان.

(ب) توزيع (ت) للقيم :

من الضروري اخذ الحذر عندما يكون حجم العينة صغيرا اقل من (30) وذلك لأنها تتطلب إجراءات خاصة عند التحليل. فعندما يكون حجم العينة أكثر من (30) يتجه توزيع قيمها نحو التوزيع الطبيعي وبغض النظر عن التوزيع الحقيقي لقيم مجتمع الدراسة.

وبالنسبة للعينات الصغيرة الحجم فان توزيع قيمها يتأثر بطبيعة توزيع قيم المجتمع المأخوذة منه. وعندما يكون توزيع قيم المجتمع معروفا أو متوقعا أن لا يكون طبيعي حينها يجب اعتماد حجم كبير للعينة . أما إذا كان توزيع قيم المجتمع طبيعيا عندها يمكن اخذ عينات بحجم صغير ويعتمد توزيع (ت) (T) في التحليل و المقارنة.

يتشابه توزيع قيم (ت) مع شكل الجرس بزيادة الحجم حتى يتطابق معه عندما يتعدى العدد (30) فشكل توزيع (ت) للقيم لا يختلف كثيرا عن التوزيع الطبيعي إلا في الأعداد القليلة، وكلاهما متماثل النصفين لذا يعتمد كبديل له في القيم القليلة العدد.

ولتوزيع (ت) جداول للقيم الحرجة منظمة على شكل اسطر اعتمادا على درجة الحرية التي تقاس بـ (حجم العينة - 1). أما الأعمدة فتمثل درجة الاحتمالية Probability، وتتناقص القيم الحرجة بتزايد درجة الحرية (حجم العينة). ودرجة الحرية تفضل على حجم العينة في الأحجام الصغيرة للعينة لأنها تقلل من الانحياز في تقدير خصائص مجتمع الدراسة.

ليس هناك اختلاف في حساب قيم المعدل والانحراف المعياري بين العينات الكبيرة الحجم أو الصغيرة، يبدأ الاختلاف مع تحديد حدود الثقة حيث تستبدل قيمة (Z) لتحل مكانها قيمة (T).

إن ارتفاع القيم الجدولية في توزيعات (ت) قياسا بنظائرها في توزيع (Z) تعكس الدرجة العالية لعدم ضمان تمثيل العينات الصغيرة لخصائص مجتمعها.

العوامل المحددة لحجم العينة:

درجة التباين في خصائص مجتمع الدراسة: يلعب التباين في خصائص مجتمع الدراسة دورا مهما في تحديد درجة دقة نتائج العينة، فكلما كان التباين كبيرا تطلب الأمر زيادة حجم العينة ليكون تمثيلها للتباين في المجتمع صحيحا.

طريقة التحليل المعتمدة: عند إقرار حجم العينة، من الضروري تحديد الحجم الأصغر المقبول للعينة في المجاميع الثانوية ضمن مجتمع الدراسة ، إذ أن بعض الاختبارات الإحصائية تتطلب عددا معينا كحد أدنى لكل فئة أو صنف لتكون النتائج ذات معنى.

حجم المعلومات المطلوبة: فكلما كانت المعلومات المطلوبة من العينة (الواحدة) كثيرة وتفصيلية كان حجم العينة صغيرا، ما لم يكن المشروع البحثي كبيرا وتتوفر له المصادر البشرية والمادية اللازمة. إن الدقة في المعلومات المطلوبة من العينة أهم بكثير من حجم العينة ، فحجم العينة لا يتحدد بحجم مجتمع الدراسة فقط، بل وبالدقة المتوخاة والتفاصيل المطلوبة.

المصادر المالية والبشرية المتوفرة: تتطلب الدراسة الميدانية توفر مصادر مالية وبشرية لتغطية تكاليفها التي تكون في الغالب باهظة لتأثيراتها على تحديد حجم منطقة الدراسة، مجتمع الدراسة وبالتالي حجم العينة. إن مضاعفة حجم العينة يتطلب زيادة في كمية المصادر المالية والجهد البشري.

حدود الثقة في تقديرات خصائص مجتمع الدراسة: لزيادة الدقة في النتائج يعمد البعض إلى تقليص حدود الثقة (المدى الذي يفترض أن يقع ضمنه المعدل المتوقع للمجتمع). إن إنقاص حدود الثقة من (6%) إلى (4%) يتطلب زيادة حجم العينة بنسبة (225%)، وكلما كان المدى كبيرا كان حجم العينة صغيرا، والعكس صحيح.

حالات الإخفاق وعدم الاستجابة: العامل الآخر الذي يحدد حجم العينة هي حالات الإخفاق في الحصول على المعلومات وعدم الاستجابة أو المعلومات غير الوافية.

توزيع المعاينة

وهو ذلك التوزيع التكراري لأحد التوابع الإحصائية المحسوب من بيانات العينات العشوائية ذات الحجم الواحد والتي يمكن سحبها من مجتمع إحصائي واحد.

نفرض أننا أخذنا عينه حجمها n من مجتمع ما، ثم سحبنا منها بعض المقاييس الإحصائية مثل المتوسط الحسابي، التباين، ... فإن كل مقياس من هذه المقاييس يعتبر متغير عشوائي في ذاته يختلف من عينه إلى أخرى – هذا المتغير العشوائي يخضع لتوزيع معين – هذا التوزيع يسمى بتوزيع العينة

فمثلاً نقول إن توزيع المعاينة للمتوسط الحسابي وهو عبارة عن توزيع جميع المتوسطات الحسابية للعينات المأخوذة من نفس هذا المجتمع ذو الحجم n ، وكذلك فإن توزيع المعاينة للتباين هو توزيع جميع التباينات المحسوبة من عينات لها نفس الحجم n ومأخوذة من نفس المجتمع ، وهكذا ...

أ- توزيع المعاينة للمتوسط:

إذا أخذنا عينات متكررة من مجتمع ما وقمنا بحساب متوسط كل عينة، فإننا نجد أن معظم هذه المتوسطات تختلف عن بعضها البعض، ويسمى التوزيع الاحتمالي لمتوسطات العينات **"توزيع المعاينة للوسط"**

مثال:

إذا أخذنا عينات متكررة حجمها n من مجتمع ما فإن لها متوسط وانحراف معياري كالتالي:

$$\mu_{\overline{X}} = \mu$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

- إذا كان الإحصائي المستخدم هو الوسط الحسابي، فان توزيع المعاينة هو للوسط الحسابي ويسمى توزيع المعاينة للوسط الحسابي.
- إذا كان الإحصائي المستخدم هو الوسيط، فان توزيع المعاينة هو للوسيط ويسمى توزيع المعاينة للوسيط.
- في توزيع العينة أو المعاينة، الوسط الحسابي (الإحصائي) يتطابق مع قيمة المعلمة، وهذا يعني أن الأخطاء تم تحييدها.

١. الانحراف المعياري لتوزيع العينة أو المعاينة يسمى **الخطأ المعياري للإحصائي** ويشير إلى درجة خطأ التقديرات، ويختلف عن الانحراف المعياري للمجتمع.
٢. إذا كان الخطأ المعياري للوسط الحسابي، نرمز له σ_x.
٣. إذا كان الخطأ المعياري للوسيط، نرمز له σ_{Mdn}.
٤. لكي تمثل العينة المجتمع تمثيلاً صادقاً، سواء كان السحب بدون إرجاع أو بالإرجاع، يجب أن نراعي الآتي:
   ١. الوسط الحسابي للعينة (الإحصائي) = الوسط الحسابي للمجتمع (المعلمة)
   $X = \mu$
   ٢. الانحراف المعياري للعينة = الخطأ المعياري للمجتمع
   $S_x=\sigma_x$

مثال:

إذا كان لدينا مجتمع إحصائي مؤلف من العاملين في منشأة صغيرة حجمه N=4 مفردات، ومكون من القيم (عدد ابناء العامل): {0,2,4,6}

$$\mu = \frac{0+2+4+6}{4} = 3$$

$$\sigma^2 = \frac{(0-3)^2+(2-3)^2+(4-3)^2+(6-3)^2}{4} = 5$$

نفرض أننا سحبنا جميع العينات الممكنة مع الإعادة ذات الحجم n=2 ثم حسبنا متوسطاتها

عدد العينات الممكن سحبها مع الإعادة يعطى بالعلاقة: $N^n = 4^2 = 16$

وإن متوسطات العينات العشوائية المسحوبة تتأرجح بين (0 ، 6) أنظر الجدول التالي:

رقم العينة	العينة		المتوسط	رقم العينة	العينة		المتوسط
1	0	0	0	9	4	0	2
2	0	2	1	10	4	2	3
3	0	4	2	11	4	4	4
4	0	6	3	12	4	6	5
5	2	0	1	13	6	0	3
6	2	2	2	14	6	2	4
7	2	4	3	15	6	4	5
8	2	6	4	16	6	6	6

وإن الجدول الاحتمالي لتوزيع معاينة الأوساط الحسابية $\overline{X}_i$

المتوسط	0	1	2	3	4	5	6
(P)	1/16	2/16	3/16	4/16	3/16	2/16	1/16

ولو رسمنا المدرج التكراري، نلاحظ أن توزيع المعاينة للأوساط الحسابية للعينات يمكن أن يقترب وبشكل جيد من منحني التوزيع الطبيعي والذي يرمز لمتوسطه بـ μ_x وتشتته بـ :

$$\mu_{\bar{x}} = \sum \overline{X}_i P(\overline{X}_i) = 0.\left(\tfrac{1}{16}\right)+1.\left(\tfrac{2}{16}\right)+2.\left(\tfrac{3}{16}\right)+3.\left(\tfrac{4}{16}\right)+4.\left(\tfrac{3}{16}\right)+5.\left(\tfrac{2}{16}\right)+6.\left(\tfrac{1}{16}\right)=3$$

وهي نفس قيمة μ

إذاً: $\mu_{\overline{X}} = \mu$

	مع الإرجاع	بدون ارجاع
العينة	$\overline{X}=\frac{\sum x}{n}$	$\overline{X}=\frac{\sum x}{n}$
	$\overline{X}=\frac{\sum fx}{n}$	$\overline{X}=\frac{\sum fx}{n}$
	$S=\sqrt{\frac{\sum\left(x-\overline{X}\right)^2}{n}}$	$S=\sqrt{\frac{\sum\left(x-\overline{X}\right)^2}{n}}$
	$S=\sqrt{\frac{\sum f\left(x-\overline{X}\right)^2}{n}}$	$S=\sqrt{\frac{\sum f\left(x-\overline{X}\right)^2}{n}}$
المجتمع	$\mu=\frac{\sum x}{n}$	$\mu=\frac{\sum x}{n}$
	$\sigma=\sqrt{\frac{\sum(x-\mu)^2}{n}}$	$\sigma=\sqrt{\frac{\sum(x-\mu)^2}{n}}$

مثال:

يتكون مجتمع إحصائي من 4 أشخاص، وتم قياس ميزة معينة لهم وحصلنا على القيم التالية:

1 2 3 4

نود اختيار عينات من هذا المجتمع بدون إرجاع وبالإرجاع، مع العلم بان حجم العينة =2

الحل:

السحب بدون إرجاع: "العينة"

١. في العمود الأول من الجدول (1)، نسحب العينات، وفي العمود الثاني نحسب الوسط الحسابي لكل عينة.
٢. نقوم بتكوين جدول (2) يحتوي على البيانات التالية: العمود الأول ونمثل فيه البيانات (X) وتعني متوسط العينات التي تم سحبها، والعمود الثاني يمثل التكرار للمتوسطات (f)

بعد ذلك نحسب الوسط الحسابي والانحراف المعياري لهذه البيانات

جدول (1)

العينة	الوسط الحسابي X
(1,2)	1.5
(1,3)	2
(1,4)	2.5
(2,3)	2.5
(2,4)	3
(3,4)	3.5

جدول (1)

X	f	f.X	$\overline{X}$	(X - $\overline{X}$)	(X - $\overline{X}$)^2	f.(X - $\overline{X}$)^2
3.5	1	3.5	2.5	1	1	1
3	1	3	2.5	0.5	0.25	0.25
2.5	2	5	2.5	0	0	0
2	1	2	2.5	-0.5	0.25	0.25
1.5	1	1.5	2.5	-1	1	1
	6	15	2.5			2.5

$$\overline{X} = \frac{\sum X}{n} = \frac{15}{6} = 2.5$$

$$S = \sqrt{\frac{\sum f\left(X-\overline{X}\right)^2}{N}} = \sqrt{\frac{2.5}{6}} = 0.645$$

<u>السحب بدون إرجاع: "المجتمع"</u>

نكون جدول للبيانات من النوع الأول ونحسب الوسط الحسابي والانحراف المعياري.

X	(X-μ)	(X-μ)^2
4	1.5	2.25
3	0.5	0.25
2	-0.5	0.25
1	-1.5	2.25
	0	5

$$\mu = \frac{\sum X}{n} = \frac{10}{4} = 2.5$$

$$\sigma = \sqrt{\frac{\sum (X-\mu)^2}{N}} = \sqrt{\frac{5}{4}} = 1.1180$$

<u>السحب بالإرجاع: "العينة"</u>

١. في العمود الأول من الجدول (1)، نسحب العينات، وفي العمود الثاني نحسب الوسط الحسابي لكل عينة.
٢. نقوم بتكوين جدول (2) يحتوي على بيانات من النوع الثاني، ونمثل فيه البيانات (X) و تكرارها (f).

بعد ذلك نحسب الوسط الحسابي والانحراف المعياري لهذه البيانات

العينة	الوسط الحسابي X
(1,1)	1
(1,2)	1.5
(1,3)	2
(1,4)	2.5
(2,1)	1.5
(2,2)	2
(2,3)	2.5
(2,4)	3
(3,1)	2
(3,2)	2.5
(3,3)	3
(3,4)	3.5
(4,1)	3.5
(4,2)	3
(4,3)	3.5
(4,4)	4

X	f	f.X	$(X - \overline{X})$	$(X - \overline{X})^2$	$f.(X - \overline{X})^2$
4	1	4	1.5	2.25	2.25
3.5	2	7	1	1	2
3	3	9	0.5	0.25	0.75
2.5	4	10	0	0	0
2	3	6	-.05	0.25	0.75
1.5	2	3	-1	1	2
1	1	1	-1.5	2.25	2.25
	16	40			10

$$\overline{X} = \frac{\sum f.X}{n} = \frac{40}{16} = 2.5$$

$$S = \sqrt{\frac{\sum f\left(X - \overline{X}\right)^2}{N}} = \sqrt{\frac{10}{16}} = 0.7905$$

<u>السحب بالإرجاع: "المجتمع"</u>

نكون جدول لبيانات من النوع الأول ونحسب الوسط الحسابي والانحراف المعياري

X	$(X-\mu)$	$(X-\mu)^2$
4	1.5	2.25
3	0.5	0.25
2	-0.5	0.25
1	-1.5	2.25
	0	5

$$\mu = \frac{\sum X}{n} = \frac{10}{4} = 2.5$$

$$\sigma = \sqrt{\frac{\sum (X-\mu)^2}{N}} = \sqrt{\frac{5}{4}} = 1.1180$$

وبالمقارنة بين النتائج نجد أن:

الوسط الحسابي للعينة $\bar{x}$ 2.5 = الوسط الحسابي للمجتمع μ 2.5

المحاضرة 8

التقدير الإحصائي (الجزء الأول)

التقدير:

التقدير هو عملية استنتاج أو تقدير أحد معالم المجتمع (مثل الوسط الحسابي أو الانحراف المعياري) بناءً على بيانات عينة مسحوبة من المجتمع.

وهناك نوعان (أو أسلوبان) للتقدير:

- الأول تقدير النقطة (أو القيمة الواحدة)
- الثاني تقدير الفترة (أو فترة التقدير أو الثقة)

1- التقدير بنقطة

التقدير بنقطة هو عبارة عن عدد واحد، ويكون هذا التقدير بنقطة **غير متحيز** إذا كانت القيمة المتوقعة أو القيمة الوسطى للإحصاء المناظر، عند تكرار المعينة العشوائية، مساوية لمعلمة المجتمع . فمثلا $\bar{x}$ هي تقدير (بنقطة) غير متحيز للمعلمة μ لأن $\mu x = \mu$ حيث μx هي القيمة المتوقعة للمتوسط $\bar{x}$ أما الانحراف المعيارى **s** للعينــة فهو تقديــر غير متحيز للمعلمــة σ والنسبة فى العينة **p** هى تقدير غير متحيز للمعلمة **p** (وهى نسبة المفردات التى لها خاصية معينة فى المجتمع كله) .

فالتقدير بنقطة يعني بالتالي أن نحصل على قيمة واحدة من العينة، وتستخدم هذه القيمة الواحدة كتقريب أو كتقدير لمعلمة المجتمع المجهولة.

فمثلاً لو أخذنا الوسط الحسابي للدخل في العينة كتقدير لمتوسط الدولة نكون قد حصلنا على تقدير نقطة لمتوسط دخل الدولة.
وكمثال آخر لو أخذنا نسبة الناخبين في العينة الذين يؤيدون مرشحا معيناً كتقدير لهذه النسبة في المجتمع نكون حصلنا على تقدير نقطة للنسبة في مجتمع الناخبين

2- التقدير بفترة:

أما التقدير بفترة فنحصل من خلاله على مدى Range أو فترة تتحدد بحدين (حد أدنى وحد أعلى) - نحصل عليهما من العينة.
ونلاحظ هنا أن فترة التقدير (أو تقدير الفترة) تحتوي على أكثر من قيمة بل قد يكون عدد القيم غير محدود أو لا نهائياً في كثير من الحالات.

مثلاً:

إذا قدرنا أن الوسط الحسابي لأعمار الناخبين يتراوح بين (6 - 40) و (6 + 40) سنة أي يتراوح بين 34 سنة كحد أدنى و 46 سنة كحد أعلى نكون قد حصلنا على تقدير فترة للوسط الحسابي لأعمار الناخبين في المجتمع - ونلاحظ أن هذه الفترة (34, 46) تحتوي على عدد لا نهائي من الأعمار، بمعنى أن العدد لا يقتصر فقط على الأعداد الصحيحة والتي تشمل السنوات، ولكنها تشمل أيضا كسور السنوات، والأيام والشهور، والساعات.. الخ

وتتميز تقديرات الفترة بالإضافة إلى أنها تحتوي على عدد كبير جداً من القيم، بأنه يمكن حساب احتمال أن يكون التقدير صحيحاً، وبالتالي فإنه يمكن معرفة مدى دقة التقديرات. لذا فإن فترات التقدير تسمى أيضاً " فترات الثقة " Confidence intervals لأن هذه الفترات تعتمد في تكوينها الإحصائي على درجات أو مستويات ثقة معينة Confidence Levels مثل 95% أو 99 % وغيرها، بمعنى أن احتمال أن تكون فترة التقدير صحيحة هو 0.95 أو 0.99 وهكذا...

مثلا:

لو كان متوسط أعمار الناخبين يتراوح ما بين 46 و 34 سنة، ودرجة الثقة هي 95 % فإن هذا معناه أنه لو تكررت التجربة مائة مرة، فإن التقدير سيكون محصوراً بين هذين الرقمين في 95 مرة من الحالات (أي احتمال أن يكون صحيحاً هو 95%).

التقدير الإحصائي لفترة الثقة للوسط الحسابي لعينة إحصائية مأخوذة من توزيع طبيعي

إذا كان المجتمع المسحوب منه العينة ذو توزيعاً طبيعي وتباينه معروفاً أو كانت العينة كبيرة (أي حجمها ثلاثون مفردة أو أكثر) فإن توزيع المعاينة للوسط الحسابي يكون توزيعاً طبيعياً

وطالما نحن نتكلم عن تقدير متوسط المجتمع فإن أول ما نفكر فيه هو الوسط الحسابي للعينة، وفترة التقدير (أو الثقة) للوسط الحسابي للمجتمع تأخذ الشكل التالي :

تقدير متوسط المجتمع = الوسط الحسابي للعينة ± الخطأ المعياري للوسط

وبالرموز فإن: $\hat{\mu} = \overline{X} \pm 1\sigma^2 \overline{x}$

حيث أن:

- $\hat{\mu}$ = تقدير الوسط الحسابي للمجتمع،
- $\overline{X}$ = الوسط الحسابي للعينة،
- $\sigma^2 \overline{x}$ = الخطأ المعياري للوسط،
- (+) تشير للجمع فنحصل على الحد الأعلى لفترة التقدير،
- (-) تشير للطرح فنحصل على الحد الأدنى،

ولكن احتمال أن يكون هذا الكلام صحيحا هو % 68.26 فقط، أي أن درجة الثقة هنا لا تتعدى % 68.26

فإذا أضفنا وطرحنا ضعف الخطأ المعياري يرتفع الاحتمال إلى 95.44% أي ترتفع درجة الثقة إلى % 95.44 وفي هذه الحالة تأخذ فترة الثقة الشكل التالي:

$$\hat{\mu} = \overline{X} \pm 2\sigma^2 \overline{x}$$

وإذا أضفنا وطرحنا ثلاثة أمثال الخطأ المعياري يصبح الاحتمال % 99.72 أي ترتفع درجة الثقة إلى 99.72% وتأخذ فترة الثقة الشكل التالي:

$$\hat{\mu} = \overline{X} \pm 3\sigma^2 \overline{x}$$

أي أنه بزيادة درجة الثقة يزيد طول الفترة، ومما سبق نلاحظ ما يلي:

1- أن هناك علاقة وثيقة بين درجة الثقة والرقم أو "المعامل المضروب في الخطأ المعياري فهو إما 1 أو 2 أو 3 على حسب درجة الثقة :

معامل الضرب في الخطأ المعياري (معامل الثقة)	درجة الثقة
1	68.26%
2	95.44%
3	99.72 %

ولذلك فإن هذا المعامل هو الذي يسمى "معامل الثقة" فبناء على درجة الثقة المطلوبة يتحدد معامل الثقة.

2- إن درجات ومعاملات الثقة التي ذكرناها تخص التوزيع الطبيعي، وأن المعاملات 1 أو 2 أو 3 ما هي إلا الدرجة المعيارية (Z) والتي نحصل عليها من جدول التوزيع الطبيعي المعياري وذلك بقسمة درجة الثقة (أو الاحتمال) على 2 (حيث أن المساحة موزعة بالتساوي على يمين ويسار الوسط) ثم بالكشف في المساحات تحت المنحنى الطبيعي المعياري عن خارج القسمة (أو أقرب رقم له) فنحصل على Z المقابلة وهذا يرجع إلى أن توزيع المعاينة للوسط هو التوزيع الطبيعي.

3- يمكن الحصول على فترات تقدير بأي درجة ثقة أخرى (غير الثلاث التي ذكرناها) وذلك بقسمة درجة الثقة المطلوبة - كما ذكرنا - على 2 ثم الكشف في المساحات حتى نحصل على Z المناسبة.

4- أن فترة التقدير (أو الثقة) للوسط الحسابي للمجتمع تأخذ الشكل التالي:

$$\hat{\mu} = \overline{X} \pm Z\sigma\overline{x}$$

حيث تحدد Z درجة الثقة المطلوبة

وحيث أن الخطأ المعياري للوسط $\sigma_{\overline{x}}$ يأخذ الشكل التالي:

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

ولذا فإن فترة تقدير الوسط تأخذ الشكل النهائي التالي:

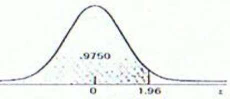
$$\widehat{\mu} = \overline{X} \pm Z\frac{\sigma}{\sqrt{n}}$$

ويمكن تلخيص خطوات تقدير الوسط الحسابي للمجتمع فيما يلي :

١. احسب الوسط الحسابي للعينة $\overline{X}$
٢. احسب الخطأ المعياري للوسط والذي يساوي: $\frac{\sigma}{\sqrt{n}}$
٣. أضرب الخطأ المعياري للوسط في معامل الثقة المناسب (أو الدرجة المعيارية) حسب درجة الثقة المطلوبة أي أحسب: $Z\frac{\sigma}{\sqrt{n}}$
٤. فعندما نطرح حاصل الضرب السابق من الوسط الحسابي للعينة نحصل بالتالي على الحد الأدنى لفترة التقدير، وعندما نجمع حاصل الضرب مرة أخرى على الوسط الحسابي للعينة نحصل بالتالي على الحد الأعلى لفترة التقدير

جدول التوزيع الطبيعي المعياري:

جدول التوزيع الطبيعي المعياري



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	z
0.00	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359	0.00
0.10	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753	0.10
0.20	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141	0.20
0.30	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517	0.30
0.40	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879	0.40
0.50	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224	0.50
0.60	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549	0.60
0.70	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852	0.70
0.80	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133	0.80
0.90	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389	0.90
1.00	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621	1.00
1.10	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830	1.10
1.20	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015	1.20
1.30	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177	1.30
1.40	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319	1.40
1.50	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441	1.50
1.60	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545	1.60
1.70	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633	1.70
1.80	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706	1.80
1.90	.9713	.9719	.9726	.9732	.9738	.9744	**.9750**	.9756	.9761	.9767	1.90
2.00	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817	2.00
2.10	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857	2.10
2.20	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890	2.20
2.30	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916	2.30
2.40	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936	2.40
2.50	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952	2.50
2.60	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964	2.60
2.70	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974	2.70
2.80	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981	2.80
2.90	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986	2.90
3.00	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990	3.00
3.10	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993	3.10
3.20	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995	3.20
3.30	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997	3.30
3.40	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998	3.40
3.50	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	3.50
3.60	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	3.60
3.70	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	3.70
3.80	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	3.80

z	−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	−0.03	−0.02	−0.01	0.00	z
−3.80	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	−3.80
−3.70	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	−3.70
−3.60	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	−3.60
−3.50	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	−3.50
−3.40	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	−3.40
−3.30	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005	−3.30
−3.20	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007	−3.20
−3.10	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010	−3.10
−3.00	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013	−3.00
−2.90	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019	−2.90
−2.80	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026	−2.80
−2.70	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035	−2.70
−2.60	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047	−2.60
−2.50	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062	−2.50
−2.40	.0064	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.0080	.0082	−2.40
−2.30	.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104	.0107	−2.30
−2.20	.0110	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136	.0139	−2.20
−2.10	.0143	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174	.0179	−2.10
−2.00	.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222	.0228	−2.00
−1.90	.0233	.0239	.0244	.0250	.0256	.0262	.0268	.0274	.0281	.0287	−1.90
−1.80	.0294	.0301	.0307	.0314	.0322	.0329	.0336	.0344	.0351	.0359	−1.80
−1.70	.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427	.0436	.0446	−1.70
−1.60	.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526	.0537	.0548	−1.60
−1.50	.0559	.0571	.0582	.0594	.0606	.0618	.0630	.0643	.0655	.0668	−1.50
−1.40	.0681	.0694	.0708	.0721	.0735	.0749	.0764	.0778	.0793	.0808	−1.40
−1.30	.0823	.0838	.0853	.0869	.0885	.0901	.0918	.0934	.0951	.0968	−1.30
−1.20	.0985	.1003	.1020	.1038	.1056	.1075	.1093	.1112	.1131	.1151	−1.20
−1.10	.1170	.1190	.1210	.1230	.1251	.1271	.1292	.1314	.1335	.1357	−1.10
−1.00	.1379	.1401	.1423	.1446	.1469	.1492	.1515	.1539	.1562	.1587	−1.00
−0.90	.1611	.1635	.1660	.1685	.1711	.1736	.1762	.1788	.1814	.1841	−0.90
−0.80	.1867	.1894	.1922	.1949	.1977	.2005	.2033	.2061	.2090	.2119	−0.80
−0.70	.2148	.2177	.2206	.2236	.2266	.2296	.2327	.2358	.2389	.2420	−0.70
−0.60	.2451	.2483	.2514	.2546	.2578	.2611	.2643	.2676	.2709	.2743	−0.60
−0.50	.2776	.2810	.2843	.2877	.2912	.2946	.2981	.3015	.3050	.3085	−0.50
−0.40	.3121	.3156	.3192	.3228	.3264	.3300	.3336	.3372	.3409	.3446	−0.40
−0.30	.3483	.3520	.3557	.3594	.3632	.3669	.3707	.3745	.3783	.3821	−0.30
−0.20	.3859	.3897	.3936	.3974	.4013	.4052	.4090	.4129	.4168	.4207	−0.20
−0.10	.4247	.4286	.4325	.4364	.4404	.4443	.4483	.4522	.4562	.4602	−0.10
−0.00	.4641	.4681	.4721	.4761	.4801	.4840	.4880	.4920	.4960	.5000	0.00

ويمكن سرد أشهر وأهم درجات ومعاملات الثقة (للتوزيع الطبيعي) في الجدول التالي (مع ملاحظة أن 95%، 99% هي أشهرها على الإطلاق):

معامل الثقة Z	درجة الثقة
1	68.26%
1.65	90%
1.96	**95 %**
2	95.44%
2.58	**99%**
3	99.72%

مع ملاحظة أنه إذا كان الانحراف المعياري للمجتمع غير معروف - وهو غالبا ما يحدث في الواقع – فيمكن استخدام الانحراف المعياري للعينة S بدلا منه طالما كان حجم العينة كبيرا بدرجة كافية وتصبح فترة تقدير الوسط الحسابي للمجتمع كما يلي:

$$\widehat{\mu} = \overline{X} \pm Z \frac{S}{\sqrt{n}}$$

ولإيضاح هذه النقطة بشيء من التفصيل نأخذ المثال التالي:

مثال:

لو أردنا معرفة متوسط الدخل اليومي لمجموعة من الناخبين في دولة ما، فإن ذلك يبدو أمرا صعباً من الناحية العملية نظراً لكبر حجم مجتمع الناخبين، إضافة إلى طول الوقت والتكاليف. لذا فإن الأسلوب العلمي المتبع في حالة كهذه هو اختيار عينة عشوائية نستطيع من خلال معرفة نتائجها لتقدير متوسط دخول الناخبين في هذه الدولة.

فلو سحبت عينة عشوائية من مجموع مجتمع الناخبين في دولة ما حجمها 100 ناخب، فإذا كان الوسط الحسابي والانحراف المعياري للدخل السنوي للناخبين بالعينة هما على الترتيب 90 ألف ريال و 25 ألف ريال.

المطلوب:

أوجد فترة تقدير للوسط الحسابي للدخل السنوي لمجموع الناخبين في هذه الدولة بدرجة ثقة 95% ؟

الحل:

بما أن فترة تقدير الوسط الحسابي للمجتمع هي: $\widehat{\mu} = \overline{X} \pm Z \frac{S}{\sqrt{n}}$

والمعلومات المعطاة هي :

حجم العينة n = 100

الوسط الحسابي للعينة = $\overline{X} = 90$

والانحراف المعياري للعينة S = 25

وحيث أن درجة الثقة هي 95% فإن: Z = 1.96 حسب ما هو موضح في الجدول السابق. وبالتالي فإن فترة تقدير الوسط الحسابي للدخل السنوي لمجتمع الناخبين بدرجة ثقة 95% هي :

$$\hat{\mu} = 90 \pm 1.96 \frac{25}{\sqrt{100}}$$
$$= 90 \pm 1.96(2.5) = 90 \pm 4.9$$
$$\mu = \begin{cases} 85.1 \\ 94.9 \end{cases}$$

أي أن الوسط الحسابي للدخل السنوي لمجتمع الناخبين يتراوح بين 85.1 ألف ريال كحد أدنى، 94.9 ألف ريال كحد أعلى، وذلك بدرجة ثقة 95 %.

مثال:

أخذت عينة عشوائية حجمها 144 بوسط مقداره 100 وانحراف معيارى مقداره 60 وبالتالي فإن فترة تقدير الوسط الحسابي للمجتمع بدرجة ثقة 95% هي :

$$\hat{\mu} = \overline{X} \pm 1.96 S_{\overline{X}}$$
$$= \overline{\mathrm{X}} \pm 1.96 \frac{S}{\sqrt{n}}$$
$$= 100 \pm 1.96 \frac{60}{\sqrt{144}} = 100 \pm 1.96(5) = 100 \pm 9.8$$

أى أن $\hat{\mu}$ تقع بين 90.2 , 109.8 بدرجة ثقة 95%. وكثيرا ما تستخدم أيضاً درجات الثقة 90 , 99% وهى مناظرة لقيمة z=1.64 , z=2.58 على الترتيب

تحديد حجم العينة لتقدير الوسط الحسابي للمجتمع

يعتبر تحديد حجم العينة المناسب من المشاكل المهمة والشائعة التي تواجه الباحثين في مختلف المجالات، وبالذات عند دراسة الظواهر السياسية والإجتماعية ...الخ ، ويختلف تحديد حجم العينة باختلاف الهدف من التقدير.

فإذا كان المطلوب هو تقدير الوسط الحسابي للمجتمع، فإن فترة تقدير الوسط هي كما سبق وأن أوضحنا: $\hat{\mu} = \overline{X} \pm Z \frac{\sigma}{\sqrt{n}}$

ومنها نجد أن حجم العينة يأخذ الشكل التالي: $n = \frac{Z^2 \sigma^2}{e^2}$

- Z = هو معامل الثقة (أو الدرجة المعيارية) المقابل لدرجة الثقة المطلوبة، ونحصل عليا من جدول التوزيع الطبيعي المعياري.
- $\{\sigma^2\}$ = هو تباين المجتمع (أو هو مربع الانحراف المعياري).
- e = هو أقصى خطأ مسموح به في تقدير الوسط، وهو عادة ما يحدده الباحث، وتتوقف قيمته على أهمية الموضوع أو الظاهرة السياسية المراد دراستها، ومدى الدقة المطلوبة في التقدير، ويسمى اختصاراً "الخطأ في تقدير الوسط".

ولتوضيح كيفية تحديد حجم العينة المناسب عند تقدير الوسط الحسابي للمجتمع، نأخذ المثال التالي :

مثال:

إذا كانت دخول الأفراد اليومية في إحدى دول العالم النامية تتبع التوزيع الطبيعي بانحراف معياري $\sigma = 15$ دولاراً، فما هو حجم العينة المناسب لتقدير متوسط دخول الأفراد في هذه الدولة بحيث لا يتعدى الخطأ في تقدير متوسط الدخل اليومي 5 دولارات، وذلك بدرجة ثقة 99 %؟

الحل :

في هذا المثال نجد أن :

درجة الثقة % 99 أي أن : Z = 2.58

أقصى خطأ مسموح به هو 5 دولارات، أي أن: e = 5

والانحراف المعياري للمجتمع: $\sigma = 15$

وبالتعويض بهذه القيم في المعادلة التي تحدد حجم العينة وهي: $n = \frac{Z^2 \sigma^2}{(e)^2}$

فإن حجم العينة مقرباً لأقرب عدد صحيح هو:

$$n = \frac{(2.58)^2 (15)^2}{5^2} = \frac{(6.65)(225)}{25} = \frac{1496.25}{25} = 59.85 \approx 60$$

أي أنه يجب على الباحث أن يأخذ عينة لا يقل حجمها عن 60 فرداً حتى يكون لديه تقديراً دقيقاً عن متوسط دخول الأفراد في هذه الدولة بحيث لا يتعدى الخطأ في تقديره لمتوسط الدخل عن خمس دولارات، وذلك بدرجة ثقة % 99.

مثال:

يرغب أحد مدراء إحدى المصانع فى تقدير متوسط عدد الدقائق التى يأخذها العمال لإنجاز عملية صناعية معينة بحيث لا يتعدى الخطأ في تقدير متوسط الأداء في حدود ± 3 دقيقة وبدرجة ثقة 90% . ويعلم المدير من خبرته الماضية أن الانحراف المعيارى σ هو 15 دقيقة .

الحل:

في هذا المثال نجد أن :

درجة الثقة 90% أي أن : Z = 1.65

أقصى خطأ مسموح به هو 3 دقائق، أي أن : e = 3

والانحراف المعياري للمجتمع: $\sigma = 15$

وبالتعويض بهذه القيم في المعادلة التي تحدد حجم العينة وهي: $n = \frac{Z^2 . \sigma^2}{(e)^2}$

فإن حجم العينة مقرباً لأقرب عدد صحيح هو :

$$n = \frac{(1.65)^2 (15)^2}{3^2} = \frac{(2.72)(225)}{9} = \frac{612}{9} = 68$$

أي أنه يجب على المدير أن يأخذ عينة لا يقل حجمها عن 68 فرداً حتى يكون لديه تقديراً دقيقاً لعدد الدقائق التى يأخذها العمال لإنجاز عملية صناعية معينة بحيث لا يتعدى الخطأ في تقديره لمتوسط الإنجاز عن ثلاث دقائق، وذلك بدرجة ثقة % 90.

المحاضرة 9

التقدير الإحصائي (الجزء الثاني)

المقصود بالتقدير هو تقدير معالم المجتمع الإحصائي (أو التوزيع الاحتمالي) والتي غالباً ما تكون مجهولة ويكون المطلوب هو الحصول على تقديرات لها -عادة– من بيانات العينة فقد يكون المطلوب تقدير متوسط دخل الدولة، أو تقدير متوسط عمر الناخب... وغيرها.

وكما أوضحنا في المحاضرة السابقة أن هناك نوعان (أو أسلوبان) للتقدير **يسمى الأول** تقدير النقطة (أو القيمة الواحدة)، **ويسمى الثاني** تقدير الفترة (أو فترة التقدير أو الثقة).

ففي حالة تقدير النقطة نحصل على قيمة واحدة من العينة، وتستخدم هذه القيمة الواحدة كتقريب أو كتقدير لمعلمة المجتمع المجهولة. فمثلاً لو أخذنا الوسط الحسابي للدخل في العينة كتقدير لمتوسط الدولة نكون قد حصلنا على تقدير نقطة لمتوسط دخل الدولة.

أما في حالة تقدير الفترة أو فترة التقدير فنحصل على مدى Range أو فترة تتحدد بحدين (حد أدنى وحد أعلى) - نحصل عليهما من العينة. ونلاحظ هنا أن فترة التقدير (أو تقدير الفترة) تحتوي على أكثر من قيمة بل قد يكون عدد القيم غير محدود أو لا نهائياً في كثير من الحالات.

فترات الثقة للمتوسط باستخدام توزيع t

تناولنا فيما سبق التقدير الإحصائي للوسط الحسابي للمجتمع في الحالات التي يكون فيها الانحراف المعياري للمجتمع معلوماً، و (أو) أن العينة كبيرة بدرجة كافية.

ولكن إذا كانت العينة صغيرة بمعنى أن حجمها أقل من (30) مفردة، والانحراف المعياري للمجتمع الطبيعي غير معلوم، فإن التوزيع الإحصائي المتبع في مثل هذه الحالات هو ما يطلق عليه " توزيع t "

توزيع ***t*** هو توزيع متماثل حول متوسط الصفر ولكنه منبسط عن التوزيع الطبيعي القياسي ، ولهذا فان جزاءاً أكبر من مساحته تقع عند الأطراف. وبينما يوجد توزيع طبيعى قياسى واحد ، فإن هناك توزيعاً ***t*** مختلفاً لكل حجم للعينة **n**. ولكن مع تزايد **n** فإن توزيـع ***t*** يقترب من التوزيع الطبيعى القياسى (أنظر الشكل التالي) إلى أن تكون $n \geq 30$ ، وعندئذ يتساويان تقريباً.

ولعل الاختلاف الأساسي بين توزيع t والتوزيع الطبيعي هو أن الانحراف المعياري للعينة هو المستخدم في الأول بدلا من الانحراف المعياري للمجتمع في الثاني، وفيما عدا ذلك فالتوزيعان متماثلان وكلما زادت قيمة n كلما اقترب توزيع t من توزيع z ويعتمد توزيع t على ما يعرف بدرجات الحرية DEGREES OF FREEDOM .

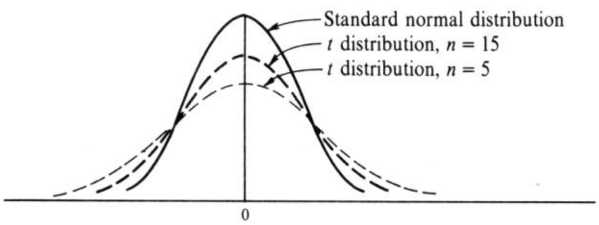


Fig. 4-3

نلاحظ من خلال الشكل السابق أنه كلما كبر حجم العينة n كلما قرب توزيع t من التوزيع الطبيعي

ويعطى جدول توزيع *t* قيم ***t*** التى على يمينها تمثل المساحة تحت المنحنى **10%**، **5%**، **2.5%**، **1%** و **0.5%** من المساحة الكلية تحت المنحنى **لدرجات الحرية** المختلفة .

درجات الحرية :

تعرف درجات الحرية بأنها عدد المشاهدات المستقلة في العينة والتي تساوي حجم العينة مطروحاً منه عدد القيود أو معالم المجتمع التي يتم تقديرها من بيانات العينة.

وكمثال مبسط لشرح فكرة درجات الحرية نفترض أن لدينا 3 قيم واشترطنا أن مجموع القيم يساوي 10 فإن لدى الباحث في هذه الحالة حرية في اختيار الرقم الأول (وليكن 2) والثاني (وليكن 3) لذلك فإن قيمة الثالثة لابد وأن تكون (5) بالتالي نستطيع القول بأن درجة الحرية المتاحة لدى الباحث هي (2) أي 1 - 3 = 2 أي أن درجات الحرية في هذه الحالة هي :

n - 1

حيث n تساوي حجم العينة (والتي تساوي في المثال السابق 3)

والرقم (1) والذي طرحناه يعني الشرط الذي يحتم أن مجموع القيم = 10 وبصفة عامة إذا كان عدد القيود k فإن درجات الحرية تساوي n - k

وفترة الثقة **95%** لوسط المجتمع غير المعلوم عند استخدام توزيع ***t*** هى:

$$P(\overline{X} - t\frac{s}{\sqrt{n}} < \mu < \overline{X} + t\frac{s}{\sqrt{n}}) = 0.95$$

حيث تشير ***t*** الى قيمة ***t*** التى تقع عندها **2.5%** من المساحة الكلية للمنحنى عند كل طرف (عند درجات الحرية المستخدمة)، وتستخدم $s/\sqrt{n}$ بدلا من $\sigma_{\bar{x}} = \sigma/\sqrt{n}$

شروط توزيع t :

ويمكن تحديد الشروط الثلاثة لاستخدام توزيع t كما يلي:

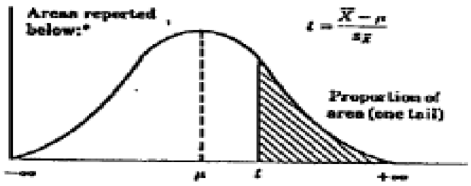
١. أن يكون المجتمع المسحوبة منه العينة له توزيع طبيعي.
٢. والانحراف المعياري للمجتمع غير معروف (أو مجهول).
٣. والعينة صغيرة (حجمها أقل من 30 مفردة).

جدول توزيع t:

الجدول أدناه يعطي ــــ قيمة = t

المقابلة للمساحة المظللة وقيمتها ∝

Proportions of Area for the t Distributions



df	0.10	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898

df	0.10	0.05	0.025	0.01	0.005
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

مثال على العينة أقل من 30 و المجتمع يتبع التوزيع الطبيعي :

سحبت عينة عشوائية من n=10 بطارية فلاش متوسطها 5 ساعات، والانحراف المعيارى للعينة s=1 ساعة من خط إنتاج من المعروف أنه ينتج بطاريات عمرها موزع طبقاً للتوزيع الطبيعى .

المطلوب :

إيجاد فترة الـ 95% ثقة للمتوسط غير المعلوم لعمر البطاريات فى المجتمع كله.

الحل:

لإيجاد فترة الـ 95% ثقة للمتوسط غير المعلوم لعمر البطاريات فى المجتمع كله، فإننا نوجد أولا قيمة t 0.025 و التى تكون معها 2.5% من المساحة عند الأطراف لدرجات حرية n-1=9 . ونحصل على هذه القيمة من خلال الرجوع إلى جدول t بالتحرك تحت عمود 0.025 حتى درجات حرية 9 والقيمة التى سيتم التحصل عليها هى 2.262 إذن:

$$\hat{\mu} = \overline{X} \pm 2.262\frac{s}{\sqrt{n}} = 5 \pm 2.262\frac{1}{\sqrt{10}} \cong 5 \pm 2.262(0.316) \cong 5 \pm 0.71$$

وتقع $\hat{\mu}$ بين **4.29 , 5.71** ساعة بدرجة ثقة **95%** (أنظر الشكل التالي):

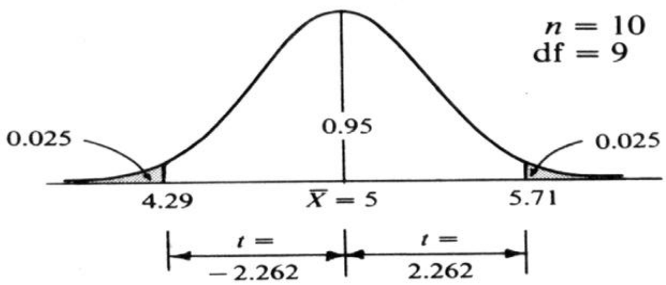


Fig. 4-4

مثال:

إذا رغبت جهة معينة إجراء دراسة على بيانات محددة، أوجد التالي:

١. كيف يمكن إيجاد قيمة t التى تناظر 10% من المساحة عند الأطراف وبدرجات حرية 9 ؟
٢. كيف تفسر قيم t مختلفاً عن تفسير قيم z ؟
٣. أوجد قيم t المناظرة لنسب 5، 0.5 ، 2.5 % من المساحة عند الأطراف لعدد 9 من درجات الحرية ؟
٤. أوجد قيم t المناظرة لنسب 5، 0.5 ، 2.5 % عند الأطراف لحجم عينة n كبير جداً أو لا نهائي ؟
٥. كيف تقارن بين t هذه بقيم z المناظرة ؟

الحل:

١. يمكن الحصول على قيمة t المناظرة لنسبة **10%** من المساحة **عند الأطراف** بالتحرك عبر العمود الذى رأسه **0.10** فى جدول t حتى نصل إلى درجات حرية **9**. وهذا يعطى قيمة t تساوى **t=+1.383**. وبالتماثل، فإن **10%** من المساحة لتوزيع t بدرجات حرية **9** تقع عند الطرف الأيسر وذلك إلى اليسار من **t =-1.383.**
٢. تشير قيم t فى جدول توزيع t إلى المساحات (الاحتمالات) عند أطراف توزيع t المقابلة لدرجات الحرية المعينة، أما قيم z فى جدول التوزيع الطبيعي فإنها تشير إلى المساحات (الاحتمالات) تحت المنحنى الطبيعي المعياري التي **تقع بين المتوسط وبين قيم z المحددة**.

٣. بالتحرك عبر الأعمدة التى رؤوسها **0.05**، **0.025**، و **0.005** فى جدول توزيع t حتى نصل إلى **9 df** ،نحصل على قيم **t =+1.833**، **t=+2.262** و **t=+3.250** على الترتيب . وكنتيجة للتماثل فإن **5**، **2.5**، و **0.5%** من المساحة تقع فى الطرف الأيسر لتوزيع ***t*** لدرجات حرية **9** إلى اليسار من **t=-1.833** ، **t=-2.262**، **t=-3.350** على الترتيب .

٤. عندما يكون حجم العينات (ودرجات الحرية) كبيرة جـــداً أو نهائية فإن قيمة $t_0.05 = 1.645$ ، $t_0.005 = 2.576$ و $t_0.025 = 1.960$ (من الصف الأخير فى جدول توزيع t) . وهذه تتطابق مع قيم **z** المناظرة فى جدول التوزيع الطبيعي . وبالتحديد $t_0.05 = 1.645$ تعنى أن **2.5%** من المساحة تحت توزيع ***t*** بدرجات حرية ∞ تقع عند الطرف الأيمن ، إلى اليمين من **t=1.96** . وبالمثل ، فإن **z=1.96** تعطى (من جدول التوزيع الطبيعي) **0.4750** من المساحة تحت التوزيع الطبيعي القياسى من $\mu = 0$ إلى **z=1.96**

وعليه لدرجات حرية هي $df = n-1 = \infty$ فإن توزيع ***t*** يتطابق مع التوزيع القياسي الطبيعي

فترة تقدير النسبة للمجتمع (فترة الثقة للنسبة)

إن تقدير النسبة في المجتمع تعتبر من الحالات المهمة لقياس الظواهر السياسية، وبالذات الوصفية منها كقياس اتجاهات الرأي العام، وقياس نسبة قتلى الحروب، ونسبة الدول التي أوفت بالتزاماتها في المنظمات الدولية أو الإقليمية... وغيرها ونظراً لأنه من الصعوبة بمكان في كثير من الأحيان حساب هذه النسبة مباشرة من المجتمع، فإننا غالبا ما نلجأ لتقدير هذه النسبة من عينة عشوائية مسحوبة من هذا المجتمع.

خطوات تقدير النسبة في المجتمع:

فلو افترضنا أن نسبة المؤيدين للسياسة الاقتصادية التي تنتهجها دولة ما هي *P* وأن العينة العشوائية كبيرة بدرجة كافية وأن نسبة مؤيدي هذه السياسة في العينة هي $\hat{P}$ فإن خطوات تقدير النسبة في المجتمع تكون كما يلي:

١. حساب النسبة في العينة $\hat{P}$

٢. حساب الخطأ المعياري للنسبة والتي تساوي في هذه الحالة: $\sigma_{\hat{P}} = \sqrt{\dfrac{\hat{P}(1-\hat{P})}{n}}$

٣. ضرب الخطأ المعياري للنسبة في معامل الثقة المناسب Z (حسب درجة الثقة المطلوبة) والتي نحصل عليها من جدول التوزيع الطبيعي المعياري (أو من الجدول الذي يحوي أهم درجات ومعاملات الثقة والذي ذكرناه آنفاً). أي نحسب:

$$Z\sigma_{\hat{P}} = Z\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

٤. للحصول على الحد الأدنى لتقدير النسبة نطرح حاصل الضرب (السابق) من نسبة العينة $\hat{P}$ وللحصول على الحد الأعلى نجمع حاصل الضرب مع النسبة في العينة فنحصل على فترة تقدير النسبة. وبالتالي فإن فترة تقدير النسبة تكون في شكلها النهائي كما يلي: $P = \hat{P} \pm Z\sqrt{\dfrac{\hat{P}(1-\hat{P})}{n}}$

مثال:

عينة عشوائية حجمها 144 ناخباً سحبت من إحدى المدن فوجد أن عدد المؤيدين في العينة لمرشح معين هو 60 ناخباً، أنشئ فترة تقدير لنسبة المؤيدين لهذا المرشح في المدينة كلها بدرجة ثقة % 95.

الحل:

نحسب أولاً نسبة المؤيدين للمرشح في العينة $\hat{P}$ التي نحصل عليها بقسمة عدد المؤيدين له على العدد الكلي للعينة (حجم العينة)
أي أن: $\hat{P} = \frac{60}{144} = 0.42$

وحيث أن درجة الثقة المطلوبة هي % 95 فإن معامل الثقة المناســـب هو: Z = 1.96 وفترة تقدير نسبة المؤيدين لهذا المرشح في المدينة تأخذ الشكل التالي: $P = \hat{P} \pm z\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$

وبالتعويض عن حجم العينة n = 144

والنسبة في العينة $\hat{P} = 0.42$، $1-\hat{P} = 1-0.42 = 0.58$

ومعامل الثقة Z = 1.96

نحصل بعدها على:

$$\begin{aligned}
P &= 0.42 \pm 1.96\sqrt{\frac{0.42 \times 0.58}{144}} \\
&= 0.42 \pm (1.96)(0.0411) \\
&= 0.42 \pm 0.08
\end{aligned}$$

$$\therefore P \begin{cases} 0.34 \\ 0.50 \end{cases}$$

أي أن نسبة المؤيدين للمرشح في المدينة تتراوح بين 0.34 , 0.50 وذلك بدرجة ثقة % 95 ، بمعنى آخر أن نسبة مؤيدي هذا المرشح في هذه المدينة لا تتجاوز % 50 كحد أعلى، وبالتالي ففرصته في الفوز كمرشح قد لا تكون كبيرة وذلك بدرجة ثقة % 95 بمعنى أن هذا الحكم لا تتجاوز نسبة الخطأ فيه % 5.

المحاضرة 10

اختبار الفروض الإحصائية

يعتبر اختبارات الفروض الإحصائية واحدة من أهم التطبيقات التي قدمها علم الإحصاء كحل للمشاكل العلمية المختلفة بشتى فروع العلم ورغم أهمية موضوع تقدير المعالم إلا انه غالبا ما يكون الاهتمام مركز ليس علي مجرد تقدير المعالم ولكن علي عملية وضع قواعد تمكن من التوصل إلي قرار بقبول أو رفض خاصية أو بالمعني الإحصائي فرض عن معالم مجتمع واحد أو أكثر وهذا ما يسمي اختبارات الفروض الإحصائية، ومن خلالها يمكن لأي شخص أن يتخذ قرار برفض أو قبول فرض معين أو مجموعة من الفروض المتعلقة بمشكلة معينة موجودة في الحياة العامة.

وبصفه عامه فان قبول او رفض اى قرار يجب ان يمر بعدة مراحل وهى:

١. سحب عينة من المجتمع بحيث تكون ممثله للمجتمع (عشوائية)
٢. تجميع البيانات المتعلقة بالمشكلة من العينة
٣. تطبيق قواعد معينه لاختبار الفروض الموضوعه عن طريق الباحث وهى مشكله تحتاج لخبره احصائية
٤. اتخاذ القرار بناء على ما توصل اليه الباحث من نتائج.

اختبارات الفروض الاحصائية Testing Statistical Hypotheses

من المعروف ان اتخاذ اى قرار لا يتم الا من خلال اختبارات الفروض الاحصائية التى تعتمد بدورها كما سبق على الاحتمالات وتوزيعات المعاينه وهذا يؤكد اهمية الدور الذى تلعبه نظريه الاحتمالات فى التنبؤ والتخطيط واتخاذ القرارات بالاضافه الى اهميتها فى تقدير معالم المجتمع المجهوله والتى تعتبر احد اهتمامات الباحثين.

تبدأ مشكله التعرف على معالم المجتمع المجهوله بما يسمى بالاستدلال الإحصائى (Statistical Inferences) حيث ينقسم الى فرعين: الفرع الأول يهتم بتقدير (Estimation) معالم المجتمع والفرع الأخر يختص باجراء اختبارات فروض (Testing Hypotheses) تدور حول معالم المجتمع المجهوله.

الاستدلال الاحصائي يتم باستخدام عينة عشوائية مسحوبه من المجتمع وذلك لاستحالة التعامل مع المجتمع ككل، فالاحصاءات التحليلية قدمت القوانين التى سهلت هذه العملية وجعلتها تتم بأقل الاخطاء الممكنه.

اختبارات الفروض ترتكز اساسا على فكرة ان هناك عينه اخرى غير مسحوبه من المجتمع المسحوب منه العينات فان الفرق بين الوسط الحسابي المحسوب من هذه العينه وبين المعلمه المجهوله قد يكون فرقا معنويا Significant غير راجع للصدفه. واشتقت اسمها منها حيث عن طريقها نستطيع ان نحدد وبسهوله هل الفروق بين المعلومات المحسوبة من العينة وبين المعلومات المفروضه لمجتع معين فرقا يرجع الى الصدفه ام فرق حقيقى، وبأسلوب أخر هل هو فرق معنوى او فرق غير معنوى؟ وبذلك سميت هذه الاختبارات باسم اختبارات المعنويه Test of Significant

مفهوم الاختبارات الإحصائية المعلمية واللامعلمية

الاختبارات الاحصائيه قد تدور حول معالم المجتمع المجهوله مثل الفروض المتعلقة بالوسط الحسابى، النسبه، التباين، معامل الارتباط،... وفى هذه الحاله يطلق على هذه الاختبارات اسم الاختبارات المعلميه Parametric Tests

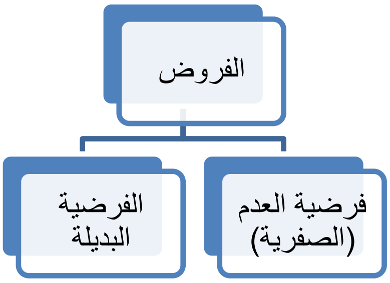
وقد تكون فروضا لا تتعلق بمعالم المجتمع ولكن تتعلق بأشياء اخرى قد تكون وصفية مثل العلاقة بين التعليم والتدخين، خضوع نتائج معينه لنظريه معينه، العلاقة بين لون العينين ولون الشعر،.... وفى هذه الحالة يسمى الاختبار باسم الاختبار اللامعلمي Non Parametric Test

فالإحصاء المعلمي: اساليب احصائية تتطلب افتراضات محددة عن طبيعة التوزيعات الاحتمالية للمجتمع (مثل حساب الوسط الحسابي كمقياس للنزعة المركزية)

أما الإحصاء اللامعلمي: اساليب احصائية تتطلب افتراضات أقل عن طبيعة التوزيعات الاحتمالية للمجتمع (مثل حساب الوسيط كمقياس للنزعة المركزية)

الفروض الإحصائية

تعتبر الفروض الإحصائية بمثابة اقتراح عن معالم المجتمع موضوع الدراسة، والتي ما زالت غير معلومة للباحث، فهي حلول ممكنة لمشكلة البحث



الفرضية الصفرية (فرضية العدم) H_O (Null Hypothesis) :

هي الفرضية حول معلمة المجتمع التي نجري اختبار عليها باستخدام بيانات من عينة والتي تشير أن الفرق بين معلمة المجتمع والإحصائي من العينة ناتج عن الصدفة ولا فرق حقيقي بينهما. وهي الفرضية التي ننطلق منها ونرفضها عندما تتوفر دلائل على عدم صحتها، وخلاف ذلك نقبلها وتعني كلمة Nul انه لا يوجد فرق بين معلمة المجتمع والقيمة المدعاة (إحصائية العينة).

الفرضية البديلة Alternative Hypothesis (H_a) :

هي الفرضية التي يضعها الباحث كبديل عن فرضية العدم و نقبلها عندما نرفض فرضية العدم باعتبارها ليست صحيحة بناء على المعلومات المستقاة من العينة.

مستوى الدلالة الإحصائية ومنطقة الرفض

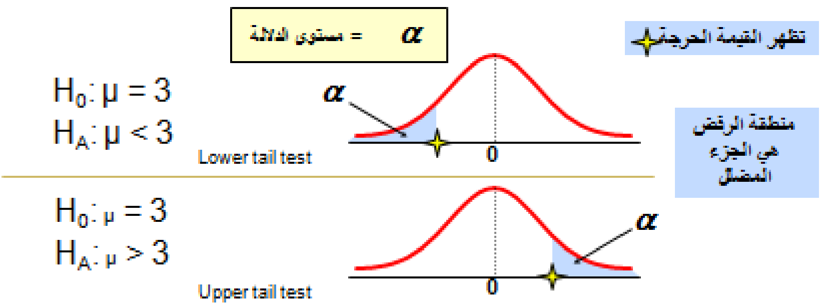
عندما نقبل الفرضية الصفرية (فرضية العدم) فإننا نقبلها بنسبة دقة 90% أو 95% أو 99% أو غير ذلك وتسمى مستويات الدلالة أو الثقة Significance Levels أي يوجد نسبة خطأ معين في قبولنا للفرضية المبدئية بمعنى أننا نقبل صحة الفرضية الصفرية وهي خاطئة وهذا الخطأ هو α ويسمى مستوى المعنوية.

أي إذا كان مستوى الثقة 95% α -1 فان مستوى المعنوية α تساوي 5% وهي عبارة عن مساحة المنطقة التي تقع تحت منحنى التوزيع والتي تمثل منطقة الرفض، وتكون أما على صورة ذيل واحد جهة اليمين أو اليسار أو ذيلين متساويين في المساحة واحد جهة اليمين والثاني جهة اليسار.

اختبار الفرضيات من طرف واحد:

هو الاختبار الذي تبين فيه الفرضية البديلة بأن معلمة المجتمع اكبر أو اصغر من معلمة المجتمع المفترضة.

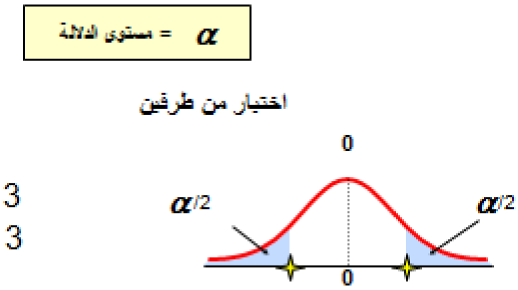
مستوى الدلالة الإحصائية ومنطقة الرفض من طرف واحد



اختبار الفرضيات من طرفين:

هو الاختبار الذي لا تبين فيه الفرضية البديلة بأن معلمة المجتمع اكبر أو اصغر من معلمة المجتمع المفترضة بل مجرد أنها تختلف.

مستوى الدلالة الإحصائية ومنطقة الرفض من طرفين



$H_0: \mu = 3$
$H_A: \mu \neq 3$

اختبار الفروض عن خصائص المجتمع:

إن اختبار الفروض عن خصائص المجتمع (مثل μ و σ) هو جانب أساسي أخر من جوانب الاستدلال والتحليل الإحصائي. وفي اختبار الفروض نبدأ بعمل فرض ما عن خاصية المجتمع غير المعلومة. ثم نأخذ عينة عشوائية من المجتمع، وعلى أساس الخاصية المناظرة في العينة، أما أن نقبل وإما أن نرفض الفرض بدرجة ثقة محددة.

وفي اختبار الفروض يمكن أن نرتكب نوعين من الخطأ:

الخطأ من النوع الأول Type I error: الخطأ من النوع الأول هو "رفض الفرض العدمي بينما هو صحيح ". أي أنه على الرغم من أن الفرض العدمي في الواقع صحيح وكان من الواجب قبوله فقد تم أخذ قرار خاطئ برفضه. وباختصار شديد فإن الخطأ من النوع الأول هو: " رفض فرض صحيح". ويرمز له بالرمز α .

الخطأ من النوع الثاني Type II error: في المقابل فإن الخطأ من النوع الثاني يعني " قبول الفرض العدمي بينما هو خاطئ " أي أنه على الرغم من أن الفرض العدمي خاطئ وكان من الواجب رفضه فقد تم أخذ قرار خاطئ بقبوله وباختصار شديد فإن الخطأ من النوع الثاني هو " قبول فرض خاطئ ". ويرمز له بالرمز β .

<u>ويمكن أن نمثل ذلك في الجدول التالي:</u>

القرار \ الفرضية	صحيحة (H_O)	خاطئة (H_a)
قبول (H_O)	صواب	خطأ 2 بيتا (B)
رفض (H_O)	خطأ 1 ألفا (α)	صواب

١. فرضية صحيحة نتائج العينة تؤيد صحتها. (قبول صواب)
٢. فرضية صحيحة نتائج العينة غير مؤيدة لصحتها. (رفض خطأ) وهذا يعطينا خطأ من النوع الأول ألفا (α) ويمكن أن يقلل برفع مستوى الدلالة
٣. فرضية خاطئة نتائج تؤيد صحتها (قبول خطأ) وهذا يعطينا خطأ من النوع الثاني بيتا (B) ويمكن أن يقلل بزيادة حجم العينة
٤. فرضية خاطئة نتائج غيرمؤيدة صحتها (رفض صواب)

مستوى المعنوية Level of Significance

يعتبر مصطلح **"مستوى المعنوية"** واحدا من أهم المصطلحات المستخدمة في دراسة نظرية اختبارات الفروض.
والمقصود بمستوى المعنوية α هو "احتمال حدوث الخطأ من النوع الأول". أو نسبة حدوثه "أي احتمال رفض الفرض العدمي بينما هو صحيح".

وعادة ما يرمز إلى مستوى المعنوية بالرمز اللاتيني ألفا α وأشهر قيمتين لمستوى المعنوية α هما 5%، 1 %، ولكن ليس هناك ما يمنع من أن يأخذ قيما أخرى.

ومن الملاحظات المهمة هنا هو أن "مستوى المعنوية" والذي يسمى أحياناً "مستوى الدلالة" هو المكمل لدرجة الثقة، بمعنى أن مجموعهما يساوي 100% أو واحد صحيح. فإذا كانت درجة الثقة 95% فإن مستوى المعنوية يساوي 5%. والعكس صحيح فإذا كان مستوى المعنوية 5% فإن هذا يعني أن درجة الثقة 95 %.

ولعل من أهم الملاحظات هنا هو استخدام تعبير "مستوى المعنوية" في حالات اختبارات الفروض، بينما يستخدم مصطلح "درجة أو مستوى الثقة" في حالات التقدير.

ويمكننا ضبط أو تحديد احتمال ارتكاب خطأ من النوع الأول α. ولكن إذا خفضنا α فسوف نضطر إلى قبول احتمال أكبر لارتكاب خطأ من النوع الثاني β، اللهم إلا إذا رفعنا حجم العينة.

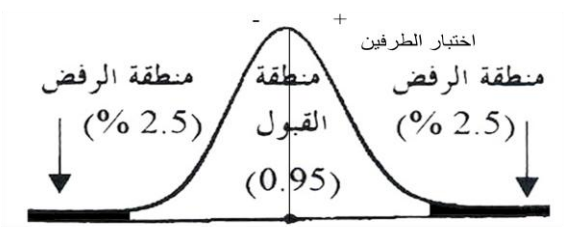
وتمسى α مستوى المعنوية، و $1-\alpha$ مستوى الثقة للاختبار.

والفكرة الأساسية في اختبار الفرض هي تقسيم المساحة تحت المنحنى إلى منطقتين: أحداهما تسمى "منطقة القبول" أي منطقة قبول الفرض العدمي. والأخرى تسمى "منطقة الرفض" أي منطقة رفض الفرض العدمي والتي تسمى أحيانا " بالمنطقة الحرجة Critical region ".

والنقطة الجديرة بالملاحظة هنا هي أن <u>منطقة القبول تمثل درجة الثقة</u>، بينما تمثل <u>منطقة الرفض مستوى المعنوية</u>.

هناك ثلاث حالات مختلفة لمنطقتي القبول والرفض (كما أوضحنا ذلك من قبل عند الحديث عن أنواع اختبارات الفروض) وهي:

الأولى: إذا كان الفرض البديل يأخذ شكل "لا يساوي" كأن يكون الفرض في هذه الحالة هو أن متوسط دخل الفرد لا يساوي 200 دولاراً فإن منطقة الرفض تكون موزعة على طرفي المنحنى بالتساوي، ويسمى الاختبار في هذه الحالة "اختبار الطرفين"، والذي يأخذ الشكل التالي (بافتراض أن):

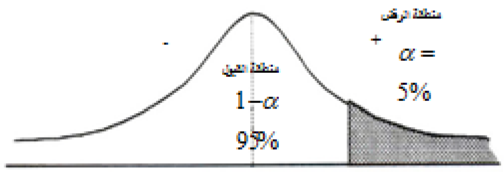


فالفرض الصفري هنا <u>يعني أن متوسط دخل الفرد يساوي 200 دولار شهريا</u>، والفرض البديل في هذه الحالة هو <u>بمعنى أن متوسط دخل الفرد لا يساوي 200 دولار شهرياً</u>.

حيث تمثل المنطقة البيضاء غير المظللة منطقة القبول والتي قد تساوي 95% وبالتالي فمنطقة الرفض مقسمة بالتساوي على طرفي المنحنى في هذه الحالة تكون قيمة كل منهما % 2.5.

والنتيجة هو أن القرار أيا كان نوعه سيكون بمستوى معنوية % 5 بمعنى أن احتمال أو نسبة الخطأ فيه من النوع الأول تساوي % 5.

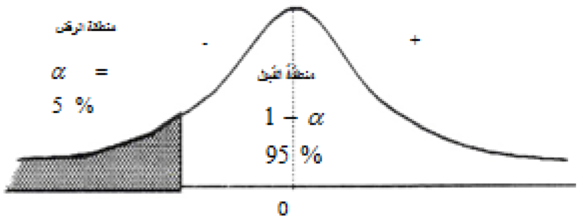
الثانية: إذا كان الفرض البديل يأخذ شكل "أكبر من" فإن منطقة الرفض تكون مركزة بالكامل في الطرف الأيمن للمنحنى. ويسمى الاختبار في هذه الحالة اختبار الطرف الأيمن. والذي يأخذ الشكل التالي:



اختبار الطرف الأيمن

فالفرض الصفري هنا نفس فرض المثال السابق، بينما الفرض البديل هو بمعنى أن متوسط دخل الفرد أكبر من 200 دولاراً شهرياً. وبالتالي فإن مستوى المعنوية والذي يساوي مثلاً 5% مركز في الطرف الأيمن من المنحنى.

الثالثة: إذا كان الفرض البديل يأخذ شكل "أقل من" فإن منطقة الرفض تكون مركزة بالكامل في الطرف الأيسر للمنحنى، ويسمى الاختبار في هذه الحالة اختبار الطرف الأيسر، والشكل التالي يوضح ذلك:



اختبار الطرف الأيسر

مع افتراض ثبات الفرض الصفري كما في المثال السابق، بينما الفرض البديل هو بمعنى أن متوسط دخل الفرد أقل من 200 دولار شهرياً، وبالتالي فإن مستوى المعنوية والذي يساوي مثلاً 5 % مركز في الطرف الأيسر من المنحنى.

خطوات إجراء الإختبار الإحصائي

الاختبار الاحصائى قد يكون متعلقا بعينة واحدة او عينتين او اكثر وقد يكون اختباراً معلمياً او غير معلمياً ويجب أن يمر الاختبار أياً كان نوعه بعدة خطوات وهذه الخطوات يمكن إيجازها في التالي:

1) صياغة الفرض الصفري H_o : والذي يأخذ – عادة – شكل " يساوي" فمثلاً إذا كان المطلوب هو اختبار ما إذا كان متوسط عمر الناخب هو 20 سنة فإن هذا الفرض يصاغ كما يلي:

$$Ho: \mu = 20$$

2) صياغة الفرض البديل H_1 : والذي يأخذ أحد أشكال ثلاثة إما :

"لا يساوي "

أو "أكبر من"

أو "أقل من"

وبالرموز فإن الفرض البديل قد يأخذ شكل أحد الصيغ التالية:

$$H1: \mu > 20 \qquad H1: \mu < 20 \qquad H1: \mu \neq 20$$

والذي يحدد شكل الفرض البديل هو مدى اقتناع الباحث بذلك أو مدى توفر المعلومات الأولية

فمثلاً: إذا كانت وجهة نظر الباحث أن متوسط عمر الناخب لا يمكن أن يقل عن 20 سنة فإنه يختار الفرض البديل "أكبر من" والعكس صحيح إذا كان يعتقد أن متوسط عمر الناخب لا يزيد عن 20 سنة فإنه يختار الفرض البديل "أقل من" أما إذا لم يكن لديه أي تصور أو أي معلومات فإنه يختار الفرض البديل "لا يساوي"

3) إحصائية الاختبار: وهي الإحصائية التي يتم حسابها من بيانات العينة بافتراض أن الفرض العدمي صحيح.
ويتوقف شكل الإحصائية على العوامل التالية:

أ - توزيع المجتمع، وهل هو طبيعي أم لا، وهل تباينه معروف أم لا.
ب - حجم العينة، وهل هو كبير أم صغير.
ج - الفرض العدمي المراد اختباره وهل هو عن الوسط أو النسبة أو التباين أو الارتباط... الخ.

والفكرة الأساسية (غالباً) في إحصائية الاختبار هي: حساب الفرق بين قيمة المعلمة التي نفترضها للمجتمع (في الفرض العدمي) والقيمة المقابلة لها في العينة أي التابع الإحصائي، ثم نقسم (أو ننسب) هذا الفرق إلى الخطأ المعياري للتابع الإحصائي.

فمثلاً: إذا كان الاختبار عن الوسط الحسابي فإنه يتم حساب الفرق بين قيمة الوسط الحسابي للمجتمع التي نفترضها وقيمة الوسط الحسابي للعينة، ثم نقسم هذا الفرق على الخطأ المعياري للوسط. وهكذا مع باقي الإحصائيات.

أ) بافتراض أن المجتمع الإحصائي المسحوبة منه العينة هو مجتمع طبيعي وانحرافه المعياري معروف، (أو) أن العينة كبيرة بدرجة كافية فإن إحصائية الاختبار والتي نرمز لها بالرمز $Z_{\bar{X}}$

الإحصائية في حالة اختبار الوسط للعينات الكبيرة: $$Z_{\bar{X}} = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

لاحظ أن البسط هو الفرق بين متوسطي المجتمع والعينة، والمقام هو الخطأ المعياري للوسط. ومن الناحية العملية فإن الانحراف المعياري للمجتمع عـادة ما يكون غير معروف ولكن طالما أن العينة كبيرة بدرجة كافية فإنه يمكن استخدام الانحراف المعياري للعينة S بدلا من الانحراف المعياري للمجتمع σ

ب) أما في حالة العينات الصغيرة وذلك عندما يكون المجتمع طبيعياً وانحرافه المعياري غير معروف فإن الإحصائية تأخذ الشكل التالي:

الإحصائية في حالة اختبار الوسط للعينات الصغيرة : $$T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$$

والتي لها توزيع t بدرجات حرية n - 1

4) تحديد منطقتي القبول والرفض: وذلك بناءً على الجداول الإحصائية والتي تعتمد على:

أ - توزيع المعاينة (وهل هو طبيعي أو t أو...)
ب - والفرض البديل (وهل هو لا يساوي أو أكبر من أو أقل من أي هل يستخدم اختبار الطرفين أو الطرف الأيمن أو الأيسر).
ج - ومستوى المعنوية (وهل هو 1% أو 5% أو غير ذلك).

5) المقارنة والقرار: بمعنى أن نقارن قيمة الإحصائية (المحسوبة من الخطوة الثالثة) بحدود منطقتي القبول والرفض (والتي حددناها في الخطوة الرابعة). فإذا وقعت قيمة الإحصائية داخل منطقة القبول فإن القرار هو: قبول الفرض الصفري. أما إذا وقعت قيمة الإحصائية في منطقة الرفض فإن القرار هو رفض الفرض الصفري، وفي هذه الحالة نقبل الفرض البديل. مع ملاحظة أن القرار مرتبط بمستوى المعنوية المحدد. بمعنى أن القرار قد يتغير إذا تغير مستوى المعنوية المستخدم (وفي بعض الحالات قد لا يتغير القرار، فهذا يتوقف على قيمة الإحصائية وما إذا كانت تقع في منطقة القبول أو منطقة الرفض).

أي أنه توجد طريقتين لاتخاذ القرار في الاختبارات الاحصائية وهي:

(i) حساب احصاء الاختبار ومقارنته بقيمة جدوليه وتحدد القيمة الجدوليه بناء على نوع الاحتبار ذو طرف واحد One Tail Test أو ذو طرفين Two Tail Test

(ii) حساب ما يسمى بالقيمة الاحتماليه P-value ويرمز لها فى بعض البرامج الإحصائية بالرمز .Sig فاذا كان الاختبار ذو طرف واحد تقارن .Sig بالقيمة α لكن اذا كان الاختبار ذو طرفين تقارن بالقيمة α/2

مثال:

عينة عشوائية حجمها 49 شخصاً اختيرت من أفراد دولة ما، فإذا كان الوسط الحسابي لدخول الأفراد الأسبوعية في العينة هو 75 دولاراً. كيف يمكن اختبار الفرض الصفري بأن متوسط الدخل الأسبوعي لمواطني هذه الدولة يساوي 72 دولاراً مقابل الفرض البديل أنه لا يساوي 72 وذلك بمستوى معنوية 5 % إذا علمت أن الانحراف المعياري لدخول الأفراد يساوي 14 دولاراً.

الحل:

١. الفرض العدمي: هو أن متوسط المجتمع يساوي 72 وبالرموز: $Ho: \mu = 72$

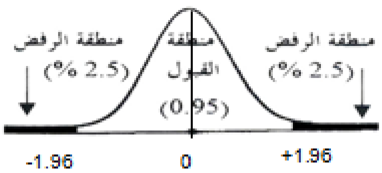
٢. الفرض البديل: هو أن المتوسط لا يساوي 72 وبالرموز: $HI: \mu \neq 72$

٣. الإحصائية: بما أن العينة كبيرة فإن الإحصائية في حالة اختبار الوسط تأخذ الشكل التالي: $Z_{\bar{X}} = \dfrac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$

حيث $n = 49, \sigma = 14, \bar{X} = 75, \mu = 72$ وبالتعويض نحصل على: $Z_{\bar{X}} = \dfrac{75-72}{\dfrac{14}{\sqrt{49}}} = \dfrac{3}{\dfrac{4}{7}} = \dfrac{3}{2} = 1.5$

أي أن قيمة الإحصائية تساوي 1.5

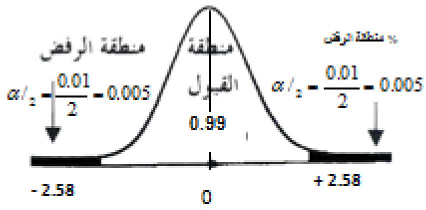
٤. حدود منطقتي القبول والرفض: نحصل عليها من التوزيع الطبيعي المعياري حيث مستوى المعنوية 5% وبما أن الفرض البديل هو: "لا يساوي" فإن ما يستخدم في هذه الحالة هو اختبار الطرفين كما في الشكل التالي :



وقد حصلنا على حدود منطقتي القبول والرفض وذلك بقسمة درجة الثقة (المكملة لمستوى المعنوية) والتي تساوي 0.95 على 2 فنحصل على 0.4750 وبالكشف في جدول التوزيع الطبيعي المعياري عن Z التي تقابل المساحة 0.4750 نجد أنها تساوي 1.96 وحيث أنها موزعة على طرفي المنحنى بالتساوي فنضع إشارة موجبة في النصف الأيمن، وإشارة سالبة في النصف الأيسر، أي أن منطقة القبول تبدأ من القيمة 1.96- وتستمر حتى القيمة 1.96 + (أي أن أي قيمة محصورة بين هاتين القيمتين تكون في منطقة القبول، وأي قيمة خارج هذه الحدود تكون في منطقة الرفض).

٥. المقارنة والقرار: وبمقارنة قيمة الإحصائية المحسوبة من الخطوة رقم 3 (والتي تساوي 1.5) بحدود منطقتي القبول والرفض (من الخطوة رقم 4) نجد أنها تقع في منطقة القبول لذلك فإن القرار هو:
قبول الفرض الصفري بأن متوسط دخول الأفراد الأسبوعية في هذه الدولة يساوي 72 دولاراً وذلك بمستوى معنوية 5 %.

لو استخدمنا مستوى معنوية 1% بدلاً من 5% كما في المثال أعلاه فإن حدود منطقتي القبول والرفض تصبح كما يلي:



وبمقارنة قيمة الإحصائية 1.5 بحدود منطقتي القبول والرفض نجد أنها تقع في منطقة القبول أي أن القرار هو نفسه قبول الفرض الصفري ولن يتغير بل يتأكد باستخدام مستوى معنوية 1%.

مثال:

أفترض أن شركة ترغب في اختبار ما إذا كان يمكنها الادعاء بأن متوسط عمر المصباح من إنتاجها هو 1000 ساعة احتراق. وأنها قامت بأخذ عينة عشوائية حجمها $n = 100$ من إنتاجها فوجدت أن متوسط العينة $X = 980$ ساعة والانحراف المعياري للعينة $s = 80$ ساعة.

فإذا أرادت الشركة القيام بالاختبار عند مستوى معنوية 5%، فعليها القيام بالتالي:

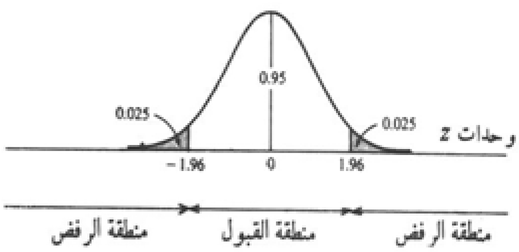
الحل:

حيث أن μ يمكن أن تساوي أو تزيد عن، أو تقل عن 1,000، فان الشركة يجب أن تضع الفرض الصفري والفرض البديل كالآتي:

$H_1 : \mu \neq 1,000$ $H_0 : \mu = 1,000$

وحيث أن $n > 30$ ، فإن توزيع المعاينة للوسط يكون تقريباً طبيعياً (ويمكن استخدام s كتقدير بدلاً من σ) . وتكون منطقة القبول للاختبار عند مستوى المعنوية 5% بين 1.96 تحت التوزيع الطبيعي القياسي وحيث أن منطقة الرفض تقع عند ذيل التوزيع، فإن الاختبار يسمى اختبار ذو ذيلين. وتكون الخطوة الثالثة إيجاد القيمة المناظرة لقيمة X :

$$z = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{\overline{X} - \mu_0}{80 / \sqrt{100}} = \frac{-20}{8} = -2.5$$



وحيث أن قيمة z المحسوبة تقع داخل منطقة الرفض، فإن على الشركة أن ترفض الفرضية الصفرية (H_0) أي أن $\mu = 1,000$ وتقبل الفرضية البديلة (H_1) أي $\mu \neq 1,000$ وذلك عند مستوى معنوية 5%.

مثال:

ترغب شركة أن تعرف بدرجة ثقة 95% ما إذا كان يمكنها الادعاء بأن صناديق الصابون المسحوق الذي تبيعها تحتوي على اكثر من 500 جرام (حوالي 1.1 رطل) من الصابون. وتعرف الشركة من الخبرة الماضية أن أوزان الصابون بالصناديق تتبع التوزيع الطبيعي. وقد أخذت الشركة عينة عشوائية حجمها $n = 25$ ووجدت أن $X = 520$ جرام و $s = 75$ جرام

الحل:

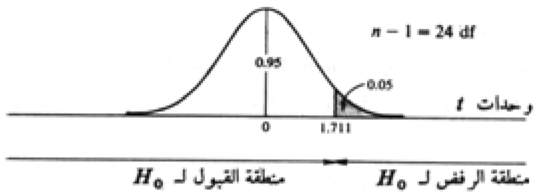
وحيث أن الشركة ترغب في اختبار ما إذا كانت $\mu > 500$، فإن:

$H_0 : \mu = 500 \qquad H_1 : \mu > 500$

وحيث أن التوزيع طبيعي ، $n < 30$، وكذلك σ غير معلومة، فعلينا أن نستخدم توزيع t (بدرجة حرية $n - 1 = 24$) لتحديد المنطقة الحرجة، أي منطقة الرفض، للاختبار بمستوى معنوية 5%. ونجد ذلك في الجدول المخصص لاختبار t ويعرضها الشكل التالي، ويسمى هذا اختبار الذيل الأيمن. وأخيراً ، حيث أن

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}} = \frac{520 - 500}{75/\sqrt{25}} = \frac{20}{15} = 1.33$$

وهي تقع داخل منطقة القبول، وتقبل H_0 أى $\mu = 500$ ، عند مستوى معنوية 5% (أو بدرجة ثقة 95%).



المحاضرة 11 (الجزء 1)

اختبار الفروض الإحصائية المعلمية

يهدف الباحث من اختبار الفرضيات حول المتوسط إلى اتخاذ قرار حول ما إذا كانت هذه الفرضية مقبولة أم مرفوضة، ويتم ذلك من خلال استخدام مختبر إحصائي مناسب، **والمختبر الإحصائي** هو متغير عشوائي ذو توزيع احتمالي يصف العلاقة بين القيم النظرية للمعلم والقيم المحسوبة من العينة، وفي العادة تقارن قيمة المختبر الإحصائي المحسوب من العينة مع قيمته المستخرجة من توزيعه الاحتمالي (باستخدام جداول خاصة) ومنها نتخذ القرار برفض أو قبول الفرضية الصفرية .

أوجه الشبه والإختلاف بين توزيع t والتوزيع المعتدل المعياري:

تحدثنا فيما سبق عن التوزيع الطبيعي المعياري وتوزيع t وذكرنا بأن توزيع t يعتمد على افتراض أن العينة مسحوبة من مجتمع توزيعه معتدل أو قريب من الإعتدال، لذلك يستخدم اختبار t لاختبار الفروض وإيجاد حدود الثقة لمتوسط المجتمع أو الفرق بين متوسطين عندما تكون احجام العينات صغيرة وتباينات المجتمعات مجهولة.

- ويشبه توزيع t التوزيع المعتدل المعياري في أنه مستمر، ناقوسي الشكل، متماثل حول الصفر.
- ويختلف عنه في أن تشتته أكبر من تشتت التوزيع المعتدل المعياري، حيث نجد أن طرفي منحنى t أكثر بعدا عن المحور الأفقي، كما أنه أكثر تسطحا في المنتصف.

وذكرنا كذلك بأن منحنى t يقترب من المنحنى المعتدل المعياري كلما زاد حجم العينة n حتى ينطبق عليه عندما تصل n إلى مالا نهاية، وهذا يعني أن الإنحراف المعياري للعينة s يقترب من الإنحراف المعياري للمجتمع σ كلما زاد حجم العينة، وعندما يزداد اقتراب n من ما لا نهاية فإن s تزداد اقترابا من σ حتى يتلاشى أي فرق بين قيم Z وقيم t .

وباختصار نجد أن متوسط توزيع t يساوي الصفر وانحرافه المعياري أكبر من الواحد الصحيح، ويقترب الإنحراف المعياري من الواحد الصحيح كلما اقترب حجم العينة من ما لا نهاية، وهذا يعني وجود عدد لانهائي من توزيعات t لها نفس الوسط الحسابي 0 ولكن إنحرافاتها المعيارية تختلف باختلاف حجم العينة n ، تذكر أن Z لها توزيع واحد فقط متوسطه 0 وانحرافه المعياري 1 .

استخدام جدول t

يعتمد استخدام جدول t على مفهوم درجات الحرية df ، ودرجات الحرية (v) وتنطق (نيو) تساوي (n-1) وتعتبر درجات الحرية المعلمة الوحيدة لتوزيع t حيث يعتمد شكل منحنى t اعتمادا كاملا عليها، أي أن توزيع t يعتمد اعتمادا كاملا على حجم العينة n .

ولغرض استخدام جدول t نجد أن الصف العلوي من الأول يحوي Q على مساحة الطرف الأيمن (أو مساحة الطرف الأيسر) لتوزيع t بدرجات حرية محددة (v) ، بينما يحتوي الصف العلوي الثاني على 2Q على مساحة الطرفين الأيمن والأيسر لتوزيع t بدرجات حرية محددة (v) ، حيث يستخدم هذا الصف عندما يكون الاختبار ذو طرفين، بينما يستخدم الصف الأول عندما يكون الاختبار ذو طرف واحد فقط.

ويحتوي العمود الأول بأقصى يسار الجدول على قيم لدرجات الحرية (v) ، ويمثل الرقم الموجود أما صف معين وتحت احتمال محدد قيمة t الحرجة التي تحدد منطقة الرفض لتوزيع t بدرجات حرية (v) ، ونتيجة لتماثل توزيع t حول الصفر، فإن جدول توزيع t يحتوي على القيم الموجبة فقط.

مثال:

افترض أن مستوى المعنوية في مشكلة معينة يساوي 0.05 ، وأن حجم العينة يساوي 20 ، أوجد قيمة T الحرجة التي تناظر التالي:

١. اختبار ذو طرف أيمن.
٢. اختبار ذو طرف أيسر.
٣. اختبار ذو طرفين.

الحل:

1- عندما تكون α = 0.05 و v = (1-20) = 19 ، نجد أن القيمة الموجودة أمام الصف 19 وتحت الإحتمال 0.05 الموجود في الصف العلوي الأول من الجدول هو 1.729 ، أي أن قيمة t الحرجة بدرجات حرية 19 ومستوى معنوية 0.05 هو 1.729 (لاختبار ذو طرف أيمن)، ويبين الجدول التالي جزء مستقطع من جدول t :

Q **V**		**0.05**	
19		1.729	

2- ونتيجة لأن توزيع t متماثل حول الصفر، فإن قيمة t الحرجة بدرجات حرية 19 والتي تكون المساحة إلى يسارها مساوية 0.05 هي 1.729- ، وهذا في حالة الاختبار ذو الطرف الأيسر.

3- إذا كان الاختبار ذو طرفين فإن قيمة α هي قيمة 2Q الموجودة في الصف العلوي الثاني من جدول t ، وبالنظر إلى الجدول نجد أن القيمة الحرجة لـ t = 2.093 وهي القيمة الموجودة أمام الصف 19 وتحت العمود 0.05 في صف 2Q العلوي الثاني، ويبين الجدول التالي جزء مستقطع من جدول t :

2Q **V**		**0.05**	
19		2.093	

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
					Confidence Level						

أنواع الاختبار (الفروض)

أنواع الاختبار (الفروض) فى حالة عينه واحده

بفرض اننا سوف نرمز للمعلم المجهول بالرمز μ ونريد اختبار الفرض القائل ان قيمته تساوى μ_0 سيكون فرض العدم على الصورة التالية: $H_0 : \mu = \mu_0$

وسيكون الفرض البديل فى حالة الاختبار ذو طرف واحد $H_a : \mu < \mu_0$ or $H_a : \mu > \mu_0$

وسيكون الفرض البديل فى حالة الاختبار ذو طرفين $H_a : \mu \neq \mu_0$

أنواع الاختبار (الفروض) فى حالة عينتين

بفرض اننا سوف نرمز للمعلم المجهول بالرمز μ ونريد اختبار الفرض القائل ان قيمته متساويه فى المجتمعين سيكون فرض العدم على الصورة التالية: $H_0 : \mu_1 = \mu_2$

وسيكون الفرض البديل فى حالة الاختبار ذو طرف واحد $H_a : \mu_1 < \mu_2$ or $H_a : \mu_1 > \mu_2$

وسيكون الفرض البديل فى حالة الاختبار ذو طرفين $H_a : \mu_1 \neq \mu_2$

أنواع الاختبار (الفروض) فى حالة أكثر من عينتين

بفرض اننا سوف نرمز للمعلم المجهول بالرمز μ ونريد اختبار الفرض القائل ان قيمته متساويه فى المجتمعات التى عددها r سيكون فرض العدم على الصورة التالية $H_0 : \mu_1 = \mu_2 = \cdots = \mu_r$

وسيكون الفرض البديل $H_a : \text{at least two are different}$

فى جميع الاختبارات يمكن قياس قوة الاختبار بما يسمى بالخطأ من النوع الثانى Type II Error والذى يستخدم بدوره فى حساب دالة القوه Power Function

الاختبارات الاحصائية لعينة واحدة One Sample Test

اختبار Z-test :

في كثير من الأحيان لا يمكن معرفة تباين المجتمع الذي سحبت منه العينة ، إلا أنه إذا كان حجم العينة كبيرا (n > 30) فإنه يمكن استخدام تباين العينة الكبيرة (S^2) عوضاً عن تباين المجتمع (σ^2) الغير معلوم، وذلك لأن (S^2) مقدر جيد (σ^2) ولأنه لا يتغير كثيراً من عينة لأخرى ما دام حجم العينة كبير، ففي هذه الحالة يمكننا استخدام اختبار (Z) لاختبار الفرضيات الصفرية موضع الدراسة وذلك من خلال المختبر الإحصائي التالي: $Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$

ويعتبر ذلك مدخل ضروري لفهم اختبار t-test .

خطوات اختبار Z

- وضع فرض العدم والفرض البديل.

مثال: ينتج مصنع دقيق قمح في عبوات زنة العبوة (2.5) كيلوجرام. فإن فرضية العدم هي: $H_0 : \mu = 2.5$

• في حين يأخذ الفرض البديل عدة أشكال حسب طبيعة الاختبار:

$H_1 : \mu \neq 2.5$	يستهدف الاختبار النظر فيما إذا كان المتوسط لا يساوي متوسط المجتمع بغض النظر عن كون الاختلاف زيادة أو نقصاً
$H_1 : \mu > 2.5$	يستهدف الاختبار النظر فيما إذا كان المتوسط اكبر من متوسط المجتمع
$H_1 : \mu < 2.5$	يستهدف الاختبار النظر فيما إذا كان المتوسط اقل من متوسط المجتمع

• تحديد مستوى الدلالة (α): وتحدد مستويات المعنوية سلفاً وهي عادة 0.05 أو 5%

• حساب إحصائية الاختبار (Z) حيث: $Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$

$\overline{X}$ الوسط الحسابي للعينة

μ_0 القيمة الفرضية للوسط الحسابي للمجتمع

α مستوى الدلالة

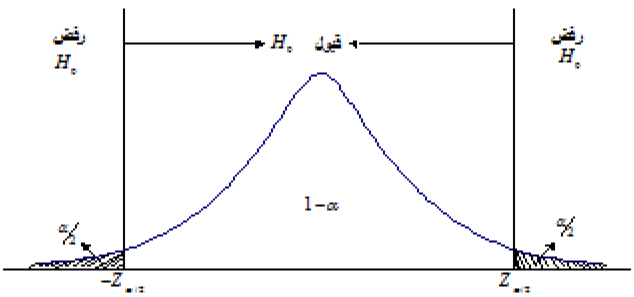
σ الانحراف المعياري للمجتمع

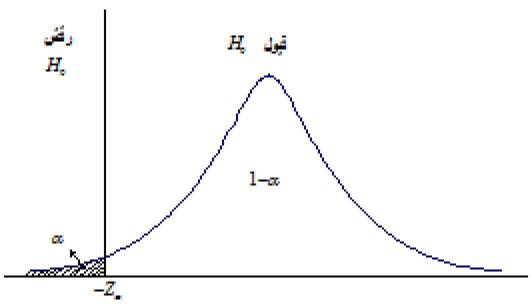
n حجم العينة

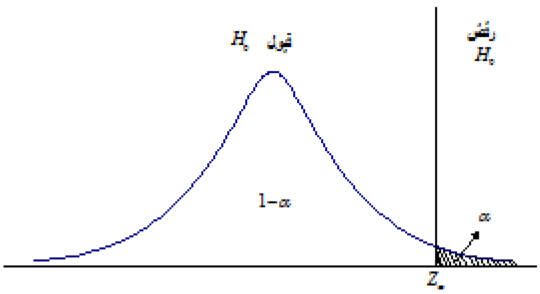
• اتخاذ قرار حول بيانات العينة من خلال مقارنة قيمة إحصائية الاختبار بالقيمة النظرية للتوزيع الطبيعي المعياري عند مستوى معنوية محدد (Z_α). وفيما يلي قاعدة القرار لرفض فرض العدم :

الفرض البديل	**قاعدة القرار: رفض فرض العدم إذا**	
$H_1 : \mu \neq \mu_0$	$\lvert Z \rvert > Z_{\alpha/2}$	إذا كانت القيمة المطلقة لـ Z أكبر من قيمة Z النظرية عند مستوي معنوية (α/2)
$H_1 : \mu > \mu_0$	$Z > Z_\alpha$	إذا كانت قيمة Z أكبر من قيمة Z النظرية عند مستوي معنوية α
$H_1 : \mu < \mu_0$	$Z < -Z_\alpha$	إذا كانت قيمة Z أقل من قيمة Z النظرية عند مستوي معنوية α

وهذه القاعدة توضحها الأشكال التالية:







مثال على اختبار Z :

إذا كان متوسط استهلاك الفرد السعودي من الدجاج حسب تقارير وزارة الصحة هو (12) كيلوجرام بانحراف معياري (6) كيلوجرامات لفترة السبعينات الميلادية. أجرى أحد الباحثين دراسة في عام 2003م من عينة قوامها (49) فرداً ووجد أن متوسط الاستهلاك للفرد هو (14) كيلوجرام. هل تشير الدراسة الحالية أن متوسط الاستهلاك أرتفع عما عليه في السبعينات.

الحل:

١. فرض العدم والفرض البديل.

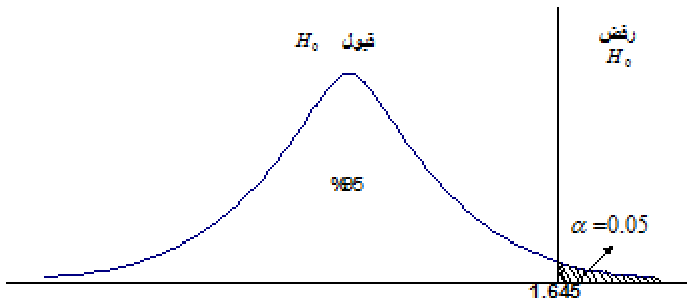
فرض العدم: $H_o: \mu=12$

الفرض البديل: $H_1: \mu>12$

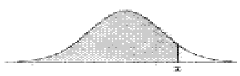
٢. مستوى الدلالة $\alpha = (0.05)$

٣. إحصائية الاختبار (Z): $Z = \dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} = \dfrac{14-12}{6/\sqrt{49}} = 2.33$

٤. تحديد قيمة Z المعيارية من الجدول عند مستوى دلالة (0.05)، نحتاج لتحديد قيمة Zα التي تقع علي اليمين وتساوي 1.645 (أنظر الشكل التالي):



Tables of the Normal Distribution



Probability Content from -oo to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

٥. بما أن القيمة المحسوبة أكبر من القيمة النظرية المستخرجة من الجدول كما يبين الشكل، فإنها تقع في منطقة الرفض. وبذلك نرفض فرض العدم حيث أن البيانات المتوفرة تقدم دليلاً كافياً على أن متوسط استهلاك الفرد من لحوم الدواجن في الوقت الحالي قد ارتفع بمستوى معنوي أو ذو دلالة عما علية في سبعينات القرن الماضي.

حساب اختبار (ت) لعينة واحدة One Sample T-Test من خلال الـ SPSS

اختبار t-test :

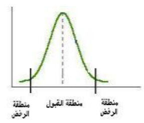
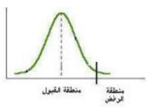
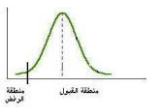
ولكن إذا كان حجم العينة صغيراً (n < 30) فإن قيمة (S^2) تتغير كثيراً من عينة إلى أخرى وبالتالي لا يمكننا هنا أن نستخدم اختبـار (Z)، مما دفع كثيراً من الإحصائيين للبحث عن البديل المناسب.

ففي سنة 1908م استطاع العالم الايرلندي وليم كوسيت W.S. Gosset من نشر بحث تحت اسم مستعار بسبب ظروف خاصة هو (استيودنت، Student) استطاع من خلاله أن يشتق معادلة للتوزيع الاحتمالي (t) الذي قيمته هي: $t = \dfrac{\overline{X} - \mu}{S/\sqrt{n}}$

وهذا الاختبار يشبه اختبار التوزيع الطبيعي (Z)، بيد أنه يختلف عنه في تضمنه للانحراف المعياري (S) للعينة بدلاً من الانحراف المعيـاري (σ) للمجتمع.

يستخدم هذا الصنف من اختبار (ت) للحكم على معنوية الفروق بين متوسط عينة ومتوسط المجتمع الذي سحبت منه. ولغرض توضيح ذلك إليك هذا العرض الموجز لخطوات اختبار (ت) حول متوسط حسابي واحد على افتراض أن تباين المجتمع σ^2 غير معلوم وأن حجم العينة صغير .

والجدول التالي يوضح خطوات اختبار حول متوسط حسابي واحد باستخدام المختبر الإحصائي (ت) t-test على افتراض أن تباين المجتمع σ^2 غير معلوم وأن حجم العينة صغير

خطوات الاختبار	اختبار ذو طرفين	اختبار ذو طرف واحد	
		طرف يمين	طرف يسار
1- الفرضية الصفرية H_0 2- الفرضية البديلة H_1	$\mu = \mu_0$ $\mu \neq \mu_0$	$\mu = \mu_0$ $\mu > \mu_0$	$\mu = \mu_0$ $\mu < \mu_0$
3- مستوى الدلالة	α		
4- منطقة الرفض	ا ت ا ≥ ت($\frac{1}{2}$ α،df) 	ت ≥ ت(α،df) 	ت ≤ ت(α،df) 
5- المختبر الإحصائي	$t = \dfrac{\overline{X} - \mu}{S/\sqrt{n}}$ هذا المختبر الإحصائي يستخدم لتوضيح أهمية الفروق بين الوسط الحسابي للعينة والوسط الحسابي للمجتمع		
6- القرار	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [ت ($\frac{1}{2}$ α،df)] الجدولية أي أن : ا ت ا ≥ ت($\frac{1}{2}$ α،df)	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [ت (α ،df)] الجدولية أي أن: ت ≥ ت(α،df)	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [-ت (α،df)] الجدولية أي أن: ت ≤ - ت(α،df)

مثال على اختبار t:

لو كانت لدينا عينة عشوائية تتكون من 250 طالب وجد أن الوسط الحسابي لأطوال طلاب العينة 155.95سم، والانحراف المعياري = 2.94 سم، علما بأن الوسط الحسابي لأطوال طلاب الجامعة يبلغ 158 سم، اختبر أهمية الفرق المعنوي بين الوسط الحسابي لأطوال طلاب العينة والوسط الحسابي لأطوال طلاب الجامعة .

الحل:

سيتم اختبار الفرضيات التالية:

الفرضية الصفرية: لا توجد فروق ذات دلالة إحصائية بين متوسط أطوال الطلاب في العينة ومتوسط أطوال الطلاب في الجامعة

($\mu = \mu_0$)

الفرضية البديلة: توجد فروق ذات دلالة إحصائية بين متوسط أطوال الطلاب في العينة ومتوسط أطوال الطلاب في الجامعة

($\mu \neq \mu_0$)

مستوى الدلالة: α = 0.05

منطقة الرفض: قيمة (ت) الجدولية عند مستوى دلالة α 0.05 ودرجات حرية 249 = 1.960

المختبر الإحصائي: $\overline{X}$ = 155.95سم ، n = 250 طالب ، S = 2.94 سم ، μ = 158سم

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{155.95 - 158}{2.94/\sqrt{250}} = -11.006$$

القرار:

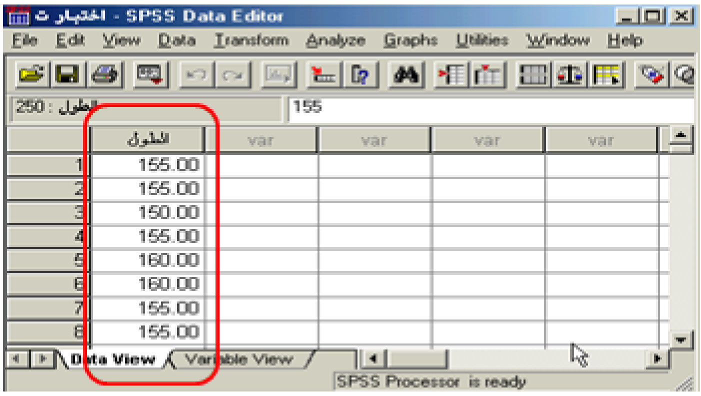
قيمة ت المحسوبة (-11.006) أكبر من قيمة ت المجدولة (1.96) عند مستوى دلالة α = 0.05 .

∴ نرفض الفرضية الصفية ونقبل البديلة .

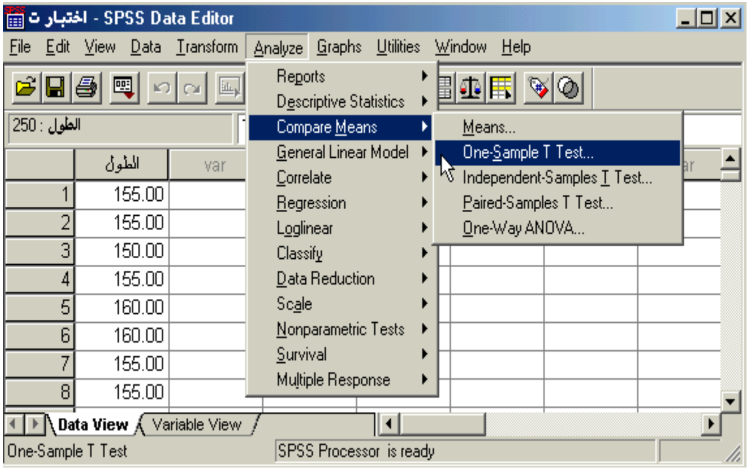
أي أنه توجد فروق ذات دلالة إحصائية بين الوسط الحسابي للعينة والوسط الحسابي لمجتمع البحث .

لغرض حساب قيمة (ت) لنفس المثال السابق من خلال استخدام برنامج الـ SPSS اتبع الخطوات التالية:

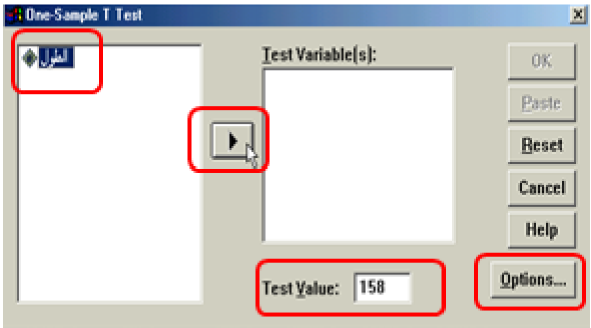
- قم بإدخال البيانات المراد تحليلها من خلال شاشة **تحرير البيانات Data Editor** بالطريقة المناسبة كالتالي:



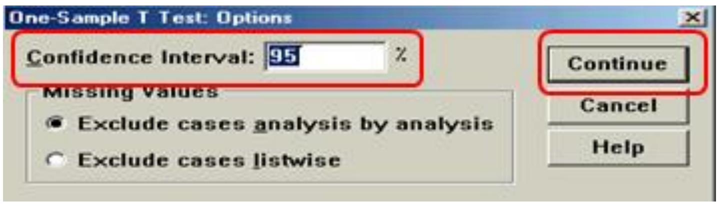
- من القائمة "تحليل" Analyze اختر الأمر "مقارنة المتوسطات" Compare Means فتظهر قائمة أوامر فرعية اختر منها "اختبار (ت) لعينة واحدة" One-Sample T Test كالتالي:



- بعد اختيار الأمر "اختبار (ت) لعينة واحدة" One-Sample T Test سوف يظهر لك صندوق الحوار التالي:



- من قائمة المتغيرات في الجهة اليسرى من صندوق الحوار انقر نقرا مزدوجا على المتغير "الطول" (أو أنقر على السهم الذي يظهر في صندوق الحوار بعد التظليل على المتغير المرغوب نقله إلى الجهة الأخرى) ستلاحظ انتقاله مباشرة في المستطيل "متغيرات الاختبار" Test Variable(s).
- في الحقل الخاص بـ "القيمة المختبرة" Test Value أكتب القيمة التي تريد أن تقارن بها متوسط العينة موضع الدراسة (في هذا المثال يتم كتابة الرقم 158 والذي يمثل متوسط أطوال الطلاب في الجامعة) .
- قم بالنقر على زر "خيارات" Options في الجهة السفلية اليمنى من صندوق الحوار السابق وذلك عند الرغبة في تغيير قيمة "فترة الثقة" Confidence Interval حيث يظهر لك صندوق الحوار التالي والذي يتيح إمكانية تغيير فترة الثقة المختبرة (بشكل تلقائي سوف تظهر القيمة 95%) ، وبعد الانتهاء من التعديل على هذا الصندوق الحواري أنقر على زر "استمرار" Continue



- أنقر بعد ذلك على زر "موافق" OK سيؤدي ذلك إلى تنفيذ الاختبار، وستلاحظ ظهور النتائج في شاشة المخرجات كالتالي:

➡ T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
الطول	250	155.9520	2.9422	.1861

One-Sample Test

	Test Value = 158					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
الطول	-11.006	249	.000	-2.0480	-2.4145	-1.6815

يتضح من النتائج أن قيمة (ت) المحسوبة t-test = 11.006- ، ودرجات الحرية df = 249 ، وقيمة Sig. (2-tailed) =0.000 ، وبما أن قيمة Sig. (2-tailed) في الجدول (0.000) أصغر من قيمة α = 0.05 فإننا بالتالي نرفض الفرضية الصفرية، أي أنه توجد فروق ذات دلالة إحصائية بين متوسط أطوال العينة ومتوسط أطوال طلاب الجامعة.

والأوامر المستخدمة لحساب اختبار (ت) لعينة واحدة One Sample T-Test من خلال برنامج الـ SPSS (مع اختلاف المتغيرات موضع الدراسة) هي:

```
T-TEST
/TESTVAL=158
/MISSING=ANALYSIS
/VARIABLES=الطول
/CRITERIA=CIN (.95).
```

المحاضرة 11 (الجزء 2)

تابع اختبار الفروض الإحصائية المعلمية

الاختبارات الاحصائية لعينتين مستقلتين Independent Samples t-test

يستخدم هذا النوع من اختبار (ت) للحكم على معنوية Significance الفروق بين متوسطي عينتين غير مرتبطتين Independent. ولغرض توضيح ذلك إليك هذا العرض الموجز لخطوات اختبار (ت) حول متوسطين على افتراض أن تباين المجتمع σ_1^2 و σ_2^2 غير معلومين ولكنهما متجانسين وأن حجم العينة صغير.

والجدول التالي يوضح خطوات اختبار الفرق بين متوسطين باستخدام المختبر الإحصائي (ت) على افتراض أن تباين المجتمع σ_1^2 و σ_2^2 غير معلومين ولكنهما متجانسين ، وأن كل من n_1 و n_2 صغير

ويستند هذا الاختبار إلى توفر عدد من الافتراضات وهذه الافتراضات هي :

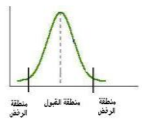
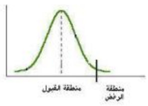
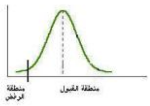
- **مستوى القياس:** يشترط لاستخدام هذا الاختبار أن تكون البيانات فئوية (فترية) أو نسبية.
- **أن يكون حجم العينة صغيراً:** يقتضي هذا الافتراض أن يكون حجم العينة أقل من (30) وأكبر من (5) (وإذا كان حجم العينة أكبر من 30 فلا بأس من ذلك) ، بحيث أن حجم ( ) لا يسمح بتقدير جدي لـ (σ) ، لذا ولتجنب الخطأ نستعيض عن (σ) بالمعلمة (S) . ويجب كذلك أن يكون الفرق بين حجم عينتي البحث صغيراً ، لأنه كلما زاد الفرق بين حجم العينتين أثر ذلك على قيمة (t) المحسوبة
- **التوزيع الطبيعي:** ويقضي هذا الافتراض أن المشاهدات ($س_1$) في المجتمع الأول تتخذ شكل التوزيع الطبيعي لوسط يساوي (μ_1) وكذلك الأمر بالنسبة للمشاهدات ($س_2$) في المجتمع الثاني يفترض فيها أن تتخذ شكل التوزيع الطبيعي لوسط يساوي (μ_2) ، وإن مخالفة هذا الافتراض ليس لها تبعات تذكر .
- **تجانس التباين في المجتمعين:** وبموجب هذا الافتراض يكون لتباين المشاهدات في كل من المجتمعين نفس القيمة (σ^2) وبذلك تكون القيمة المتوقعة للتباين في كل العينتين مساوية للمقدار (σ^2) أي يكون كل من (S_1^2 ، S_2^2) تقديراً مستقلاً لنفس المقدار (σ^2) وإن هناك اتفاقاً بين الإحصائيين على أن افتراض تجانس التباين هذا يمكن مخالفته إذا تساوت العينات في أعداد أفرادها، أي إذا كانت ($n_1 = n_2$).

أما عندما تكون ($n_1 \neq n_2$) فإن هناك وضعين هما :

١. عندما تكون العينة الكبيرة الحجم منتمية للمجتمع ذي التباين الكبير، والعينة صغيرة الحجم منتمية للمجتمع ذي التباين الصغير، فإن احتمال ارتكاب الخطأ من النوع الأول تكون أقل من (α) أي لا يصل الأمر إلى رفض الفرضية الصفية وهي صحيحة وذلك بسبب مخالفة هذا الافتراض. ومن ذلك يكون الباحث في هذه الحالة في الجانب الأمين، أي إن مخالفة تجانس التباين لا تؤثر على نتائج البحث.
٢. أما عندما تكون العينة كبيرة الحجم منتمية للمجتمع ذي التباين الصغير والعينة صغيرة الحجم منتمية للمجتمع ذي التباين الكبير، فإن احتمال ارتكاب الخطأ من النوع الأول يزداد (رفض الفرضية الصفرية وهي صحيحة) ، وفي هذه الحالة ليس هناك حاجة للقلق من مخالفة هذا الافتراض وذلك عندما تكون الفرضية الصفرية مقبولة، ولكن يجب أن يساورنا القلق إذا رفضناها ، وذلك لازدياد احتمال رفضها وهي صحيحة (أي ازدياد احتمال ارتكاب الخطأ من النوع الأول)، ولتفادي ذلك نستخدم اختبار ولتيش Weltch بالعلاقة : $t = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$

- **الاستقلالية:** ويقتضي هذا الافتراض أن (n_1) من المشاهدات قد تم الحصول عليها عشوائيا من المجتمع الأول بشكل مستقل عن (n_2) من المشاهدات والتي تم الحصول عليها عشوائياً من المجتمع الثاني. وإن الاستقلالية هنا لا تعني استقلالية البيانات بين المجتمعين فقط ، بل تعني استقلالية المشاهدات ضمن المجتمع الواحد أيضاً (مثل عملية تطبيق اختبار قبلي واختبار بعدي على مجموعة واحدة) .

والجدول التالي يوضح خطوات اختبار الفرق بين متوسطين باستخدام المختبر الإحصائي (ت) على افتراض أن تباين المجتمع σ_1^2 و σ_2^2 غير معلومين ولكنهما متجانسين ، وأن كل من n_1 و n_2 صغير

خطوات الاختبار	اختبار ذو طرفين	اختبار ذو طرف واحد	
		طرف يمين	طرف يسار
1- الفرضية الصفرية H_0 2- الفرضية البديلة H_1	$\mu = \mu_0$ $\mu \neq \mu_0$	$\mu = \mu_0$ $\mu > \mu_0$	$\mu = \mu_0$ $\mu < \mu_0$
3- مستوى الدلالة	α		
4- منطقة الرفض	ا ت ا ≤ ت($\frac{1}{2}$ α،df) 	ت ≤ ت(α،df) 	ت ≥ ت(α،df) 
5- المختبر الإحصائي	$$t = \frac{\overline{X_1} - \overline{X_2}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ هذا المختبر الإحصائي يستخدم لتوضيح أهمية الفروق بين متوسطين للعينات المستقلة وعند تجانس التباين		
6- القرار	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [ت ($\frac{1}{2}$ α،df)] الجدولية أي أن : ا ت ا ≤ ت($\frac{1}{2}$ α،df)	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [ت (α ،df)] الجدولية أي أن: ت ≤ ت(α،df)	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [-ت (α،df)] الجدولية أي أن: ت ≥ - ت(α،df)

ولتوضيح ما ورد في الجدول السابق دعنا نتناول هذا المثال:

أراد باحث أن يعرف أثر استخدام نظم مساندة القرارات على كفاءة القرارات التي تتخذها الإدارة بمساعدة تلك النظم، فوزع 50 مديرا لمنشآت صناعية عشوائيا في مجموعتين، ثم عين أحدهما بطريقة عشوائية لتكون مجموعة تجريبية والأخرى ضابطة، وفي نهاية التجربة وزع على المجموعتان استقصاء يقيس درجة فاعلية القرار وكفاءته عندما يتم اتخاذه باستخدام نظم مساندة القرارات بدلا من الطريقة التقليدية فكانت النتائج كما يلي:

المجموعة التجريبية	المجموعة الضابطة
$n_1 = 25$	$n_2 = 25$
$\overline{X_1} = 7.60$	$\overline{X_2} = 6.0$
$S_1^2 = 2.27$	$S_2^2 = 1.78$

فهل تدل هذه البيانات على أن أداء المجموعة التجريبية كان أفضل من أداء المجموعة الضابطة عند مستوى α = 0.05 ؟

الحل:

سيتم اختبار الفرضيات التالية:

الفرضية الصفرية : لا توجد فروق ذات دلالة إحصائية بين متوسط المجموعة التجريبية ومتوسط المجموعة الضابطة ($\mu_1=\mu_2$).

الفرضية البديلة : توجد فروق ذات دلالة إحصائية بين متوسط المجموعة التجريبية ومتوسط المجموعة الضابطة لصالح المجموعة التجريبية ($\mu_1>\mu_2$)

<u>مستوى الدلالة :</u> α = 0.05

<u>منطقة الرفض:</u> قيمة مستوى الدلالة α = 0.05 والاختبار بذيل واحد ، ودرجات الحرية = 25 + 25 - 2 = 48 ، بذلك تكون قيمة (ت) الجدولية = 1.68

<u>المختبر الإحصائي:</u> $t = \dfrac{\overline{X_1} - \overline{X_2}}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

ولتطبيق هذه العلاقة يلزمنا حساب قيمة الانحراف المعياري (S) من خلال العلاقة التالية:

$$S^2 = \frac{\left[(n_1 - 1)(S_1^2)\right] + \left[(n_2 - 1)(S_2^2)\right]}{(n_1 + n_2) - 2}$$

إذا التباين يساوي: $S^2 = \dfrac{\left[(25-1)(2.27)^2\right] + \left[(25-1)(1.78)^2\right]}{(25+25)-2} = 4.16$

إذن الإنحراف المعياري يساوي: $S = \sqrt{S^2} = \sqrt{4.16} = 2.04$

ثم نحسب قيمة (ت) من خلال تطبيق العلاقة التالية: $t = \dfrac{\overline{X_1} - \overline{X_2}}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \dfrac{7.60 - 6.0}{2.04\sqrt{\dfrac{1}{25} + \dfrac{1}{25}}} = 2.77$

<u>القرار :</u>

قيمة (t) المحسوبة (2.77) أكبر من قيمـــة (ت) المجدولة (1.68) عند مستوى دلالة α = 0.05 .

∴ نرفض الفرضية الصفرية ونقبل البديلة

أي أن المجموعة التي خضعت للتجربة يصبح أداؤهم أفضل في عملية اتخاذ القرار من الذين لم يخضعون للتجربة وذلك عند مستوى دلالة α0.05 .

<u>أما في حالة عدم تجانس التباين</u> فإننا نستخدم طريقة وليتش Weltch لحساب قيمة (t) وذلك من خلال تطبيق العلاقة:

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

وللتحقق من تجانس التباين يتم ذلك من خلال تطبيق العلاقة التالية : $F = \dfrac{S_g^2}{S_l^2}$

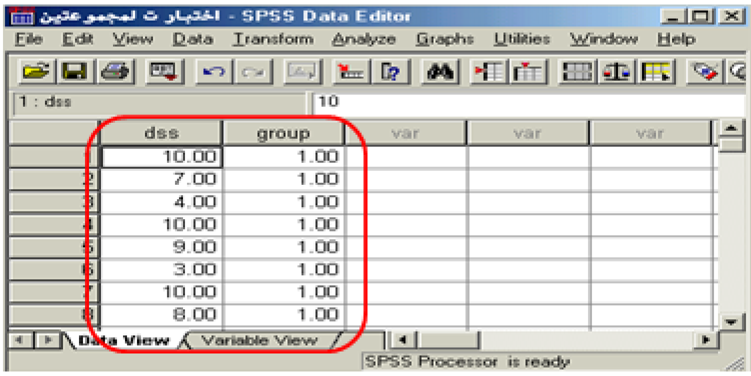
حيث: S_g^2 تعني التباين الأكبر ، S_l^2 تعني التباين الأصغر.

ومن ثم مقارنة قيمة (F) المحسوبة (لقياس تجانس التباين) بقيمة (F) المجدولة عند درجة حرية ($1-n_1$ و $1-n_2$)

حساب اختبار (ت) للعينات المستقلة Independent Samples T-Test من خلال الـ SPSS

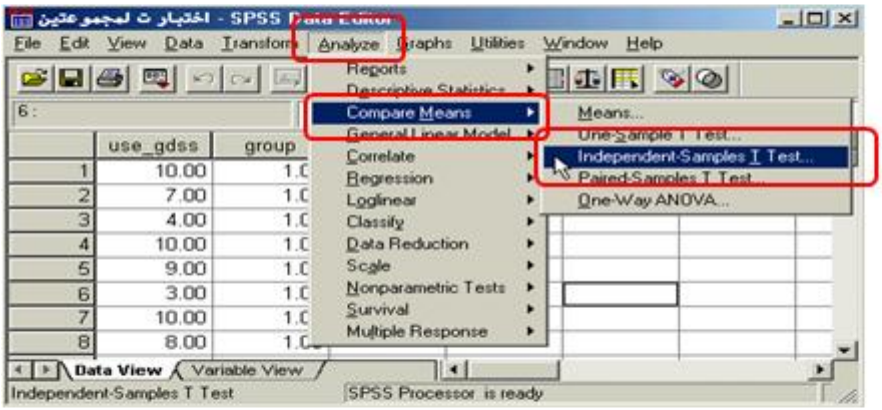
لغرض حساب قيمة (ت) لنفس المثال السابق من خلال استخدام برنامج الـ SPSS نتبع الخطوات التالية:

• قم بإدخال البيانات المراد تحليلها من خلال شاشة تحرير البيانات Data Editor بالطريقة المناسبة كالتالي:

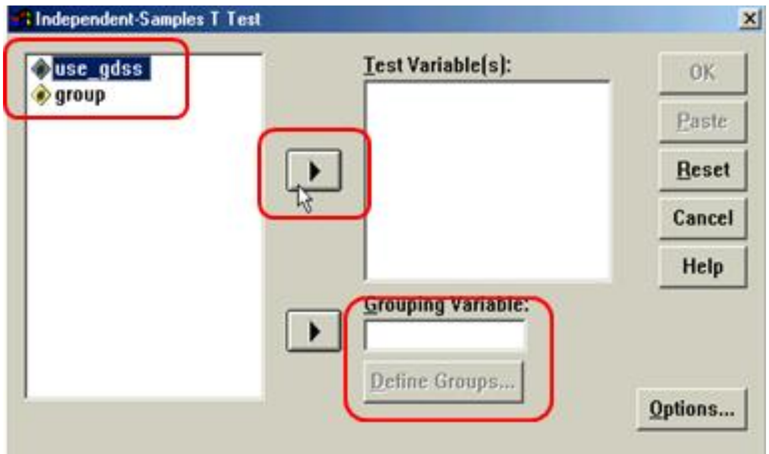


لاحظ أنه تم إدخال النتائج المتحصل عليها تحت متغير واحد باسم dss ، وتم إنشاء متغير آخر بمسمى group ليحوي رمز للمجموعة التجريبية والمجموعة الضابطة.

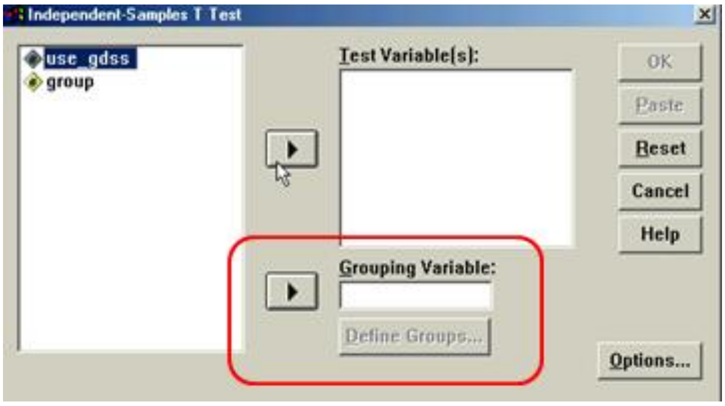
• من القائمة "تحليل" Analyze اختر الأمر "مقارنة المتوسطات" Compare Means فتظهر قائمة أوامر فرعية اختر منها "اختبار (ت) للعينات المستقلة" Independent Samples T-Testكالتالي:



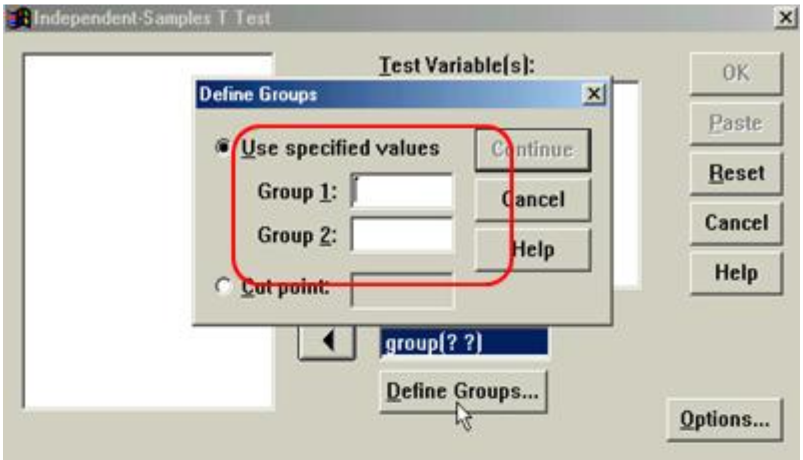
• بعد اختيار الأمر "اختبار (ت) للعينات المستقلة" Independent Samples T-Testسوف يظهر لك صندوق الحوار التالي:



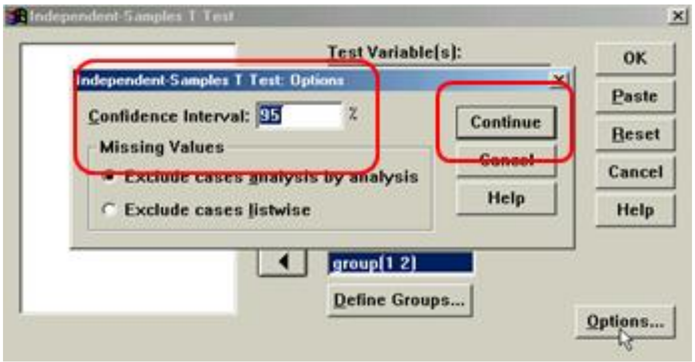
• من قائمة المتغيرات في الجهة اليسرى من صندوق الحوار حدد المتغير المراد نقله إلى المستطيل الخاص بـ "متغيرات الاختبار" Test Variable(s) ومن ثم أنقر على السهم الذي يظهر مقابل المستطيل الخاص بـ "متغيرات الاختبار" ، ستلاحظ انتقال المتغير مباشرة في المستطيل "متغيرات الاختبار" Test Variable(s) ، واعمل نفس الشيء مع المتغير group لنقله إلى الحقل الخاص بـ "مجاميع المتغيرات" Grouping Variable



- أنقر على زر "تعريف المجموعات" Define Groups في أسفل صندوق الحوار السابق ، سيؤدي ذلك إلى فتح صندوق حوار صغير يتيح الفرصة للمستخدم لتعريف قيم المجموعة الأولى (والتي يمثلها الرقم 1) وقيم المجموعة الثانية (والتي يمثلها الرقم 2).



- أنقر على زر "خيارات" Options في الجهة السفلية اليمنى من صندوق الحوار السابق وذلك عند الرغبة في تغيير قيمة "فترة الثقة" Confidence Interval حيث يظهر لك صندوق الحوار التالي والذي يتيح إمكانية تغيير فترة الثقة المختبرة (بشكل تلقائي سوف تظهر القيمة 95%) ، وبعد الانتهاء من التعديل على هذا الصندوق الحواري أنقر على زر "استمرار" Continue



- أنقر بعد ذلك على زر "موافق" OK سيؤدي ذلك إلى تنفيذ الاختبار، وستلاحظ ظهور النتائج في شاشة المخرجات كالتالي:

T-Test

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
USE_GDSS	USE DSS	25	7.6000	2.2730	.4546
	NOT USE DSS	25	6.0000	1.7795	.3559

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. E Differe
USE_GDSS	Equal variances assumed	1.095	.301	2.771	48	.008	1.6000	[illegible]
	Equal variances not assumed			2.771	45.386	.008	1.6000	[illegible]

يتضح من النتائج أن قيمة (F) = 1.095 ومستوى دلالتها 0.301 وهذه القيمة أكبر من 0.05 ، مما يدل على أنها غير دالة (وهذا يعني أن هناك تجانس بين تباين المجموعتين)، وهذا يدفعنا إلى قراءة نتائج اختبار (ت) المقابلة للعبارة "افتراض تساوي التباين" Equal variances assumed، من هذه النتائج نلاحظ أن قيمة (ت) المحسوبة t-test = 2.771 ، ودرجات الحرية df = 48 ، وقيمة Sig. (2-tailed) =0.008 ، وبما أن قيمة Sig. (2-tailed) في الجدول (0.008) أصغر من قيمة α = 0.05

فإننا بالتالي نرفض الفرضية الصفرية، أي أنه توجد فروق ذات دلالة إحصائية بين متوسط المجموعة التجريبية ومتوسط المجموعة الضابطة لصالح المجموعة الضابطة (وذلك بسبب حصولها على متوسط حسابي أكبر = 7.60).

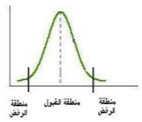
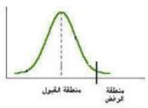
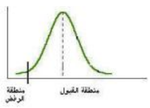
والأوامر المستخدمة لحساب اختبار (ت) للعينات المستقلة Independent Samples T-Test من خلال برنامج الـ SPSS (مع اختلاف المتغيرات موضع الدراسة) هي:

```
T-TEST

GROUPS=group(1 2)

/MISSING=ANALYSIS

/VARIABLES=use_gdss

/CRITERIA=CIN(.95)
```

الاختبارات الاحصائية لعينتين غير مستقلتين (العينات المرتبطة) Paried Samples t-test

يستخدم هذا النوع للحكم على دلالة الفروق ومعنويتها Significance بين متوسطي عينتين مرتبطتين Correlated Data، مثل اختبار دلالة الفروق بين متوسط أداء الموظفين قبل التدريب وبعد التدريب. ولغرض توضيح ذلك إليك هذا العرض الموجز لخطوات اختبار (ت) حول متوسطين مرتبطين على افتراض أن تباين المجتمع و غير معلومين وأن حجم العينة صغير .

والجدول التالي يوضح خطوات اختبار الفرق بين متوسطين مرتبطين باستخدام المختبر الإحصائي (ت) على افتراض أن تباين المجتمع σ_1^2 و σ_2^2 غير معلومين، وأن حجم العينة صغير

خطوات الاختبار	اختبار ذو طرفين	اختبار ذو طرف واحد	
		طرف يمين	طرف يسار
1- الفرضية الصفرية H_0 2- الفرضية البديلة H_1	$\mu = \mu_0$ $\mu \neq \mu_0$	$\mu = \mu_0$ $\mu > \mu_0$	$\mu = \mu_0$ $\mu < \mu_0$
3- مستوى الدلالة	α		
4- منطقة الرفض	ا ت ا ≤ ت($\frac{1}{2}$ α، df) 	ت ≤ ت(α،df) 	ت ≥ ت(α،df) 
5- المختبر الإحصائي	$$t=\frac{\overline{X_1}-\overline{X_2}}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}-2r\left(\frac{S_1}{\sqrt{n_1}}\right)\left(\frac{S_2}{\sqrt{n_2}}\right)}}$$ هذا المختبر الإحصائي يستخدم لتوضيح أهمية الفروق بين متوسطين للعينات المرتبطة		
6- القرار	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [ت ($\frac{1}{2}$ α، df)] الجدولية أي أن : ا ت ا ≤ ت($\frac{1}{2}$ α، df)	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [ت (α ،df)] الجدولية أي أن: ت ≤ ت(α،df)	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [-ت (α،df)] الجدولية أي أن: ت ≥ - ت(α،df)

ولتوضيح ما ورد في الجدول السابق دعنا نتناول هذا المثال:

أراد باحث أن يعرف أثر برنامج التدريب الصيفي في الميدان على أداء الطلاب وتحصيلهم في كلية العلوم الإدارية، ولغرض تحقيق ذلك قام الباحث باختبار الطلاب قبل وبعد البرنامج التدريبي، ولكون نفس الطلاب أخذوا الاختبارين، فإن الباحث يتوقع معامل ارتباط موجب بين تحصيل الطلبة في كلا القياسين. ولغرض اختبار مدى دلالة الفروق بين الاختبار القبلي والاختبار البعدي، لا بد على الباحث أن يتأكد من قيمة الارتباط بين الاختبارين والتي كانت r = 0.46 ، وقد كانت النتائج التي تم التوصل إليها كما يلي:

الإختبار القبلي	الإختبار البعدي
$n_1 =$ 100	$n_2 =$ 100
$\overline{X_1} =$ 54.28	$\overline{X_2} =$ 58.66
$S_1^2 =$ 49	$S_2^2 =$ 64

فهل تدل هذه البيانات على أن أداء الطلاب التحصيلي في الكلية بعد أخذ البرنامج التدريبي كان أفضل من أدائهم قبل أخذ البرنامج التدريبي عند مستوى α = 0.05 ؟

الحل:

سيتم اختبار الفرضيات التالية:

الفرضية الصفرية : لا توجد فروق ذات دلالة إحصائية بين متوسط تحصيل الطلاب قبل وبعد البرنامج التدريبي ($\mu_1=\mu_2$)

الفرضية البديلة : توجد فروق ذات دلالة إحصائية بين متوسط تحصيل الطلاب قبل وبعد البرنامج التدريبي ($\mu_2 \neq \mu_1$)

مستوى الدلالة : α = 0.05

منطقة الرفض : قيمة مستوى الدلالة α = 0.05 والاختبار بذيلين، ودرجات الحرية = 100-1 =99 ، بذلك تكون قيمة (ت) الجدولية = 1.980

المختبر الإحصائي:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} - 2r\left(\frac{S_1}{\sqrt{n_1}}\right)\left(\frac{S_2}{\sqrt{n_2}}\right)}}$$

إذا قيمة (ت) تساوي:

$$t = \frac{58.66 - 54.28}{\sqrt{\frac{64.0}{100} + \frac{49.0}{100} - 2(0.46)\left(\frac{8}{\sqrt{100}}\right)\left(\frac{7}{\sqrt{100}}\right)}} = 5.57$$

في هذه المعادلة ليس هناك مانع من الابتداء بـ X_1 أو X_2 في الترتيب ، لأن الإشارة ليس لها أي تأثير على النتيجة المتحصلة

القرار :

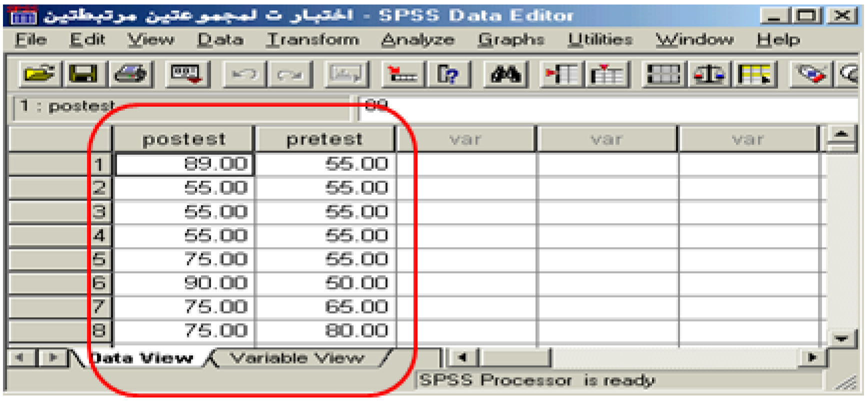
قيمة (ت) المحسوبة (5.57) أكبر من قيمـــة (ت) المجدولة (1.980) .عند مستوى دلالة α = 0.05 .

∴ نرفض الفرضية الصفرية ونقبل البديلة، أي أن للبرنامج التدريبي تأثير إيجابي على تحصيل الطلاب وأدائهم في الكلية وذلك عند مستوى دلالة 05.α

حساب اختبار (ت) لعينتين غير مستقلتين (العينات المرتبطة) Paired Samples T-Test من خلال الـ SPSS

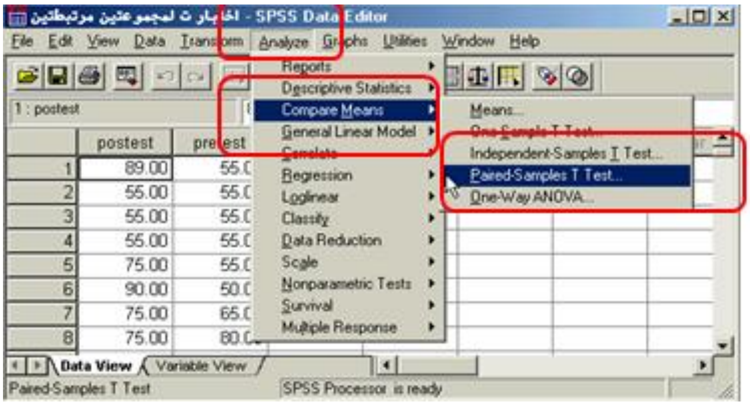
لغرض حساب قيمة (ت) لنفس المثال السابق من خلال استخدام برنامج الـ SPSS نتبع الخطوات التالية:

- قم بإدخال البيانات المراد تحليلها من خلال شاشة تحرير البيانات Data Editor بالطريقة المناسبة كالتالي:

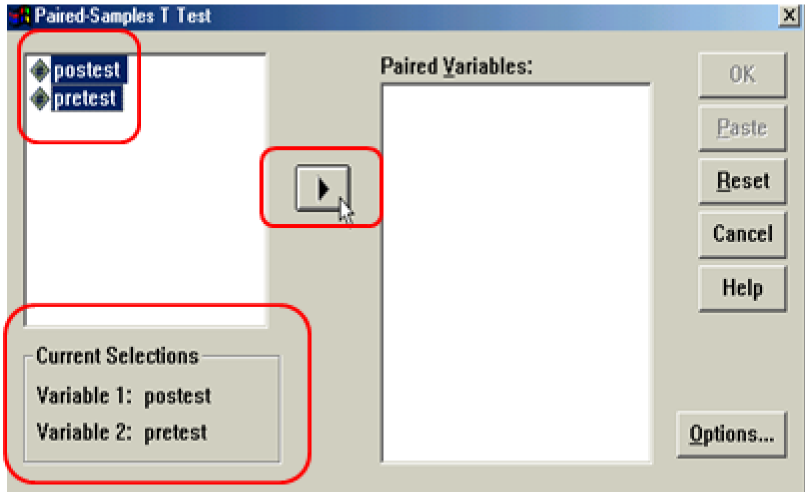


لاحظ أنه تم إدخال البيانات بطريقة مختلفة عن ما تم اتباعه في حالة العينتين المستقلتين، هنا لابد من إدخال بيانات كل متغير في عمود منفصل عن الآخر، وقد تم إعطاء كل متغير اسم مختلف عن الآخر posttest و pretest

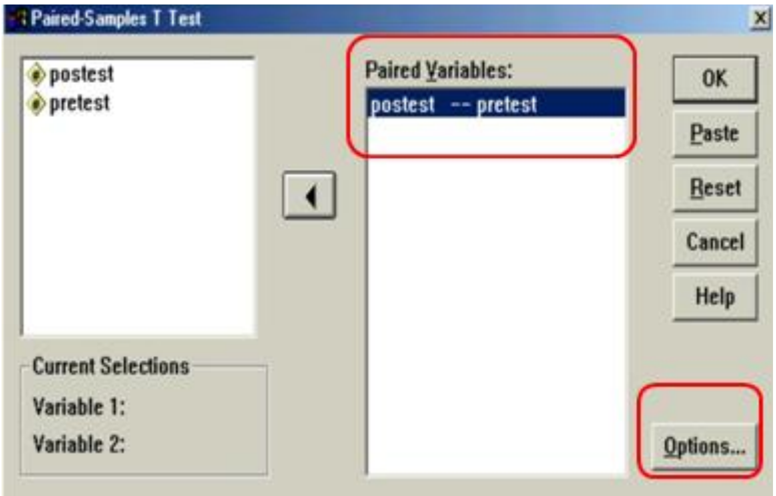
- من القائمة "تحلل" Analyze اختر الأمر "مقارنة المتوسطات" Compare Means فتظهر قائمة أوامر فرعية اختر منها "اختبار (ت) للعينات المرتبطة" Paired-Samples T-Test كالتالي:



- بعد اختيار الأمر "اختبار (ت) للعينات المرتبطة" Paired-Samples T-Test سوف يظهر لك صندوق الحوار التالي:

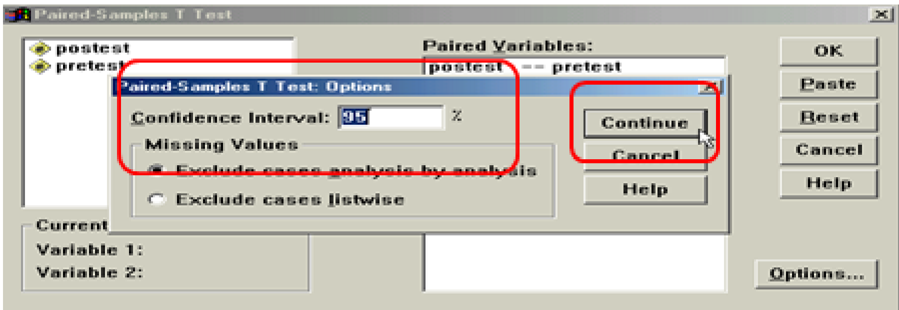


- من قائمة المتغيرات في الجهة اليسرى من صندوق الحوار حدد المتغيرين المرتبطين مع بعضها لتحليلها كأزواج، ونقلها إلى المستطيل الخاص بـ "المتغيرات الزوجية" Paired Variables (سوف تلاحظ أثناء التحديد ظهور اسم المتغير الأول واسم المتغير الثاني بعد كل عملية تحديد في المربع اسفل قائمة المتغيرات)، ثم بعد ذلك أنقر على السهم الذي يظهر مقابل المستطيل الخاص بـ "متغيرات الاختبار" ، ستلاحظ انتقال المتغير مباشرة في المستطيل "المتغيرات الزوجية" Paired Variable(s)، كرر نفس الإجراء مع المتغيرات الزوجية الأخرى والمراد تحليلها



- أنقر على زر "خيارات" Options في الجهة السفلية اليمنى من صندوق الحوار السابق وذلك عند الرغبة في تغيير قيمة "فترة الثقة" Confidence Interval حيث يظهر لك صندوق الحوار التالي والذي يتيح إمكانية تغيير فترة الثقة

المختبرة (بشكل تلقائي سوف تظهر القيمة 95%) ، وبعد الانتهاء من التعديل على هذا الصندوق الحواري أنقر على زر "استمرار" Continue



- أنقر بعد ذلك على زر "موافق" OK سيؤدي ذلك إلى تنفيذ الاختبار، وستلاحظ ظهور النتائج في شاشة المخرجات كالتالي:

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	POSTEST	58.6600	100	8.0000	.8000
	PRETEST	54.2800	100	7.0000	.7001

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	POSTEST & PRETEST	100	.458	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	POSTEST - PRETEST	4.3800	7.8570	.7857	2.8210	5.9390	5.575	99	.000

نلاحظ أن برنامج الـ SPSS قام مباشرة بحساب الإحصاءات الأساسية للبيانات مثل المتوسط الحسابي للمتغير Posttest (58.660) والانحراف المعياري لنفس المتغير (8.00) ، أما المتغير Pretest فقد كان المتوسط الحسابي (54.280) والانحراف المعياري (7.00) . بالإضافة إلى ذلك تم حساب معامل ارتباط بيرسون للمتغيرات موضع الدراسة Paired Sample Correlation وقد كانت قيمته (0.458).

ثم بعد ذلك قام البرنامج بحساب قيمة (ت) للمتغيرات موضع الدراسة في الجدول المعنون بـ "اختبار العينات المرتبطة" Paired Sample Test ، ومن هذه النتائج نلاحظ أن قيمة (ت) المحسوبة t-test = 5.575 ، ودرجات الحرية df = 99 ، وقيمة Sig. (2-tailed) =0.000 ، وبما أن قيمة Sig. (2-tailed) في الجدول (0.000) أصغر من قيمة α = 0.05 فإننا بالتالي نرفض الفرضية الصفرية ، أي أن أداء الطلاب في الكلية بعد أخذ البرنامج التدريبي كان أفضل من أدائهم قبل أخذ البرنامج التدريبي عند مستوى α = 0.05 .

والأوامر المستخدمة لحساب اختبار (ت) للعينات المرتبطة Paired Samples T-Test من خلال برنامج الـ SPSS (مع اختلاف المتغيرات موضع الدراسة) هي:

```
T-TEST
PAIRS= postest  WITH pretest (PAIRED)
/CRITERIA=CIN(.95)
/MISSING=ANALYSIS.
```

المحاضرة 12 (الجزء 1)

تابع اختبار الفروض الإحصائية المعلمية

الاختبارات الاحصائية لأكثر من عينتين مستقلتين

ناقشنا في المحاضرة السابقة طرق الاستدلال الإحصائي عن متوسط المجتمع والفرق بين متوسطين، وسنناقش في هذه المحاضرة طرق الاستدلال الإحصائي للفرق بين ثلاث متوسطات أو أكثر وذلك من خلال توزيع F

سمي توزيع F بهذا الاسم تخليدا للعالم رونالد فيشر R.A. Fisher الذي يعتبر أول من اشتق هذا التوزيع ووصفه وذلك في العشرينات والثلاثينات من القرن العشرين لذلك تعرف أحيانا بتحليل فيشر للتباين.

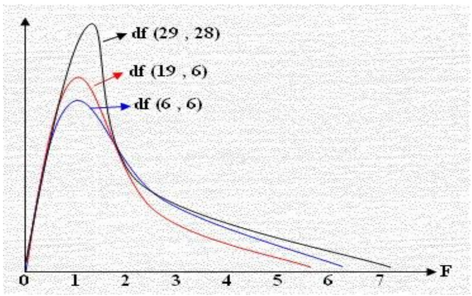
ويستخدم توزيع F أساسا لاختبار تساوي تبايني مجتمعين، ومن المثير للإنتباه ملاحظة أن اختبار تساوي التباينين يستخدم لاختبار تساوي ثلاث متوسطات أو أكثر.

وتسمى طريقة الاستدلال الاحصائي عن تساوي ثلاث متوسطات أو أكثر بتحليل التباين Analysis of Variance.

وتوزيع F عبارة عن مجموعة من المنحنيات التكرارية يتميز كل منها عن الآخر برقمين لدرجات الحرية أحدهما يمثل درجة حرية للبسط والآخر درجة حرية للمقام.

وقيمة F هي قيمة توضح نسبة التباين Variance ratio لعينتين والرمز F إشارة إلي العالم Fisher الذي قام بعمل هذا الاختبار والمعروف باختبار F وقد قام العالم Snedecor بحساب جداول خاصة لتوزيع F وفيها درجات الحرية التي في أعلى الجدول تخص البسط أما درجات الحرية على العمود الجانبي فتخص المقام.

وتوزيع F هو توزيع ملتو جهة اليمين بمعلمتين تتمثلان بدرجتي حرية (البسط ، المقام) وهما k – 1 للبسط ، n – k للمقام حيث n مجموع أحجام العينات و k هي عدد المجموعات موضع الدراسة.



أهم استخدامات توزيع F (ف) هي:

- تقدير فترة الثقة لـ σ_1^2/σ_2^2
- اختبار فرضيات حول تساوي تباينين أي: $H_0 : \sigma_1^2 = \sigma_2^2$
- اختبار فرضيات حول تساوي أكثر من متوسطين أي: $H_0 : \mu_1 = \mu_2 = \mu_3 = ...$

تحليل التباين الأحادي One Way ANOVA

تحليل التباين هو عملية يقصد بها تقسيم مجموع مربعات الانحرافات عن المتوسط الحسابي إلي مكوناته إرجاع كل من هذه المكونات إلي مسبباتها. وطريقة تحليل التباين تفيد في مقارنة عدد من المعاملات يزيد عن أثنين كما تمتاز طريقة تحليل التباين بأنه يمكن فيها استعمال كل البيانات المأخوذة من التجربة في حساب قيمة واحدة للانحراف القياسي يمكن بها مقارنة المجموعات أو المعاملات التجريبية.

فهي مجموعة من النماذج الإحصائية (statistical model) مع إجرءات مرافقة لهذه النماذج تمكن من مقارنة المتوسطات لمجتمعات إحصائية مختلفة عن طريق تقسيم التباين variance الكلي الملاحظ بينهم إلى أجزاء مختلفة.

تتلخص طريقة تحليل التباين في:

١. حساب المجموع الكلي لمربعات انحرافات كل المفردات في التجربة عن المتوسط العام .
٢. تقسيم هذا المجموع الكلي لمربعات الانحرافات Total Sum Squares إلي مكوناته طبقا للمصادر المسببة لها والذي يختلف عددها طبقا للتصميم المستعمل في التجربة.
٣. تقسم درجات الحرية الكلية طبقا للمصادر السابقة أيضا.
٤. تدون النتائج في جدول يسمي جدول تحليل التباين ANOVA ترتب فيه مصادر الاختلافات حسب التصميم الإحصائي المستعمل ويسهل هذا الجدول عمل اختبار معنوية المعاملات.

سبب الحاجة لاختبار تحليل التباين:

فالاختبار الآنف الذكر (t) يهتم باختبار الفرضيات حول اختلاف وسطي مجتمعين إحصائيين، ولكن عندما نرغب في اختبار الفروق بين أكثر من مجتمعين (ثلاثة مثلاً) فهل من الممكن أن نستخدم اختبار (Z أو t) ؟ .

إنه من الممكن استخدام هذه الاختبارات وذلك بين كل زوجين من المتوسطات مع مراعاة خطأ "بانفروني" Bonferroni، أي إنه يمكن أن نختبر الفرق بين كل متوسطين معاً ، ومن ثم نقسم قيمة على عدد المقارنات ، أي انه إذا كان لدينا ثلاث متوسطات ونريد أن نقارن بينها باستخدام اختبار (t) فإن المقارنات ستكون كالتالي: م1 ، م2 ثم م1 ، م3 ثم م2 ، ، م3

وإذا كان مستوى الدلالة الدلالة = 0.05 فإننا نقوم بقسمته على 3 ومن ثم نختبر قيمة t المحسوبة على مستوى الدلالة الجديد وذلك لتحاشي الخطأ من النوع الأول (رفض الفرضية الصفرية عندما تكون صحيحة)

إلا أننا عندما نُقدمْ على ذلك (أي نستخدم اختبار *Z* أو t بدلا من اختبار *F*) فإننا نكون قد ابتعدنا عن طريق الصواب لأسباب منها:

١. كلما ازداد عدد المتوسطات كلما ازداد عدد الأزواج التي يجب أن نجري عليها الاختبار ، وقد يصبح العدد كبيراً جداً ، مثال ذلك أننا رغم وجود ثلاثة أزواج من المتوسطات لثلاث عينات، إلا أن هذا العدد من الأزواج يرتفع إلى (10) أزواج من المتوسطات إذا أصبح عدد العينات خمس ، أما إذا أصبح عدد العينات (8) فإننا نجد أن لدينا (28) زوجاً من المتوسطات ، وبذلك تتزايد كمية الجهد المبذول في إجراء العمليات الحسابية المختلفة تزايداً سريعاً كلما ازداد عدد العينات ، وبشكل عام إذا كان عدد الأوساط (x) فإن عدد المقارنات المختلفة بين أزواج هذه الأوساط = $\frac{x(x-1)}{2}$
٢. عند المقارنة بين كل زوج من الأوساط فإننا نستخدم فقط المعلومات عن المجموعتين المقارنتين ونهمل المعلومات المتوفرة في باقي المجموعات والتي تجعل المقارنة أقوى فيما لو استعملت
٣. إن هذا الاستخدام المتعدد لاختبار (t) مثلاُ يزيد من الخطر في ارتكاب الخطأ من النوع الأول (رفض الفرضية الصفرية عندما تكون صحيحة) . فإذا كان عدد المقارنات التي تستخدم فيها اختبار (t) = ك ، α وكان مستوى الدلالة المستخدم في المقارنة = α ، فإن احتمال ارتكاب خطأ واحد أو أكثر من النوع الأول في هذه المقارنات يكون بالعلاقـــــــة =[1 - (1 -)^{ك}] حيث "ك" تعني عدد المقارنات

فإذا كان عدد المقارنات = 3 ، وكان مستوى الدلالة = 0.05 فإن احتمال ارتكاب خطأ واحد من النوع الأول = [1 - (1 - 0.05)^{3}] = 0.14 . أما إذا كان عدد الأوساط = 20 وسطا، فإن عدد المقارنات = 110 مقارنة ، واحتمال ارتكاب الخطأ من النوع الأول = 1 تقريباً . وهذه الزيادة في احتمال ارتكاب الخطأ من النوع الأول تعني ازدياد الاحتمال في ادعاء وجود فروق ذات دلالة إحصائية في حالة مقارنة واحدة على الأقل بين هذه الأوساط، في حين أن هذه الفروق ليست في الحقيقة ذات دلالة.

٤. إن الباحث كثيرا ما يرغب في اختبار مشكلة أوسع من مجرد معرفة ما إذا كانت أزواج المتوسطات مختلفة عن بعضها البعض

وفي ضوء هذه الأسباب ، فإن التوجه يجب أن يكون نحو استخدام اختبار أفضل من اختباري (Z أو t) وأقوى منهما ، وذلك لإجراء المقارنة بين ثلاث أوساط فأكثر، وتحليل التباين هو أكثر هذه الاختبارات شهرة واستخداماً لمثل هذا الغرض، وهذا الاختبار يعتبر أداة فعالة وقوية في يد الباحث يفيده في بناء تصميم تجاربه بكفاءة، كما ويساعده في فحص أثر التفاعل بين متغيرات الدراسة .

تحليل التباين الأحادي (مستوى واحد)

هو طريقة لاختبار معنوية الفرق بين المتوسطات لعدة عينات بمقارنة واحدة، ويعرف أيضاً بطريقة تؤدي لتقسيم الاختلافات الكلية لمجموعة من المشاهدات التجريبية لعدة أجزاء للتعرف على مصدر الاختلاف بينها ولذا فالهدف هنا فحص تباين المجتمع لمعرفة مدى تساوى متوسطات المجتمع ولكن لا بد من تحقيق ثلاثة أمور قبل استخدامه وهي:

١. العينات عشوائية ومستقلة.
٢. مجتمعات هذه العينات كلاً لها توزيع طبيعي.
٣. تساوي تباين المجتمعات التي أخذت منها العينات العشوائية المستقلة.

ولتوضيح ما سبق بمقارنة متوسطات ثلاث مجتمعات باستخدام ثلاث عينات (تحقق فيها الشروط الثلاثة السابقة) موضحة بالجدول الآتي:

العينة الثالثة	العينة الثانية	العينة الأولى
33 32 33.5 31.5	27 28 26.5 26.5	40 41 40.5 38.5
$\bar{X}_3 = 32.5$ $S_3 = 0.91$	$\bar{X}_2 = 27$ $S_2 = 0.71$	$\bar{X}_1 = 40$ $S_1 = 1.08$

السؤال: هل في البيانات ما يكفي لوجود فرق بين المتوسطات؟

الجواب: نعم (بمجرد النظر) فالتشتت (التباين) ظاهر 40، 27، 32.5 (المتوسطات) بمقارنته بالتشتت بين العينات (وحداتها 40 ، 41 ، 40.5، 38.5) فيبدو معدوماً.

فلذا أخذنا البيانات الآتية:

العينة الثالثة	العينة الثانية	العينة الأولى
10 60 27.5	50 20 11	40 15 65
$\bar{X}_3 = 32.5$ $S_3 = 25.4$	$\bar{X}_2 = 27$ $S_2 = 20.4$	$\bar{X}_1 = 40$ $S_1 = 25$

فالبيانات هنا لها نفس المتوسطات في البيانات السابقة ولكن التشتت (داخل العينات) كبيراً بما هو عليه في المتوسطات.

فالدليل على وجود الفرق بين متوسطات الجدول الأول واضح ولا يظهر ذلك بوضوح في بيانات الجدول الثاني بالرغم من تساوي المتوسطات في الحالتين

ولذا يتبين لنا القصد من تحليل التباين والذي يعني الفرق بين المتوسطات والذي يقاس بالتشتت داخل البيانات .

الافتراضات الاساسية لاختبار تحليل التباين:

يستند اختبار تحليل التباين إلى توفر عدد من الافتراضات، ومن هذه الافتراضات ما يلي:

١. **مستوى القياس:** يشترط لاستخدام هذا الاختبار أن تكون البيانات فترية (فئوية) أو نسبية.
٢. **حجم العينة:** يقتضي هذا الافتراض أن يكون حجم العينة كبيراً.
٣. **التوزيع الطبيعي للمجتمع الإحصائي:** يقتضي هذا الافتراض أن تكون المشاهدات في كل مجتمع من المجتمعات موزعة بشكل طبيعي، ولكن يرى الإحصائيون أن اختبار (F) لا يتأثر كثيراً بعدم توفر هذا الشرط وذلك عندما يكون حجم الخلية كبيراً والتوزيع ليس طبيعياً
٤. **تجانس التباين:** أي أن يكون للمجتمعات في مستويات المعالجة المختلفة نفس التباين (S^2) بالرغم من أن لها بالطبع أوساطاً مختلفة، وإن الإخلال بهذا الافتراض في حالة تساوي أحجام الخلايا لا يؤثر على النتائج بشكل كبير، أما عند عدم تساوي أحجام الخلايا وانتماء الخلايا ذات الحجم الصغير إلى مجتمعات ذات تباين كبير، فإن الإخلال بهذا الافتراض يؤدي إلى ارتفاع احتمال الخطأ من النوع الأول ، أما عندما تنتمي الخلايا ذات الحجم الكبير للمجتمعات ذات التباين الكبير فإنه يقلل من احتمال ارتكاب الخطأ من النوع الأول.

مثال:

إذا كان لدينا ثلاث منتجات لإحدى الشركات الصناعية ، وتم تقييمها من قبل مجموعة من المستهلكين وحصلنا على النتائج التالية :

المنتج (1) X_1	المنتج (2) X_2	المنتج (3) X_3
7	4	2
10	6	2
10	7	3
11	9	7
12	9	6
50	35	20

المطلوب: هل هناك فروق ذات دلالة بين المنتجات الثلاثة ؟

الحل:

لكون لدينا ثلاث متغيرات فترية، ولرغبة الشركة معرفة الفروق بين هذه المتغيرات موضع الدراسة، فإن أنسب أسلوب إحصائي هنا هو تحليل التباين الأحادي One Way ANOVA ، ولغرض حساب تحليل التباين الأحادي، **علينا اتباع الخطوات التالية:**

- نجمع قيم كل متغير للحصول على $\sum X$ لكل متغير
- نربع كل درجة في كل متغير للحصول على X^2 لكل متغير
- نجمع قيم مربع كل درجة للحصول على $\sum X^2$ لكل متغير
- نربع مجموع كل متغير للحصول على $\left(\sum X\right)^2$ لكل متغير
- نحسب متوسط كل متغير من خلال العلاقة: $\overline{X} = \frac{\sum x}{n}$
- نحسب مجموع المربعات الكلي Total Sum of Squares وذلك من خلال العلاقة التالية:

$Total..SS = \sum X^2 - \frac{\left(\sum X\right)^2}{n}$ ، حيث n تعني مجموع أعدد الأفراد في جميع المجموعات

أو يمكن حساب مجموع المربعات الكلي من خلال العلاقة التالية: *Total SS = Between SS + Within SS*

- نحسب مجموع المربعات بين المجموعات Between Sum of Squares وذلك من خلال العلاقة التالية:

$$Between..SS = \sum \frac{\left(\sum X_g\right)^2}{n_g} - \frac{\left(\sum X\right)^2}{n}$$

حيث:

n_g تعني عدد الأفراد في كل مجموعة

n تعني مجموع أعدد الأفراد في جميع المجموعات

$\sum \frac{\left(\sum X_g\right)^2}{n_g}$ تعني مربع مجموع قيم كل مجموعة مقسوما على عدد أفراد تلك المجموعة

$\frac{\left(\sum X\right)^2}{n}$ تعني مربع مجموع قيم كل المجموعات مقسوما على مجموع أعداد الأفراد في جميع المجموعات

- نحسب مجموع المربعات داخل المجموعات Within Sum of Squares وذلك من خلال العلاقة التالية:

$$Within..SS = \sum \left[\sum X_g^2 - \frac{\left(\sum X_g\right)^2}{n_g} \right]$$

أو يمكن حساب مجموع المربعات داخل المجموعات من خلال العلاقة التالية:

Within SS = Total SS - Between SS

- نحسب درجات الحرية بين المجموعات Between groups degrees of freedom من خلال العلاقة التالية:

$K-1$ حيث *K* تعني عدد المجموعات

ودرجات الحرية داخل المجموعات Within groups degrees of freedom من خلال العلاقة التالية:

$n-K$ حيث *n* تعني عدد الأفراد أو الاستجابات في المجموعات موضع الدراسة ، و*K* تعني عدد المجموعات.

ودرجات الحرية الكلية Total degrees of freedom من خلال العلاقة التالية:

$n-1$ حيث *n* تعني عدد الأفراد أو الاستجابات في المجموعات موضع الدراسة .

- نحسب التباين بين المجموعات أو ما يسمى متوسط المربعات بين المجموعات Between mean square وذلك من خلال العلاقة التالية: $Beween..groups..mean..square = \frac{Between..SS}{K-1}$

- نحسب التباين داخل المجموعات أو ما يسمى متوسط المربعات داخل المجموعات Within mean square وذلك من خلال العلاقة التالية: $Within..groups..mean..square = \frac{Within..SS}{(n-K)}$

- نحسب قيمة *F* من خلال العلاقة التالية : $F = \frac{Between..groups..mean..square}{Within..groups..mean..square}$

- نقارن بعد ذلك قيمة *F* المحسوبة بقيمة *F* المجدولة لاتخاذ القرار المناسب اتجاه الفرضية موضع الدراسة .

نقوم الآن بتطبيق جميع الخطوات السابقة لحساب تحليل التباين الأحادي على بيانات المثال السابق وذلك حتى يسهل علينا استخلاص بعض القيم المطلوبة لحساب هذا النوع من الاختبارات الإحصائية وتطبيق المعادلات السابقة.

المنتج (3) X_3		المنتج (2) X_2		المنتج (1) X_1	
X_3^2	X_3	X_2^2	X_2	X_1^2	X_1
4	2	16	4	49	7
4	2	36	6	100	10
9	3	49	7	100	10
49	7	81	9	121	11
36	6	81	9	144	12
102	20	**263**	35	**514**	50

☒ وضع فرض العدم والفرض البديل

صياغة الفرضية الصفرية كالتالي: $H_0 : \mu_1 = \mu_2 = \mu_3$

في حين تفترض الفرضية البديلة التالي: **متوسطان على الأقل غير متساويين : H_A**

☒ تحديد مستوى الدلالة (α): وتحدد مستويات المعنوية سلفاً وهي عادة 0.05 أو 0.01

☒ حساب إحصائية الاختبار (F) وذلك من خلال اتباع الخطوت التالية:

- المتوسط الحسابي لـ X_1 = $\overline{X} = \frac{50}{5} = 10$
- المتوسط الحسابي لـ X_2 = $\overline{X} = \frac{35}{5} = 7$
- المتوسط الحسابي لـ X_3 = $\overline{X} = \frac{20}{5} = 4$
- مجموع المربعات الكلي Total Sum of Squares = $Total..SS = \sum X^2 - \frac{\left(\sum X\right)^2}{(n_g)(k)} = 879 - \frac{(105)^2}{15} = 144$

حيث n_g تعني عدد أفراد المجموعة المحددة ، K تعني عدد المجموعات موضع الدراسة

- مجموع المربعات بين المجموعات Between Sum of Squares =

$$Between..SS = \sum \frac{\left(\sum X_g\right)^2}{n_g} - \frac{\left(\sum X\right)^2}{(n_g)(k)} = \frac{(50)^2}{5} + \frac{(35)^2}{5} + \frac{(20)^2}{5} - \frac{(105)^2}{15} = 90$$

- مجموع المربعات داخل المجموعات Within Sum of Squares =

$$Within..SS = \sum \left[\sum X_g^2 - \frac{\left(\sum X_g\right)^2}{n_g} \right]$$

$$\sum x_1^2 = 514 - \frac{(50)^2}{5} = 14$$

$$\sum x_2^2 = 263 - \frac{(35)^2}{5} = 18$$

$$\sum x_3^2 = 102 - \frac{(20)^2}{5} = 22$$

نقوم بعد ذلك بجمع نواتج هذه المعادلات لنحصل على مجموع المربعات داخل المجموعات كالتالي:

Within sum of squares = 14+18+22=54

- نحسب درجات الحرية:

درجات الحرية بين المجموعات Between groups degrees of freedom $(K-1) = 3-1 = 2$

درجات الحرية داخل المجموعات Within groups degrees of freedom $(n-K)=15-3=12$
درجات الحرية الكلية Total degrees of freedom $(n-1)=15-1=14$

- التباين بين المجموعات أو ما يسمى متوسط المربعات بين المجموعات Between mean square =

$$Beween..groups..mean..square=\frac{Between..SS}{K-1}=\frac{90}{2}=45$$

- التباين داخل المجموعات أو ما يسمى متوسط المربعات داخل المجموعات Within mean square =

$$Within..groups..mean..square=\frac{Within..SS}{(n-K)}=\frac{54}{12}=4.5$$

- قيمة F = $F=\frac{Between..groups..mean..square}{Within..groups..mean..square}=\frac{45}{4.5}=10$

- نقوم بعد ذلك بتفريغ ما تم الحصول عليه من معلومات في جدول تحليل التباين كالتالي:

مصدر التباين	مجموع المربعات SS	درجات الحرية df	متوسط المربعات Means	قيمة F
بين المجموعات Between groups	90	2	45	10
داخل المجموعات Within groups	54	12	4.5	
الكلي (المجموع) Total	144	14		

وبالرجوع إلى جدول توزيع *F* نجد أن القيمة الحرجة لـ *F* بدرجات حرية للبسط تساوي 2 ودرجات حرية للمقام تساوي 12 وباستخدام مستوى= 0.05 نجد أن القيمة الحرجة تساوي 3.88 ،

وحيث أن القيمة المحسوبة لـ *F* = 10 وهي بالتالي أكبر من القيمة الحرجة المجدولة، نستنتج أن الفرضية الصفرية تكون مرفوضة، أي يوجد اختلاف بين متوسطي مجتمعين على الأقل من المجتمعات التي قيّمة من المستهلكين ولمعرفة بين أي من المنتجات تكون الفروق ينبغي علينا اللجوء إلى أسلوب المقارنات المتعددة Multiple Comparisons .

Table F for Alfa = 0.10

\	$df_1=1$	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
$df_2=1$	39.86346	49.50000	53.59324	55.83296	57.24008	58.20442	58.90595	59.43898	59.85759	60.19498	60.70521	61.22034	61.74029	62.00205	62.26497	62.52905	62.79428	63.06064	63.32812
2	8.52632	9.00000	9.16179	9.24342	9.29263	9.32553	9.34908	9.36677	9.38054	9.39157	9.40813	9.42471	9.44131	9.44962	9.45793	9.46624	9.47456	9.48289	9.49122
3	5.53832	5.46238	5.39077	5.34264	5.30916	5.28473	5.26619	5.25167	5.24000	5.23041	5.21562	5.20031	5.18448	5.17636	5.16811	5.15972	5.15119	5.14251	5.13370
4	4.54477	4.32456	4.19086	4.10725	4.05058	4.00975	3.97897	3.95494	3.93567	3.91988	3.89553	3.87036	3.84434	3.83099	3.81742	3.80361	3.78957	3.77527	3.76073
5	4.06042	3.77972	3.61948	3.52020	3.45298	3.40451	3.36790	3.33928	3.31628	3.29740	3.26824	3.23801	3.20665	3.19052	3.17408	3.15732	3.14023	3.12279	3.10500
6	3.77595	3.46330	3.28876	3.18076	3.10751	3.05455	3.01446	2.98304	2.95774	2.93693	2.90472	2.87122	2.83634	2.81834	2.79996	2.78117	2.76195	2.74229	2.72216
7	3.58943	3.25744	3.07407	2.96053	2.88334	2.82739	2.78493	2.75158	2.72468	2.70251	2.66811	2.63223	2.59473	2.57533	2.55546	2.53510	2.51422	2.49279	2.47079
8	3.45792	3.11312	2.92380	2.80643	2.72645	2.66833	2.62413	2.58935	2.56124	2.53804	2.50196	2.46422	2.42464	2.40410	2.38302	2.36136	2.33910	2.31618	2.29257
9	3.36030	3.00645	2.81286	2.69268	2.61061	2.55086	2.50531	2.46941	2.44034	2.41632	2.37888	2.33962	2.29832	2.27683	2.25472	2.23196	2.20849	2.18427	2.15923
10	3.28502	2.92447	2.72767	2.60534	2.52164	2.46058	2.41397	2.37715	2.34731	2.32260	2.28405	2.24351	2.20074	2.17843	2.15543	2.13169	2.10716	2.08176	2.05542
11	3.22520	2.85951	2.66023	2.53619	2.45118	2.38907	2.34157	2.30400	2.27350	2.24823	2.20873	2.16709	2.12305	2.10001	2.07621	2.05161	2.02612	1.99965	1.97211
12	3.17655	2.80680	2.60552	2.48010	2.39402	2.33102	2.28278	2.24457	2.21352	2.18776	2.14744	2.10485	2.05968	2.03599	2.01149	1.98610	1.95973	1.93228	1.90361
13	3.13621	2.76317	2.56027	2.43371	2.34672	2.28298	2.23410	2.19535	2.16382	2.13763	2.09659	2.05316	2.00698	1.98272	1.95757	1.93147	1.90429	1.87591	1.84620
14	3.10221	2.72647	2.52222	2.39469	2.30694	2.24256	2.19313	2.15390	2.12195	2.09540	2.05371	2.00953	1.96245	1.93766	1.91193	1.88516	1.85723	1.82800	1.79728
15	3.07319	2.69517	2.48979	2.36143	2.27302	2.20808	2.15818	2.11853	2.08621	2.05932	2.01707	1.97222	1.92431	1.89904	1.87277	1.84539	1.81676	1.78672	1.75505
16	3.04811	2.66817	2.46181	2.33274	2.24376	2.17833	2.12800	2.08798	2.05533	2.02815	1.98539	1.93992	1.89127	1.86556	1.83879	1.81084	1.78156	1.75075	1.71817
17	3.02623	2.64464	2.43743	2.30775	2.21825	2.15239	2.10169	2.06134	2.02839	2.00094	1.95772	1.91169	1.86236	1.83624	1.80901	1.78053	1.75063	1.71909	1.68564
18	3.00698	2.62395	2.41601	2.28577	2.19583	2.12958	2.07854	2.03789	2.00467	1.97698	1.93334	1.88681	1.83685	1.81035	1.78269	1.75371	1.72322	1.69099	1.65671
19	2.98990	2.60561	2.39702	2.26630	2.17596	2.10936	2.05802	2.01710	1.98364	1.95573	1.91170	1.86471	1.81416	1.78731	1.75924	1.72979	1.69876	1.66587	1.63077
20	2.97465	2.58925	2.38009	2.24893	2.15823	2.09132	2.03970	1.99853	1.96485	1.93674	1.89236	1.84494	1.79384	1.76667	1.73822	1.70833	1.67678	1.64326	1.60738
21	2.96096	2.57457	2.36489	2.23334	2.14231	2.07512	2.02325	1.98186	1.94797	1.91967	1.87497	1.82715	1.77555	1.74807	1.71927	1.68896	1.65691	1.62278	1.58615
22	2.94858	2.56131	2.35117	2.21927	2.12794	2.06050	2.00840	1.96680	1.93273	1.90425	1.85925	1.81106	1.75899	1.73122	1.70208	1.67138	1.63885	1.60415	1.56678
23	2.93736	2.54929	2.33873	2.20651	2.11491	2.04723	1.99492	1.95312	1.91888	1.89025	1.84497	1.79643	1.74392	1.71588	1.68643	1.65535	1.62237	1.58711	1.54903
24	2.92712	2.53833	2.32739	2.19488	2.10303	2.03513	1.98263	1.94066	1.90625	1.87748	1.83194	1.78308	1.73015	1.70185	1.67210	1.64067	1.60726	1.57146	1.53270
25	2.91774	2.52831	2.31702	2.18424	2.09216	2.02406	1.97138	1.92925	1.89469	1.86578	1.82000	1.77083	1.71752	1.68898	1.65895	1.62718	1.59335	1.55703	1.51760
26	2.90913	2.51910	2.30749	2.17447	2.08218	2.01389	1.96104	1.91876	1.88407	1.85503	1.80902	1.75957	1.70589	1.67712	1.64682	1.61472	1.58050	1.54368	1.50360
27	2.90119	2.51061	2.29871	2.16546	2.07298	2.00452	1.95151	1.90909	1.87427	1.84511	1.79889	1.74917	1.69514	1.66616	1.63560	1.60320	1.56859	1.53129	1.49057
28	2.89385	2.50276	2.29060	2.15714	2.06447	1.99585	1.94270	1.90014	1.86520	1.83593	1.78951	1.73954	1.68519	1.65600	1.62519	1.59250	1.55753	1.51976	1.47841
29	2.88703	2.49548	2.28307	2.14941	2.05658	1.98781	1.93452	1.89184	1.85679	1.82741	1.78081	1.73060	1.67593	1.64655	1.61551	1.58253	1.54721	1.50899	1.46704
30	2.88069	2.48872	2.27607	2.14223	2.04925	1.98033	1.92692	1.88412	1.84896	1.81949	1.77270	1.72227	1.66731	1.63774	1.60648	1.57323	1.53757	1.49891	1.45636
40	2.83535	2.44037	2.22609	2.09095	1.99682	1.92688	1.87252	1.82886	1.79290	1.76269	1.71456	1.66241	1.60515	1.57411	1.54108	1.50562	1.46716	1.42476	1.37691
60	2.79107	2.39325	2.17741	2.04099	1.94571	1.87472	1.81939	1.77483	1.73802	1.70701	1.65743	1.60337	1.54349	1.51072	1.47554	1.43734	1.39520	1.34757	1.29146
120	2.74781	2.34734	2.12999	1.99230	1.89587	1.82381	1.76748	1.72196	1.68425	1.65238	1.60120	1.54500	1.48207	1.44723	1.40938	1.36760	1.32034	1.26457	1.19256
∞	2.70554	2.30259	2.08380	1.94486	1.84727	1.77411	1.71672	1.67020	1.63152	1.59872	1.54578	1.48714	1.42060	1.38318	1.34187	1.29513	1.23995	1.16860	1.00000

Table F for Alfa = 0.05

/	$df_1=1$	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
$df_2=1$	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684	238.8827	240.5433	241.8817	243.9060	245.9499	248.0131	249.0518	250.0951	251.1432	252.1957	253.2529	254.3144
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532	19.3710	19.3848	19.3959	19.4125	19.4291	19.4458	19.4541	19.4624	19.4707	19.4791	19.4874	19.4957
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720	8.5494	8.5264
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6877	5.6581	5.6281
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.3650
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6689
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4045
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2962
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9604
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.3080	2.2341	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.8780
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660	2.3210	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165	1.8657	1.8117
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.9380	1.8894	1.8380	1.7831
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747	2.2036	2.1282	2.0476	2.0050	1.9605	1.9139	1.8648	1.8128	1.7570
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424	1.7896	1.7330
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.7110
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027	1.7488	1.6906
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6717
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6376
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290	1.3519	1.2539
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.3940	1.3180	1.2214	1.0000

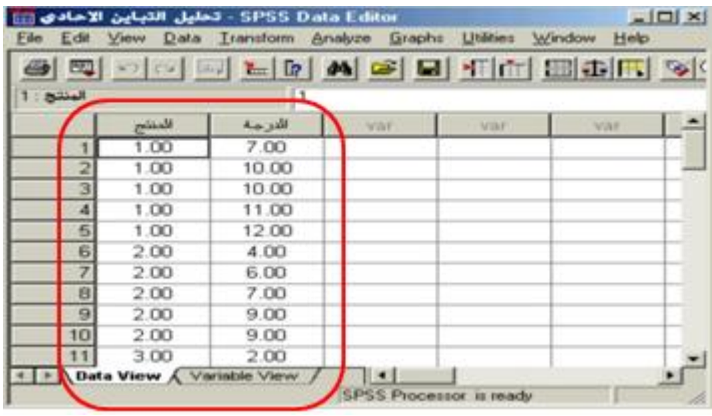
F table for Alfa = 0.025

/	df_1=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
df_2=1	647.7890	799.5000	864.1630	899.5833	921.8479	937.1111	948.2169	956.6562	963.2846	968.6274	976.7079	984.8668	993.1028	997.2492	1001.414	1005.598	1009.800	1014.020	1018.258
2	38.5063	39.0000	39.1655	39.2484	39.2982	39.3315	39.3552	39.3730	39.3869	39.3980	39.4146	39.4313	39.4479	39.4562	39.465	39.473	39.481	39.490	39.498
3	17.4434	16.0441	15.4392	15.1010	14.8848	14.7347	14.6244	14.5399	14.4731	14.4189	14.3366	14.2527	14.1674	14.1241	14.081	14.037	13.992	13.947	13.902
4	12.2179	10.6491	9.9792	9.6045	9.3645	9.1973	9.0741	8.9796	8.9047	8.8439	8.7512	8.6565	8.5599	8.5109	8.461	8.411	8.360	8.309	8.257
5	10.0070	8.4336	7.7636	7.3879	7.1464	6.9777	6.8531	6.7572	6.6811	6.6192	6.5245	6.4277	6.3286	6.2780	6.227	6.175	6.123	6.069	6.015
6	8.8131	7.2599	6.5988	6.2272	5.9876	5.8198	5.6955	5.5996	5.5234	5.4613	5.3662	5.2687	5.1684	5.1172	5.065	5.012	4.959	4.904	4.849
7	8.0727	6.5415	5.8898	5.5226	5.2852	5.1186	4.9949	4.8993	4.8232	4.7611	4.6658	4.5678	4.4667	4.4150	4.362	4.309	4.254	4.199	4.142
8	7.5709	6.0595	5.4160	5.0526	4.8173	4.6517	4.5286	4.4333	4.3572	4.2951	4.1997	4.1012	3.9995	3.9472	3.894	3.840	3.784	3.728	3.670
9	7.2093	5.7147	5.0781	4.7181	4.4844	4.3197	4.1970	4.1020	4.0260	3.9639	3.8682	3.7694	3.6669	3.6142	3.560	3.505	3.449	3.392	3.333
10	6.9367	5.4564	4.8256	4.4683	4.2361	4.0721	3.9498	3.8549	3.7790	3.7168	3.6209	3.5217	3.4185	3.3654	3.311	3.255	3.198	3.140	3.080
11	6.7241	5.2559	4.6300	4.2751	4.0440	3.8807	3.7586	3.6638	3.5879	3.5257	3.4296	3.3299	3.2261	3.1725	3.118	3.061	3.004	2.944	2.883
12	6.5538	5.0959	4.4742	4.1212	3.8911	3.7283	3.6065	3.5118	3.4358	3.3736	3.2773	3.1772	3.0728	3.0187	2.963	2.906	2.848	2.787	2.725
13	6.4143	4.9653	4.3472	3.9959	3.7667	3.6043	3.4827	3.3880	3.3120	3.2497	3.1532	3.0527	2.9477	2.8932	2.837	2.780	2.720	2.659	2.595
14	6.2979	4.8567	4.2417	3.8919	3.6634	3.5014	3.3799	3.2853	3.2093	3.1469	3.0502	2.9493	2.8437	2.7888	2.732	2.674	2.614	2.552	2.487
15	6.1995	4.7650	4.1528	3.8043	3.5764	3.4147	3.2934	3.1987	3.1227	3.0602	2.9633	2.8621	2.7559	2.7006	2.644	2.585	2.524	2.461	2.395
16	6.1151	4.6867	4.0768	3.7294	3.5021	3.3406	3.2194	3.1248	3.0488	2.9862	2.8890	2.7875	2.6808	2.6252	2.568	2.509	2.447	2.383	2.316
17	6.0420	4.6189	4.0112	3.6648	3.4379	3.2767	3.1556	3.0610	2.9849	2.9222	2.8249	2.7230	2.6158	2.5598	2.502	2.442	2.380	2.315	2.247
18	5.9781	4.5597	3.9539	3.6083	3.3820	3.2209	3.0999	3.0053	2.9291	2.8664	2.7689	2.6667	2.5590	2.5027	2.445	2.384	2.321	2.256	2.187
19	5.9216	4.5075	3.9034	3.5587	3.3327	3.1718	3.0509	2.9563	2.8801	2.8172	2.7196	2.6171	2.5089	2.4523	2.394	2.333	2.270	2.203	2.133
20	5.8715	4.4613	3.8587	3.5147	3.2891	3.1283	3.0074	2.9128	2.8365	2.7737	2.6758	2.5731	2.4645	2.4076	2.349	2.287	2.223	2.156	2.085
21	5.8266	4.4199	3.8188	3.4754	3.2501	3.0895	2.9686	2.8740	2.7977	2.7348	2.6368	2.5338	2.4247	2.3675	2.308	2.246	2.182	2.114	2.042
22	5.7863	4.3828	3.7829	3.4401	3.2151	3.0546	2.9338	2.8392	2.7628	2.6998	2.6017	2.4984	2.3890	2.3315	2.272	2.210	2.145	2.076	2.003
23	5.7498	4.3492	3.7505	3.4083	3.1835	3.0232	2.9023	2.8077	2.7313	2.6682	2.5699	2.4665	2.3567	2.2989	2.239	2.176	2.111	2.041	1.968
24	5.7166	4.3187	3.7211	3.3794	3.1548	2.9946	2.8738	2.7791	2.7027	2.6396	2.5411	2.4374	2.3273	2.2693	2.209	2.146	2.080	2.010	1.935
25	5.6864	4.2909	3.6943	3.3530	3.1287	2.9685	2.8478	2.7531	2.6766	2.6135	2.5149	2.4110	2.3005	2.2422	2.182	2.118	2.052	1.981	1.906
26	5.6586	4.2655	3.6697	3.3289	3.1048	2.9447	2.8240	2.7293	2.6528	2.5896	2.4908	2.3867	2.2759	2.2174	2.157	2.093	2.026	1.954	1.878
27	5.6331	4.2421	3.6472	3.3067	3.0828	2.9228	2.8021	2.7074	2.6309	2.5676	2.4688	2.3644	2.2533	2.1946	2.133	2.069	2.002	1.930	1.853
28	5.6096	4.2205	3.6264	3.2863	3.0626	2.9027	2.7820	2.6872	2.6106	2.5473	2.4484	2.3438	2.2324	2.1735	2.112	2.048	1.980	1.907	1.829
29	5.5878	4.2006	3.6072	3.2674	3.0438	2.8840	2.7633	2.6686	2.5919	2.5286	2.4295	2.3248	2.2131	2.1540	2.092	2.028	1.959	1.886	1.807
30	5.5675	4.1821	3.5894	3.2499	3.0265	2.8667	2.7460	2.6513	2.5746	2.5112	2.4120	2.3072	2.1952	2.1359	2.074	2.009	1.940	1.866	1.787
40	5.4239	4.0510	3.4633	3.1261	2.9037	2.7444	2.6238	2.5289	2.4519	2.3882	2.2882	2.1819	2.0677	2.0069	1.943	1.875	1.803	1.724	1.637
60	5.2856	3.9253	3.3425	3.0077	2.7863	2.6274	2.5068	2.4117	2.3344	2.2702	2.1692	2.0613	1.9445	1.8817	1.815	1.744	1.667	1.581	1.482
120	5.1523	3.8046	3.2269	2.8943	2.6740	2.5154	2.3948	2.2994	2.2217	2.1570	2.0548	1.9450	1.8249	1.7597	1.690	1.614	1.530	1.433	1.310
∞	5.0239	3.6889	3.1161	2.7858	2.5665	2.4082	2.2875	2.1918	2.1136	2.0483	1.9447	1.8326	1.7085	1.6402	1.566	1.484	1.388	1.268	1.000

حساب تحليل التباين الأحادي من خلال برنامج الـ SPSS One Way Analysis of Variance

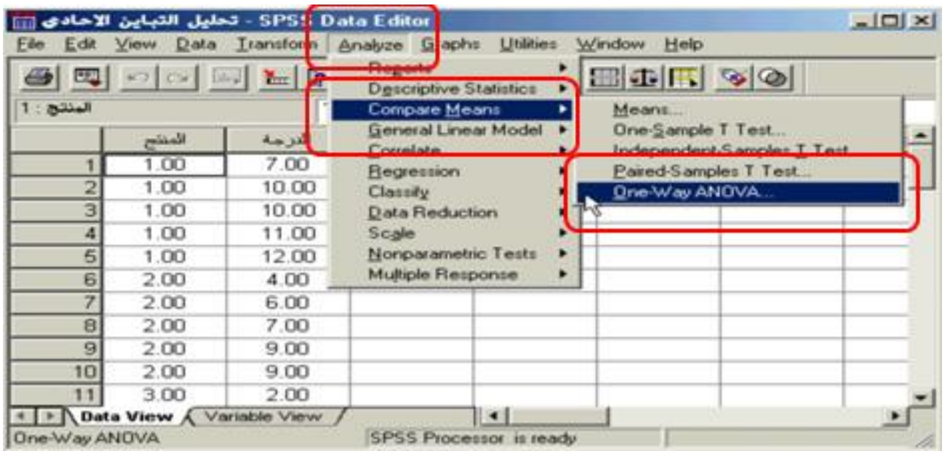
لغرض حساب قيمة تحليل التباين الأحادي One Way Analysis of Varianceلنفس المثال السابق من خلال استخدام برنامج الـ SPSS نتبع الخطوات التالية :

- قم بإدخال البيانات المراد تحليلها من خلال شاشة تحرير البيانات Data Editor بالطريقة المناسبة كالتالي:

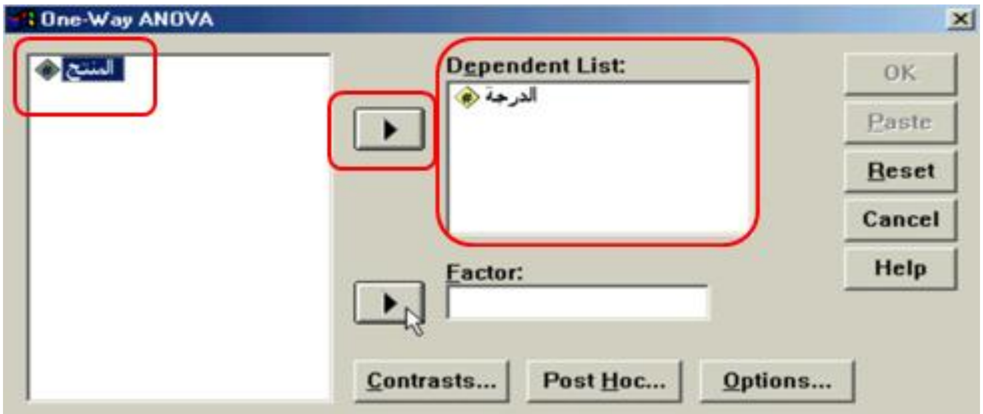


لاحظ أنه تم إدخال البيانات بطريقة مناسبة للتحليل الذي تم اختياره، حيث أدخلت مستويات المتغير المستقل في عمود وأطلق عليه اسم "المنتج"، وأدخلت درجات التقييم للمنتج تحت عمود آخر أطلق عليه اسم "الدرجة"

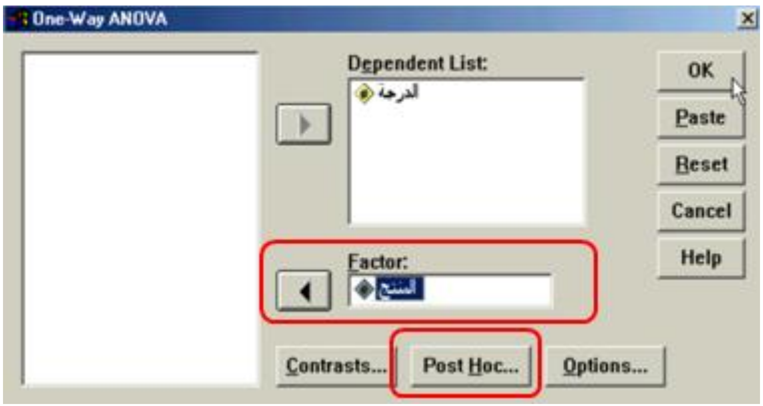
- من القائمة "تحليل" Analyze اختر الأمر "مقارنة المتوسطات" Compare Means فتظهر قائمة أوامر فرعية اختر منها "تحليل التباين الأحادي" One-Way ANOVA كالتالي:



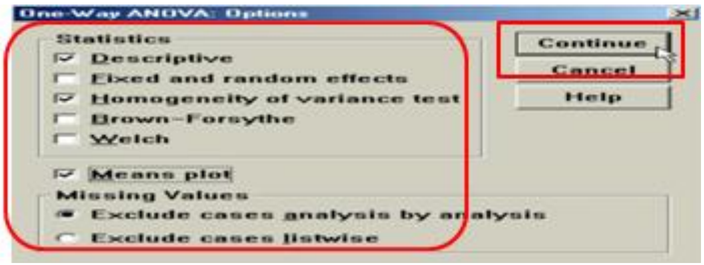
- بعد اختيار الأمر "تحليل التباين الأحادي" One-Way ANOVAسوف يظهر لك صندوق الحوار التالي:



- من قائمة المتغيرات في الجهة اليسرى من صندوق الحوار حدد المتغير المستقل والمتغير التابع المراد إجراء تحليل التباين الأحادي لها، ونقلها إلى المستطيل الخاص بـ "المتغيرات التابعة" Dependent List من خلال النقر على السهم الذي يظهر مقابل المستطيل الخاص بـ "متغيرات الاختبار"، ستلاحظ انتقال المتغير مباشرة في المستطيل "المتغيرات التابعة" Dependent List (في هذا المثال المتغير التابع هو "الدرجة")، كرر نفس الإجراء مع المتغير المستقل Independent Variable وقم بنقله إلى المستطيل الخاص بـ "العامل" Factor (في هذا المثال المتغير المستقل هو "المنتج") كما يبدوا ذلك في الشكل التالي:



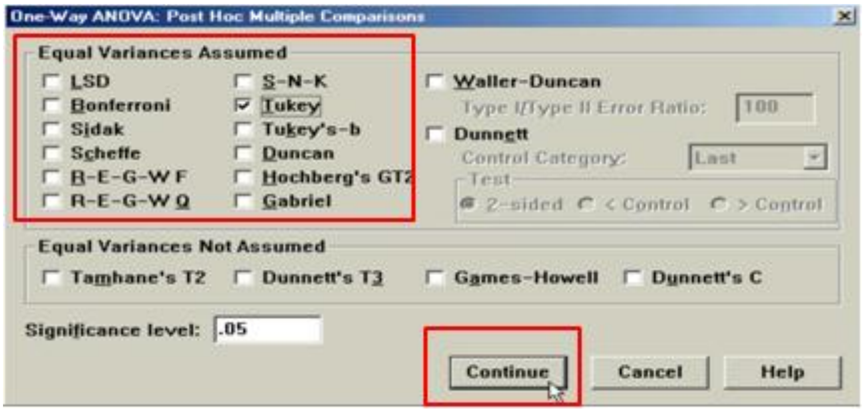
- أنقر على زر "خيارات" Options في الجهة السفلية اليمنى من صندوق الحوار السابق وذلك عند الرغبة في حساب الخصائص الأساسية للمتغيرات موضع الدراسة Statistics ، وعند الرغبة في عرض المتوسطات من خلال رسم بياني Means Plot ، وكذلك كيفية التعامل مع القيم المفقودة Missing Values ، وبعد الانتهاء من التعديل على هذا الصندوق الحواري أنقر على زر "استمرار" Continue .



- أنقر على زر "المقارنات البعدية المتعددة" Post Hoc في الجهة السفلية من صندوق الحوار السابق وذلك عند الرغبة في حساب المقارنات البعدية بين متوسطات المتغيرات موضع الدراسة والكشف عن مواقع الفروق وذلك في حالة كون قيمة *F* ذات دلالة إحصائية، وهناك العديد من الأساليب الإحصائية المعدة لهذا الغرض اشهرها:

١. طريقة شيفيه Scheffe وتستخدم هذه الطريقة في إجراء جميع المقارنات بين الأوساط وهي الطريقة المفضلة في حالة كون حجوم الخلايا غير متساوية أو عند الرغبة في إجراء مقارنات معقدة كأن نقارن ثلاث مجتمعات بمجتمع واحد، أو مجتمعين مقابل مجتمعين أو غيرها من مثل هذه المقارنات
٢. طريقة توكي Tukey وتستخدم هذه الطريقة لمقارنة جميع الأزواج الممكنة للأوساط موضع الدراسة سواء كانت حجوم الخلايا متساوية أو غير متساوية (في حالة عدم تساوي حجوم الخلايا يستخدم الوسط التوافقي لحجم الخلية)، ويعتبر هذا الاختبار أدق من اختبار شيفيه Scheffe لمقارنة أزواج الأوساط
٣. طريقة نيومن-كولز Newman-Keuls (S-N-K) وتفيد هذه الطريقة في المقارنة بين أزواج الأوساط فقط، وهي تستند كما هي الحال في طريقة توكي على توزيع مدى ستيودنتايز Studentize range ، وهي طريقة جيدة وقوية للكشف عن الفروق بين الأوساط في حالة تساوي حجوم الخلايا أو عدم تساويها (في حالة عدم تساوي حجوم الخلايا يستخدم الوسط التوافقي لحجم الخلية كما هو الحال في اختبار توكي)، ويعتبر هذا الاختبار (نيومن-كولز) أدق الاختبارات البعدية للكشف عن الفروق بين أزواج الأوساط، يليه اختبار توكي ثم بعد ذلك اختبار شفيه

وفي المثال الحالي تم اختيار طريقة توكي Tukey للمقارنة البعدية بين أزواج الأوساط (ويمكن اختيار أكثر من طريقة في وقت واحد) كما يبدوا ذلك في الشكل التالي:



وبعد الانتهاء من التعديل على هذا الصندوق الحواري أنقر على زر "استمرار" Continue ، وستنتقل إلى صندوق الحوار الرئيسي، ثم أنقر بعد ذلك على زر "موافق" OK في صندوق الحوار الرئيسي سيؤدي ذلك إلى تنفيذ الاختبار، وستلاحظ ظهور النتائج في شاشة المخرجات كالتالي:

Oneway

Descriptives

الدرجة

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
المنتج١	5	10.0000	1.87083	.83666	7.6771	12.3229	7.00	12.00
المنتج٢	5	7.0000	2.12132	.94868	4.3660	9.6340	4.00	9.00
المنتج٣	5	4.0000	2.34521	1.04881	1.0880	6.9120	2.00	7.00
Total	15	7.0000	3.20713	.82808	5.2239	8.7761	2.00	12.00

ANOVA

الدرجة

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	90.000	2	45.000	10.000	.003
Within Groups	54.000	12	4.500		
Total	144.000	14			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: الدرجة
Tukey HSD

(I) المنتج	(J) المنتج	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
المنتج١	المنتج٢	3.0000	1.34164	.105	-.5793	6.5793
	المنتج٣	6.0000*	1.34164	.002	2.4207	9.5793
المنتج٢	المنتج١	-3.0000	1.34164	.105	-6.5793	.5793
	المنتج٣	3.0000	1.34164	.105	-.5793	6.5793
المنتج٣	المنتج١	-6.0000*	1.34164	.002	-9.5793	-2.4207
	المنتج٢	-3.0000	1.34164	.105	-6.5793	.5793

*. The mean difference is significant at the .05 level.

نلاحظ أن برنامج الـ SPSS قام مباشرة بحساب الإحصاءات الأساسية للبيانات مثل المتوسط الحسابي والانحراف المعياري وغيرها من الإحصاءات ذات العلاقة.

ثم بعد ذلك قام البرنامج بحساب قيمة (*F*) للمتغيرات موضع الدراسة في الجدول المعنون بـ ANOVA ، ومن هذه النتائج نلاحظ أن قيمة (*F*) المحسوبة = 10 ، ودرجات الحرية df (2 ، 12) ، والقيمة الحرجة .Sig=0.003 ، وبما أن القيمة الحرجة لـ *F* .Sig في الجدول (0.003) أصغر من قيمة α = 0.05

فإننا نستنتج أن الفرضية الصفرية تكون مرفوضة ، أي يوجد اختلاف بين متوسطي مجتمعين على الأقل من المجتمعات التي قُيِّمة من المستهلكين. ولمعرفة بين أي من المنتجات تكون الفروق قمنا بحساب اختبار المقارنات البعدية Post Hoc Comparisons لتحديد هذه الفروق، وقد أظهرت النتائج وجود فروق ذات دلالة إحصائية بين منتج (1) والذي متوسطه 10 ومنتج (3) والذي متوسطه 4 وذلك لصالح منتج (1).

كيف يمكن حساب المقارنات المتعددة Multiple Comparisons

هي طريقة لإجراء عدد من الاختبارات الأولية لتحديد الفروق المعنوية للمتوسطات حال رفض فرضية العدم.

حال رفض فرضية العدم عند مستوى معنوية فلا دليل على وجود فروق معنوية بين المتوسطات ولا نعرف أي منها يختلف عن متوسطات (العينات) ومن حيث قيمة F معنوية ففرق المشاهدات بين المتوسطات معنوي أيضاً وعلى العموم توجد عدة طرق إحصائية لعمل اختبار بهذا الخصوص مثل:

- Least Significant difference,
- Scheffe Test,
- Student-Newman Keuls,

- Tukey's Procdure,
- Duncans New Multiple range

وسنتعرض لواحد من هذه الاختبارات بصورة مبسطة مع مثال.

طريقة Least Significant difference لمقارنة بين متوسطين

يرمز لها LSD وتحدد حال استخدام برنامج SPSS وتعرف Fisher's LSD وتختبر كل الأزواج فإن تمَ رفض فرضية العدم (تساوي المتوسطات) مقابل الفرضية البديلة (عدم تساوي المتوسطات) فإذا كان الفرق أكبر من LSD فنرفض فرض العدم، والصيغة الرياضية لحساب LSD حيث MSE متوسط مجموع المربعات (التباين) داخل المجموعات) هي:

$$LSD = t_{2n-2,\alpha/2}\sqrt{MSE \times \frac{2}{n}}$$

مثال:

إذا كانت لدينا البيانات التالية والتي تمثل بيانات أربع مجموعات تم استطلاع آراؤها حول موضوع ما:

A	B	C	D
8	4	5	6
7	3	3	5
9	6	4	6
5	5	5	4
6	2	3	3
7	7	2	4
42	27	22	28

المطلوب:هل هناك فروق بين آراء هذه المجموعات الأربع ولصالح من هذا الفرق؟

الحل:

- وضع فرض العدم والفرض البديل
  صياغة الفرضية الصفرية كالتالي: $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
  في حين تفترض الفرضية البديلة التالي: **متوسطان على الأقل غير متساويين : H_A**
- تحديد مستوى الدلالة (α): وتحدد مستويات المعنوية سلفاً وهي عادة 0.05 أو 0.01 وليكون α = 0.05
- حساب إحصائية الاختبار (F) وذلك من خلال اتباع الخطوت التالية:
  - نجمع قيم كل متغير للحصول على $\sum X$ لكل متغير
  - نربع كل درجة في كل متغير للحصول على X^2 لكل متغير
  - نجمع قيم مربع كل درجة للحصول على $\sum X^2$ لكل متغير
  - نربع مجموع كل متغير للحصول على $\left(\sum X\right)^2$ لكل متغير
  - نحسب متوسط كل متغير من خلال العلاقة: $\overline{X} = \frac{\sum x}{n}$

	A	A^2	B	B^2	C	C^2	D	D^2	
	8	64	4	16	5	25	6	36	
	7	49	3	9	3	9	5	25	
	9	81	6	36	4	16	6	36	
	5	25	5	25	5	25	4	16	
	6	36	2	4	3	9	3	9	
	7	49	7	49	2	4	4	16	
$\sum X$	42		27		22		28		119
$\sum X^2$		304		139		88		138	
$\left(\sum X\right)^2$	1764		729		484		784		3761
$\overline{X} = \frac{\sum x}{n}$	42/6 =7		27/6 =4.5		22/6 =3.7		28/6 =4.7		

- المتوسط الحسابي لـ X_1 = $\overline{X_A} = \frac{42}{6} = 7$
- المتوسط الحسابي لـ X_2 = $\overline{X_B} = \frac{27}{6} = 4.5$
- المتوسط الحسابي لـ X_3 = $\overline{X_C} = \frac{22}{6} = 3.7$
- المتوسط الحسابي لـ X_4 = $\overline{X_D} = \frac{28}{6} = 4.7$
- مجموع المربعات الكلي Total Sum of Squares =

$$Total..SS = \sum X^2 - \frac{\left(\sum X\right)^2}{(n_g)(k)} = 669 - \frac{(119)^2}{24} = 669 - 590.042 = 78.958$$

- مجموع المربعات داخل المجموعات Within Sum of Squares =

$$Within..SS = \sum\left[\sum X_g^2 - \frac{\left(\sum X_g\right)^2}{n_g}\right]$$

$$\sum x_1^2 = 304 - \frac{(42)^2}{6} = 304 - \frac{1764}{6} = 304 - 294 = 10$$

$$\sum x_2^2 = 139 - \frac{(27)^2}{6} = 139 - \frac{729}{6} = 139 - 121.5 = 17.5$$

$$\sum x_3^2 = 88 - \frac{(22)^2}{6} = 88 - \frac{484}{6} = 88 - 80.7 = 7.3$$

$$\sum x_4^2 = 138 - \frac{(28)^2}{6} = 138 - \frac{784}{6} = 138 - 130.7 = 7.3$$

- نقوم بعد ذلك بجمع نواتج هذه المعادلات لنحصل على مجموع المربعات داخل المجموعات كالتالي:

Within sum of squares = 10+17.5+7.3+7.3=42.1

- ثم نقوم بحساب مجموع المربعات بين المجموعات من خلال العلاقة التالية:

Between SS= Total SS - Within SS

Between SS = 78.958-42.1=36.858

• **ثم نقوم بعمل جدول تحليل التباين كالتالي:**

Source of Variance	SS	d f	MS	F	$f_{0.05,3,20}$
SSB	SST – SSW = 36.858	K – 1 = 4 – 1 = 3	12.264	5.818	3.10
SSW - Error	42.1	K(n – 1) = 4(6 – 1) = 20	2.108		
TOTAL	78.958	Kn – 1 = 24 – 1 = 23			

قيمة F المحسوبة أكبر من قيمة F الجدولية (7.64 > 3.1) فالمتوسطات غير متساوية،

وبإجراء اختبار المقارنات البعدية LSD نجد التالي:

$$LSD = t_{2n-2,\alpha/2}\sqrt{MSE \times \frac{2}{n}}$$

MSE × 2 / n = 2.34 × 2 ÷ 6 = 0.78 , $\sqrt{0.78} = 0.883$

$t_{2n-2,\alpha/2} = t_{10,\ 0.025} = 2.228$

LSD = 2.228 × 0.883 = 1.967

حيث أن : MSE تعني متوسط مربعات الخطأ ويتم حسابها من خلال المعادلة التالية:

$$MSE = \frac{SSW}{n-k} = \frac{42.1}{24-6} = \frac{42.1}{18} = 2.34$$

والجدول التالي يوضح متوسطات الفروق بين المجموعات، حيث وجود (*) على العدد يعني وجود فرق معنوي بين المتوسطين ولدينا أربع متوسطات (7، 4.5، 3.7، 4.7) كما أوضحنا ذلك سابقا، وتقارن هذه المتوسطات بقيمة LSD = 1.967 التي تم الحصول عليها من من المعادلة السابقة

	$\bar{X}_1 = 7$	$\bar{X}_2 = 4.5$	$\bar{X}_3 = 3.667$	$\bar{X}_4 = 4.667$
$\bar{X}_1 = 7$	0	2.5*	3.333*	2.333*
$\bar{X}_2 = 4.5$	—	0	0.833	– 0.167
$\bar{X}_3 = 3.667$	—	—	0	– 1
$\bar{X}_4 = 4.667$	—	—	—	0

المحاضرة 12 (الجزء 2)

تابع اختبار الفروض الإحصائية المعلمية

بالرغم من أن مقاييس العلاقة تختلف عما سبقها من مقاييس ، فهي تتعلق بدراسة العلاقة بين متغيرين (الإنتاجية والجودة) مثلا، بينما المقاييس السابقة فتهتم بدراسة الفروق بين المتغيرات .

وعندما نقول **مقاييس العلاقة** نعني بذلك تلك المقاييس التي تبين درجة العلاقة والارتباط بين متغيرين أو أكثر مثلا، كأن يكون الهدف معرفة هل هناك علاقة بين مستوى الإنتاجية وجودة المنتج في مصنع ما؟، أي هل كلما زادت الإنتاجية تقل جودة المنتج أو العكس .

معامل الارتباط: هو تعبير يشير إلى المقياس الإحصائي الذي يدل على مقدار العلاقة بين المتغيرات سلبية كانت أم إيجابية، وتتراوح قيمته بين الارتباط الموجب التام (+1) وبين الارتباط السالب التام (-1) .

العلاقة الطردية بين المتغيرات: هو تعبير يشير إلى تزايد المتغيرين المستقل والتابع معا، فإذا كانت الإنتاجية مرتفعة، ومستوى الجودة مرتفع، يقال حينئذ أن بينهما ارتباط موجب، وأعلى درجة تمثله هي (+1) .

العلاقة العكسية بين المتغيرات: هو تعبير يشير إلى تزايد في متغير يقابله تناقص في المتغير الآخر، فإذا كانت الإنتاجية منخفضة ومستوى الجودة مرتفع، يقال حينئذ أن بينهما ارتباط سالب، وأعلى درجة تمثله هي (-1)

ومن الطبيعي ملاحظة أن الارتباط الكامل لا وجود له في الظواهر الطبيعية، وأن معامل الارتباط الناتج في الأبحاث والدراسات الإنسانية والاجتماعية يكون عادة كسرا موجبا أو سالبا. والجدول التالي يوضح أنواع العلاقات بين المتغيرات كما يصفها معامل الارتباط:

نوع العلاقة	قيمة معامل الارتباط
طردية كاملة	1+
طردية ناقصة	+ كسر(قيمة موجبة)
صفرية	صفر
عكسية ناقصة	- كسر (قيمة سالبة)
عكسية كاملة	1-

إن معامل الارتباط التام الموجب (+1) يعنى التغير في اتجاه واحد في كلا الظاهرتين مع بقاء الأوضاع النسبية لوحدات الظاهرة ثابتة، سواء كان هذا التغير في اتجاه الزيادة (أي زيادة قيم الظاهرة الأولى تتبعها زيادة في قيم الظاهرة الأخرى)، أو في اتجاه النقص (أي نقص قيم الظاهرة الأولى يتبعها نقص في قيم الظاهرة الأخرى) .

طرق التعرف على العلاقة بين متغيرين وحسابها

أولا: طريقة شكل الانتشار Scatter Diagram :

هناك وسيلة مبدئية يعرف الباحث من خلالها نوع الارتباط بين المتغيرين وما إذا كان الارتباط قوياً وضعيفاً أو منعدماً، وما إذا كانت العلاقة خطية أو غير خطية، موجبة أو سالبة. هذه الوسيلة هي " شكل الانتشار " والتي تصلح إذا كان المتغيران كميين. وجدير بالذكر أن هذه وسيلة مبدئية تساعد فقط في معرفة نوع الارتباط ولا تعتبر بديلاً عن الطرق الإحصائية التي سوف نتناولها بالتفصيل في هذه المحاضرة.

والمقصود **بشكل الانتشار** هو تمثيل قيم الظاهرتين بيانياً على المحورين، المتغير الأول X على المحور الأفقي، والمتغير الثاني Y على المحور الرأسي، حيث يتم تمثيل كل زوج Pair من القيم بنقطة، فنحصل على شكل يمثل كيفية انتشار القيم على المستوى، وهو الذي يسمى شكل الانتشار. وطريقة انتشار القيم تدل على وجود أو عدم وجود علاقة بين المتغيرين ومدى قوتها ونوعها. فإذا كانت تتوزع بشكل منتظم دل ذلك على وجود علاقة (يمكن استنتاجها)، أما إذا كانت النقط مبعثرة ولا تنتشر حسب نظام معين دل ذلك على عدم وجود علاقة بين المتغيرين أو أن العلاقة بينهما ضعيفة.

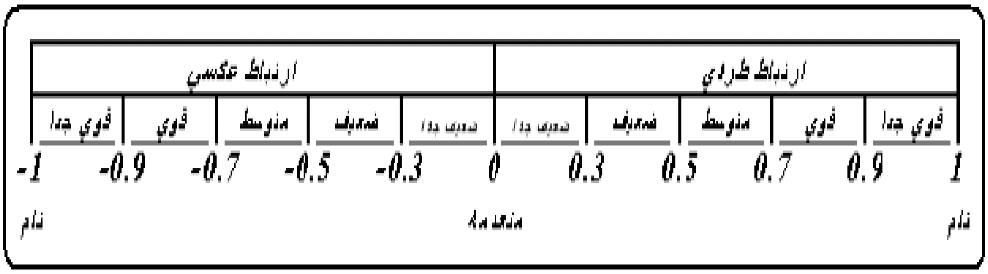
ثانيا: معامل الارتباط Correlation Coefficient :

يقاس الارتباط بين متغرين بمقياس إحصائي يسمى " معامل الارتباط " ويعكس هذا المقياس درجة أو قوة العلاقة بين المتغيرين واتجاه هذه العلاقة. وتنحصر قيمة معامل الارتباط بين + 1، 1 - .

- فإذا كانت قيمة معامل الارتباط تساوي + 1 فمعنى ذلك أن الارتباط بين المتغيرين طردي تام، وهو أقوى أنواع الارتباط الطردي بين متغيرين.
- وإذا كانت قيمة معامل الارتباط تساوي – 1 فمعنى ذلك أن الارتباط بين المتغيرين عكسي تام، وهو أقوى أنواع الارتباط العكسي بين متغيرين.
- وإذا كانت قيمة معامل الارتباط تساوي صفر، فمعنى ذلك أنه لا يوجد ارتباط بين المتغيرين.
- وكلما اقتربت قيمة معامل الارتباط من + 1 أو – 1 كلما كان الارتباط قوياً، وكلما اقترب من الصفر كلما كان الارتباط ضعيفاً.

والخلاصة:

أنه كلما كانت العلاقة قوية بين المتغيرين كلما اقترب معامل الارتباط من + 1 أو – 1 فإذا وصلت قيمة المعامل إلى + 1 أو – 1 كان الارتباط تاماً بين المتغيرين. وأنه كلما كانت العلاقة ضعيفة بين المتغيرين كلما اقترب معامل الارتباط من الصفر، فإذا وصلت قيمة المعامل إلى الصفر كان الارتباط منعدماً بين المتغيرين. ومعنى ذلك أيضاً أنه لا يوجد ارتباط بين متغيرين تكون قيمة المعامل فيه أكبر من + 1 ولا أصغر من – 1. ويمكن تمثيل قوة العلاقة بالشكل التالي:



معامل بيرسون للارتباط الخطى البسيط Simple Correlation

يفترض بيرسون Pearson أن المتغيرين كميان، وأن العلاقة بينهما خطية (أي تأخذ شكل خط مستقيم، ويرى بيرسون أن أفضل مقياس للارتباط بين متغيرين قد يختلفان في وحدات القياس و / أو في مستواهما العام (مثل الارتباط بين العمر والدخل) حيث يقاس العمر بالسنوات ويقاس الدخل بالعملة، بالريال أو الدولار.. كما أن المستوى العام للعمر – أي متوسط العمر – قد يساوي أربعين عاماً. فبينما المستوى العام – أي متوسط – الدخل الشهري قد يكون خمسة آلاف ريال مثلاً.

وبالتالي فإن أفضل مقياس للارتباط بين مثل هذين المتغيرين – حسب رأي بيرسون – هو عن طريق حساب انحرافات كل من المتغيرين عن وسطه الحسابي وقسمة هذه الانحرافات على الانحراف المعياري لكل منهما، فنحصل على ما يسمى بالوحدات المعيارية لكل متغير. ويكون معامل ارتباط بيرسون هو " متوسط حاصل ضرب هذه الوحدات المعيارية ". ومعامل الارتباط يكون بدون تمييز.

إن الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين، ويرمز له في حالة المجتمع بالرمز ρ (رو)، وفي حالة العينة بالرمز r ، وحيث أننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع، سوف نهتم بحساب معامل الارتباط في العينة كتقدير لمعامل الارتباط في المجتمع.

ويتم حساب معامل ارتباط بيرسون من خلال العلاقـة التالية:

$$r = \frac{\sum XY - \frac{\left(\sum X\right)\left(\sum Y\right)}{n}}{\sqrt{\left(\sum X^2 - \frac{\left(\sum X\right)^2}{n}\right)\left(\sum Y^2 - \frac{\left(\sum Y\right)^2}{n}\right)}}$$

حيث :

$\sum XY$ تعني مجموع حاصل ضرب كل قيمة من X في Y.

$\left(\sum X\right)$ تعني مجموع قيم المتغير X .

$\left(\sum Y\right)$ تعني مجموع قيم المتغير Y .

$\sum X^2$ تعني مجموع مربع قيم المتغير X .

$\left(\sum X\right)^2$ تعني مربع مجموع قيم المتغير X .

$\sum Y^2$ تعني مجموع مربع قيم المتغير Y .

$\left(\sum Y\right)^2$ تعني مربع مجموع قيم المتغير Y .

n عدد قيم الدراسة (عدد الأزواج المطلوب حساب الارتباط بينها) .

وقد تم اخذ طريقة حساب معامل ارتباط بيرسون في مقرر الإحصاء في الإدارة

معامل سبيرمان لارتباط الرتب Spearman’s Rank Correlation

قد يرغب الباحث في حساب معامل الارتباط بين رتب المتغيرين وليس بين القيم ذاتها، فقد يكون المتغيران وصفيين ترتيبين Ordinal أو أن يكون أحد المتغيرين كمياً بينما الأخر وصفياً ترتيبياً، أو أن يكون المتغيران كميين، ويكون اهتمام الباحث منصباً على الرتب أكثر من القيم. ففي انتخابات مجلس الشيوخ أو النواب الأمريكي مثلاً، يعتبر المرشح الأول هو من حصل على أعلى الأصوات بغض النظر عن عددها، والذي يحصل على عدد أصوات أقل منه مباشرة هو الثاني.. وهكذا.

فإذا كانت رتب المتغيرين تسير في الاتجاه نفسه: بمعنى أن الرتب الأعلى للمتغير الأول تناظرها رتب أعلى للمتغير الثاني كانت العلاقة طردية بينهما. وإذا كانت الرتب الأعلى للمتغير الأول تناظرها رتب أدنى للمتغير الثاني كانت العلاقة بينهما عكسية. ففي مثالنا السابق عن العلاقة بين دخل الناخب وعمره، كان الناخب الأكبر عمراً (بصفة عامة) هو الأعلى دخلاً، فمن الواضح أن العلاقة بينهما طردية، أما إذا كان الناخب الأكبر عمراً (بصفة عامة) هو الأقل مشاركة في العمل السياسي فإننا في هذه الحالة نكون أمام علاقة عكسية.

ولحساب معامل ارتباط الرتب هناك طرق مختلفة أهمها معاملي سبيرمان وكيندال.

لحساب معامل سبيرمان لارتباط الرتب نقوم بترتيب كل من المتغيرين ترتيباً تصاعدياً أو تنازلياً (أما تصاعدياً لكلا المتغيرين أو تنازلياً لكليهما). وفي حالة الترتيب التصاعدي تأخذ أقل قيمة من قيم المتغير الرتبة رقم 1، والقيمة الأعلى منها مباشرة الرتبة رقم 2 وهكذا (بالنسبة لكل من المتغيرين). أما في حالة الترتيب التنازلي تأخذ أكبر قيمة من قيم المتغير الرتبة رقم 1، والقيمة الأقل منها مباشرة الرتبة رقم 2 وهكذا (بالنسبة لكل من المتغيرين). وعند تساوي قيمتين (أو أكثر) من قيم المتغير نعطي كل قيمة رتبة مختلفة (كما لو كانت القيم غير متساوية) ثم نحسب متوسط هذه الرتب، ويعطى هذا المتوسط لكل من هذه القيم المتساوية.

وبعد ترتيب المتغيرين نحسب الفروق بين رتب كل من المتغيرين (ونرمز للفروق بالرمز d) ثم نقوم بتربيع هذه الفروق ونحصل على مجموعها أي نحصل على $\sum d^2$ ثم نعوض في معامل سبيرمان لارتباط الرتب والذي يأخذ الشكل التالي:

$$r_s = 1 - \frac{6\left(\sum d^2\right)}{n\left(n^2 - 1\right)}$$

حيث: $\sum d^2$ هو مجموع مربعات الفروق بين رتب المتغيرين ، n هي عدد أزواج القيم.

وقد تم اخذ طريقة حساب معامل ارتباط سبيرمان في مقرر الإحصاء في الإدارة

الإرتباط الجزئي Partial Correlation

هو عبارة عن مقياس لقوة واتجاه الارتباط بين متغيرين كميين بعد استبعاد اثر متغير كمي ثالث ، حيث يلاحظ انه بالرغم من ان قيمة معامل الارتباط بيرسون قد تكون كبيرة ولكن لا يمكن الاعتماد عليها لكونه يعتمد في قياسه على متغيرين فقط ، فقد يوجد متغير ثالث يؤثر في المتغيرين ولهذا برزت اهمية معامل الارتباط الجزئي.

فمثلا يمكن قياس قوة الارتباط بين مستوى الطلبة في الجامعات والبيئة الجامعية بعد استبعاد عدد ساعات الدراسة لكل طالب.

ففي كثي من الأحيان لا يستطيع الباحث في كثير من البحوث التي يجريها ضبط كل المتغيرات إما عن صعوبة وعوائق ميدانية أو لنسيان إجراء عملية الضبط والتثبيت للمتغيرات أثناء الخطوات الأولى من البحث. لذا فإن الباحث يحتاج إلى اسلوب إحصائي يفيده في عزل تأثير هذا المتغير أو المتغيرات التي لم يثبتها على الظاهرة المدروسة من حيث علاقاته بمتغيرات أخرى.

ويتم حساب الإرتباط الجزئي من خلال حساب الإرتباطات الثنائية بين متغيرات الدراسة (على الباحث أن يستخدم معامل الإرتباط المناسب لعدد العينة ولطبيعة توزيع المتغيرات).

أي أن بإمكان الباحث استخدام **معامل ارتباط بيرسون** أو **معامل ارتباط سبيرمان** أو غير ذلك من معاملات الارتباط تبعا كما ذكر لطبيعة توزيع متغيرات الدراسة.

مثال:

أراد باحث أن يدرس العلاقة بين التحصيل الدراسي والغياب لدى مجموعة من الطلبة، ومن المعروف أنه إلى جانب الغياب فإن طريقة التدريس للطالب تؤثر في تحصيله الدراسي أيضا، فإذا استطاع الباحث أن يضبط هذا المتغير (المتغير الخاص بطريقة التدريس) أثناء إجرائه للتجربة، ويختار الطلبة من بين الذين يتعلمون بطريقة تدريس واحدة فإنه يكون بذلك قد عزل تأثير هذا المتغير.

أما إذا لم يستطع الباحث اختيار الطلبة من الذين يخضعون لطريقة تدريس واحدة، وكان الطلبة يتلقون تدريسهم وفقا لطرق تدريس مختلفة، فإنه بذلك يكون في حاجة لمعامل الإرتباط الجزئي لكي يعزل تأثير متغير طريقة التدريس في العلاقة بين التحصيل الدراسي والغياب، والبيانات التالية توضح هذا المثال:

الطلبة	الغياب (1)	التحصيل (2)	طريقة التدريس (3)
1	70	15	13
2	110	13	20
3	120	11	55
4	95	13	80
5	105	08	06

المطلوب: حساب معامل الارتباط بين الغياب والتحصيل مع تثبيت طريقة التدريس؟

الحل:

لغرض حساب معامل الارتباط بين الغياب والتحصيل مع تثبيت طريقة التدريس لا بد من حساب معاملات الإرتباط بين المتغيرات الثلاثة السابقة كالتالي:

- معامل ارتباط بيرسون بين الغياب والتحصيل الدراسي ونرمز له 1.2 أي معامل الإرتباط بين المتغير (1) والمتغير (2)
- معامل ارتباط بيرسون بين الغياب وطريقة التدريس ونرمز له 1.3 أي معامل الإرتباط بين المتغير (1) والمتغير (3)
- معامل ارتباط بيرسون بين التحصيل الدراسي وطريقة التدريس ونرمز له 2.3 أي معامل الإرتباط بين المتغير (2) والمتغير (3)

ويتم حساب معامل ارتباط بيرسون من خلال العلاقــة التالية:

$$r = \frac{\sum XY - \frac{\left(\sum X\right)\left(\sum Y\right)}{n}}{\sqrt{\left(\sum X^2 - \frac{\left(\sum X\right)^2}{n}\right)\left(\sum Y^2 - \frac{\left(\sum Y\right)^2}{n}\right)}}$$

حيث :

$\sum XY$ تعني مجموع حاصل ضرب كل قيمة من X في Y.

$\left(\sum X\right)$ تعني مجموع قيم المتغير X .

$\left(\sum Y\right)$ تعني مجموع قيم المتغير Y .

$\sum X^2$ تعني مجموع مربع قيم المتغير X .

$\left(\sum X\right)^2$ تعني مربع مجموع قيم المتغير X .

$\sum Y^2$ تعني مجموع مربع قيم المتغير Y .

$\left(\sum Y\right)^2$ تعني مربع مجموع قيم المتغير Y .

n عدد قيم الدراسة (عدد الأزواج المطلوب حساب الارتباط بينها) .

معامل ارتباط بيرسون بين الغياب والتحصيل الدراسي ونرمز له $r_{1.2}$

XY	Y^2	X^2	التحصيل الدراسي (2) Y	الغياب (1) X	الطلبة
1050	225	4900	15	70	1
1430	169	12100	13	110	2
1320	121	14400	11	120	3
1235	169	9025	13	95	4
840	64	11025	8	105	5
5875	748	51450	60	500	**المجموع**

$$r_{1.2} = \frac{5875 - \frac{(500)(60)}{5}}{\left[\sqrt{(51450) - \frac{(500)^2}{5}}\right]\left[\sqrt{(748) - \frac{(60)^2}{5}}\right]} = \frac{5875 - 6000}{\left(\sqrt{51450 - 50000}\right)\left(\sqrt{748 - 720}\right)}$$

$$= \frac{-125}{\left(\sqrt{1450}\right)\left(\sqrt{28}\right)} = \frac{-125}{(38.08)(5.292)} = \frac{-125}{201.519} = -0.620$$

معامل ارتباط بيرسون بين الغياب وطريقة التدريس ونرمز له $r_{1.3}$

XY	Y^2	X^2	طريقة التدريس (3) Y	الغياب (1) X	الطلبة
910	1169	4900	13	70	1
2200	400	12100	20	110	2
6600	3025	14400	55	120	3
7600	6400	9025	80	95	4
630	36	11025	6	105	5
17940	10030	51450	174	500	**المجموع**

$$r_{1.3}=\frac{17940-\frac{(500)(174)}{5}}{\left[\sqrt{(51450)-\frac{(500)^2}{5}}\right]\left[\sqrt{(10030)-\frac{(174)^2}{5}}\right]}=\frac{17940-17400}{\left(\sqrt{51450-50000}\right)\left(\sqrt{10030-6055.2}\right)}$$

$$=\frac{540}{\left(\sqrt{1450}\right)\left(\sqrt{3974.8}\right)}=\frac{540}{(38.08)(63.046)}=\frac{540}{2400.73}=+0.225$$

معامل ارتباط بيرسون بين التحصيل الدراسي وطريقة التدريس ونرمز له $r_{2.3}$

XY	Y^2	X^2	طريقة التدريس (3) Y	التحصيل الدراسي (2) X	الطلبة
195	1169	225	13	15	1
260	400	169	20	13	2
605	3025	121	55	11	3
1040	6400	169	80	13	4
48	36	64	6	8	5
2148	10030	748	174	60	**المجموع**

$$r_{2.3}=\frac{2148-\frac{(60)(174)}{5}}{\left[\sqrt{(748)-\frac{(60)^2}{5}}\right]\left[\sqrt{(10030)-\frac{(174)^2}{5}}\right]}=\frac{2148-2088}{\left(\sqrt{748-720}\right)\left(\sqrt{10030-6055.2}\right)}$$

$$=\frac{60}{\left(\sqrt{28}\right)\left(\sqrt{3974.8}\right)}=\frac{60}{(5.292)(63.046)}=\frac{60}{333.639}=+0.179$$

بعد حساب معامل الارتباط الثنائي المناسب نقوم بعدها بتطبيق قانون معامل الارتباط الجزئي كالتالي:

$$r_{1.2.3}=\frac{(r_{1.2})-[(r_{1.3})(r_{2.3})]}{\sqrt{\left[1-(r_{1.3})^2\right]\left[1-(r_{2.3})^2\right]}}$$

$$r_{1.2.3}=\frac{(-.620)-[(.225)(.179)]}{\sqrt{\left[1-(.225)^2\right]\left[1-(.179)^2\right]}}=\frac{(-.620)-(.0402)}{\sqrt{(1-0.0506)(1-0.032)}}=\frac{-0.662}{\sqrt{(0.9494)(0.968)}}=\frac{-0.662}{\sqrt{0.919}}=\frac{-0.662}{0.9586}=-0.691$$

وهذه النتيجة توضح العلاقة بين الغياب والتحصيل الدراسي مع تثبيت أثر طريقة التدريس على هذه العلاقة

اختبار معنوية معامل الارتباط Significance of Correlation Coefficient

إذا كانت قيمة معامل ارتباط العينة r قريبة من + 1 أو – 1 فإن هناك علاقة خطية قوية بين المتغيرين، وإذا كانت r = 0 فإنه لا توجد علاقة خطية بينهما، أما إذا كانت قيم r متوسطة فإنه يجب اختبار معنوية (أو دلالة) معامل ارتباط العينة، وهل هناك ارتباط حقيقي بين المتغيرين في المجتمع، أم أن الارتباط بينهما زائف وغير حقيقي.

وفيما يلي نتناول بالتفصيل اختبار معنوية معامل ارتباط المجتمع والذي نرمز له بالرمز R.

اختبار أن معامل ارتباط المجتمع يساوي الصفر:

بافتراض أن المجتمع له توزيع طبيعي فإن معامل ارتباط العينة r يكون له توزيع t بوسط حسابي يساوي R وانحراف معياري يساوي $\sqrt{\frac{1-r^2}{n-2}}$ وذلك بدرجات حرية n – 2. وبالتالي تكون خطوات اختبار أن معامل ارتباط المجتمع يساوي صفر كما يلي:

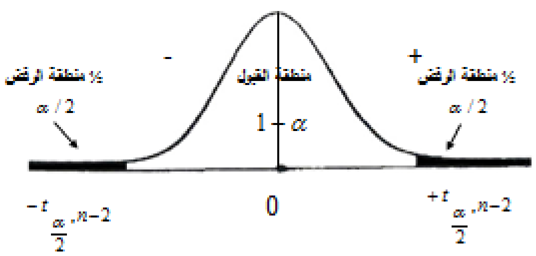
١. الفرض العدمي: أن معامل ارتباط المجتمع يساوي صفر، أي لا يوجد ارتباط بين المتغيرين. وبالرموز:
$H_0 : R = 0$

٢. الفرض البديل: معامل ارتباط المجتمع لا يساوي صفر، أي ويجد ارتباط بين المتغيرين، وبالرموز: $H_A : R \neq 0$

٣. إحصائية الاختبار: ستكون إحصائية الاختبار في هذه الحالة هي t والتي تأخذ الشكل التالي: $T = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$

والتي لها توزيع t بدرجات حرية n – 2.

٤. حدود منطقتي القبول والرفض: والتي نحصل عليها من جدول t لمستوى معنوية يساوي $\frac{\alpha}{2}$ ودرجات حرية تساوي n – 2 (اختبار الطرفين):



٥. المقارنة والقرار: حيث نقارن قيمة إحصائية الاختبار (المحسوبة في الخطوة رقم 3) بحدود منطقتي القبول والرفض (من الخطوة رقم 4). فإذا وقعت قيمة الإحصائية في منطقة القبول فإن القرار هو قبول الفرض العدمي بأن R = 0 أي لا يوجد ارتباط بين المتغيرين والعكس إذا وقعت قيمة الإحصائية في منطقة الرفض فإن القرار هو رفض الفرض العدمي، وفي هذه الحالة نقبل الفرض البديل بأن هناك ارتباط بين المتغيرين وذلك بمستوى معنوية يساوي α

مثال:

اختبر معنوية معامل الارتباط المحسوب في المثال السابق وذلك بمستوى معنوية 5 %

الحل:

لوكان لدينا البيانات التالية :

r = 0.91 ، n = 10

وتكون خطوات اختبار معنوية الارتباط كما يلي :

١. الفرض العدمي: $H_0 : R = 0$
٢. الفرض البديل: $H_A : R \neq 0$

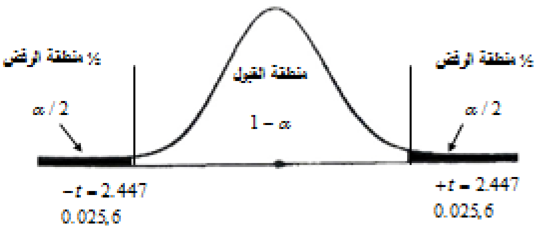
٣. <u>الإحصائية:</u>

$$t=\frac{r}{\sqrt{\frac{1-r^2}{n-2}}}=\frac{0.91}{\sqrt{\frac{1-(0.91)^2}{10-2}}}=\frac{0.91}{\sqrt{\frac{0.1719}{8}}}=\frac{0.91}{\sqrt{0.0215}}=\frac{0.91}{0.1466}=6.2074$$

إذا: $t=6.2074$

٤. <u>حدود منطقتي القبول والرفض:</u>

من جدول t حيث مستوى المعنوية $\alpha=0.05$, $\frac{\alpha}{2}=0.025$ ودرجات الحرية تساوي (n – 2 = 8 – 2=6) نجد أن قيمة t تساوي 2.447 وتكون حدود منطقتي القبول والرفض كما يلي:

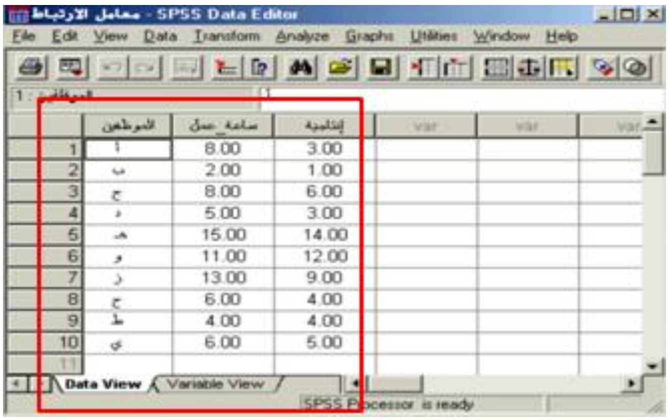


٥. <u>المقارنة والقرار:</u> بمقارنة قيمة الإحصائية المحسوبة في الخطوة رقم 3 والتي تساوي 6.2074 بحدود منطقتي القبول والرفض (أو قيم t الجدولية في الخطوة رقم 4) نجد أنها تقع في منطقة الرفض (حيث أنها أكبر من 2.447) لذلك فإن القرار هو: **رفض الفرض العدمي وقبول الفرض البديل. أي رفض أن معامل الارتباط يساوي صفر.** وقبول أن معامل الارتباط **لا** يساوي صفر أي **يوجد ارتباط** بين المتغيرين (أعمار الناخبين و دخولهم اليومية) وذلك بمستوى معنوية 5 %

حساب معامل ارتباط بيرسون Pearson Correlation Coefficient من خلال برنامج الـ SPSS

لغرض حساب قيمة معامل ارتباط بيرسون لنفس المثال السابق من خلال استخدام برنامج الـ SPSS نتبع الخطوات التالية:

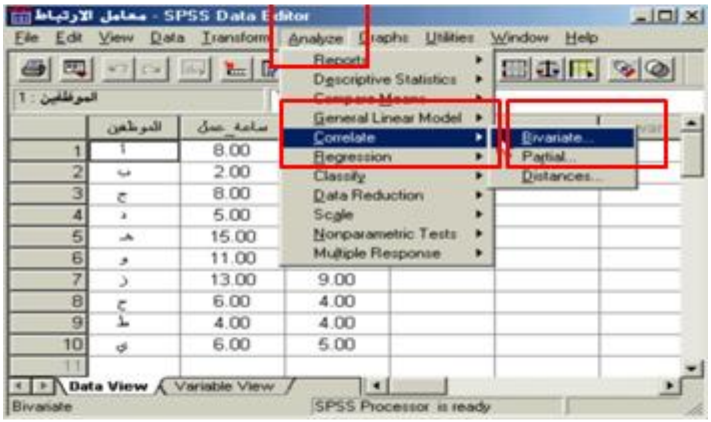
- قم بإدخال البيانات المراد تحليلها من خلال شاشة تحرير البيانات Data Editor بالطريقة المناسبة كالتالي:



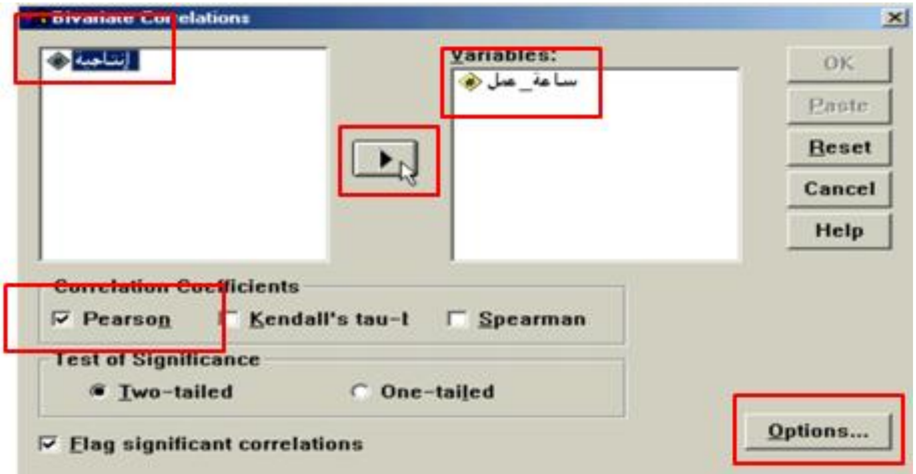
	الموظفين	ساعة_عمل	إنتاجية
1	أ	8.00	3.00
2	ب	2.00	1.00
3	ج	8.00	6.00
4	د	5.00	3.00
5	هـ	15.00	14.00
6	و	11.00	12.00
7	ز	13.00	9.00
8	ح	6.00	4.00
9	ط	4.00	4.00
10	ي	6.00	5.00

لاحظ أنه تم إدخال بيانات كل متغير في عمود منفصل عن الآخر، وقد تم إعطاء كل متغير الاسم المناسب له .

- من القائمة "تحليل" Analyze اختر الأمر "الارتباط" Correlate فتظهر قائمة أوامر فرعية اختر منها "بسيط أو فردي" Bivariate كالتالي:



- بعد اختيار الأمر "بسيط أو فردي" Bivariate سوف يظهر لك صندوق الحوار التالي:



- من قائمة المتغيرات في الجهة اليسرى من صندوق الحوار حدد المتغيرين المراد حساب قيمة الارتباط لهما ، ونقلها إلى المستطيل الخاص بـ "المتغيرات " Variables ، ثم بعد ذلك أنقر على السهم الذي يظهر مقابل المستطيل الخاص بـ "متغيرات الدراسة" ، ستلاحظ انتقال هذه المتغيرات مباشرة إلى المستطيل "المتغيرات" Variables . بعد ذلك اختر نوع معامل الارتباط المراد حسابه (حيث يظهر في اسفل صندوق الحوار هذا ثلاث أنواع من معامل الارتباط وهي معامل ارتباط بيرسون Pearson ومعامل ارتباط كاندل Kendall ومعامل ارتباط سبيرمان Spearman) وسوف نختار هنا معامل ارتباط بيرسون Pearson . كذلك يستطيع المستخدم من خلال صندوق الحوار هذا تحديد ما إذا كان المختبر الإحصائي بذيل واحد One-tailed أو بذيلين Tow-tailed
- أنقر على زر "خيارات" Options في الجهة السفلية اليمنى من صندوق الحوار السابق وذلك عند الرغبة في حساب الخصائص الأساسية للمتغيرات موضع الدراسة Statistics ، وكذلك كيفية التعامل مع القيم المفقودة Missing Values ، وبعد الانتهاء من التعديل على هذا الصندوق الحواري أنقر على زر "استمرار" Continue

- أنقر بعد ذلك على زر "موافق" OK سيؤدي ذلك إلى تنفيذ الاختبار، وستلاحظ ظهور النتائج في شاشة المخرجات كالتالي:

➡ **Correlations**

Descriptive Statistics

	Mean	Std. Deviation	N
ساعة_عمل	7.8000	4.10420	10
إنتاجية	6.1000	4.22821	10

Correlations

		ساعة_عمل	إنتاجية
ساعة_عمل	Pearson Correlation	1	.910**
	Sig. (2-tailed)	.	.000
	N	10	10
إنتاجية	Pearson Correlation	.910**	1
	Sig. (2-tailed)	.000	.
	N	10	10

**. Correlation is significant at the 0.01 level

معامل الارتباط

مستوى الدلالة

حجم العينة

ويسمى الجدول السابق مصفوفة الارتباط حيث تتكون كل خلية من 3 أرقام، الأول هو قيمة معامل الارتباط بين المتغيرين والثاني هي قيمة مستوى دلالة الاختبار والثالث هو عدد البيانات (حجم العينة) المدخلة والمستخدمة في عملية الحساب

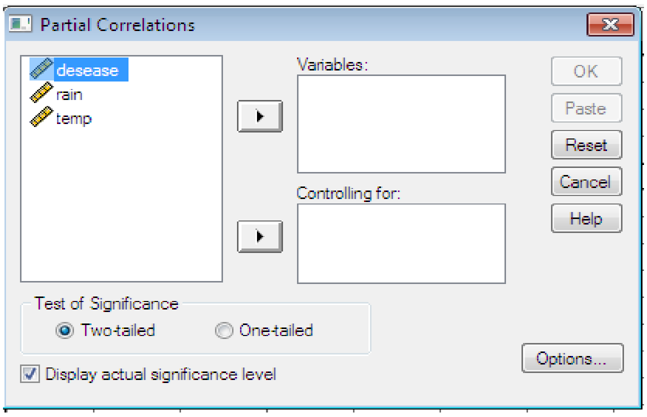
ونلاحظ أن برنامج الـ SPSS قام مباشرة بحساب الإحصاءات الأساسية للبيانات مثل المتوسط الحسابي والانحراف المعياري للمتغيرات ثم قام البرنامج بحساب معامل ارتباط بيرسون Pearson Correlation بين المتغيرين ساعات العمل ومستوى الإنتاجية والذي يساوي 0.91 وكذلك مستوى الدلالة لهذا الارتباط والذي يبدوا أنه دال عند مستوى دلالة 0.001 مما يدل على وجود علاقة إيجابية وقوية بين المتغيرين ساعات العمل في الشركة ومستوى إنتاجية الموظفين، فإننا بالتالي نرفض الفرضية الصفرية، أي توجد علاقة بين المتغيرين .

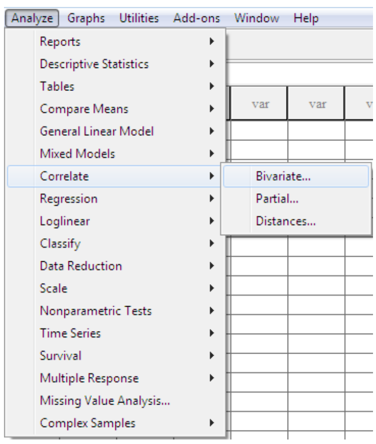
والأوامر المستخدمة لحساب معامل ارتباط بيرسون (r) Pearson Correlation Coefficient من خلال برنامج الـ SPSS (مع اختلاف المتغيرات موضع الدراسة) هي:

```
CORRELATIONS
/VARIABLES=ساعة_عمل إنتاجية
/PRINT=TWOTAIL NOSIG
/STATISTICS DESCRIPTIVES
/MISSING=PAIRWISE .
```

حساب معامل الإرتباط الجزئي Partial Correlation Coefficient من خلال برنامج الـ SPSS

- إدخال البيانات بصورة عادية في شاشة إدخال البيانات.
- بعدها نختار من correlate ___ analyze ومنها نختار partial لتظهر لنا النافذة التالية -:





ندخل متغير الغياب و التحصيل الدراسي الى المربع المعنون (VARIABLE) أما المتغير طريقة التدريس ندخله في المربع المعنون (controlling for) ، ثم نختار مستوى المعنوية من طرف واحد one- taild ثم نضغط ok.

Correlations

Control Variables			desease	temp
rain	desease	Correlation	1.000	-.493
		Significance (2-tailed)	.	.123
		df	0	9
	temp	Correlation	-.493	1.000
		Significance (2-tailed)	.123	.
		df	9	0

معامل الارتباط
مستوى الدلالة
حجم العينة

نلاحظ في النتائج أنه تم إستخدام المتغير طريقة التدريس مسيطر عليه أو بمعني أخر متغير غير نشط في الإختبار، أما المتغيرين الغياب والتحصيل الدراسي هما متغيري الإختبار وقد وجد الارتباط بينهما إرتباط عكسي وقد بلغ 0.493- بدرجات حرية n-3 ومستوي معنوية (sig = 0.123) ومن ثم هذه النتيجة تعني قبول فرض العدم.

ملحـــوظة:

١. في حالة وجود قيمة الارتباط موجبة يقوم البرنامج بأختبار الفرضية التالية:

$H_0 : P = 0 \quad VS \quad H_A : P > 0$

٢. في حالة وجود قيمة الارتباط سالبة يقوم البرنامج بأختبار الفرضية التالية:

$H_0 : P = 0 \quad VS \quad H_A : P < 0$

يمكن إستخراج معامل الارتباط البسيط لـ PEARSON بالنقر علي زر option في صندوق partial correlation وإختيار zero order correlation

Correlations

Control Variables			desease	temp	rain
-none- [a]	desease	Correlation	1.000	-.144	-.668
		Significance (2-tailed)	.	.654	.018
		df	0	10	10
	temp	Correlation	-.144	1.000	-.307
		Significance (2-tailed)	.654	.	.332
		df	10	0	10
	rain	Correlation	-.668	-.307	1.000
		Significance (2-tailed)	.018	.332	.
		df	10	10	0
rain	desease	Correlation	1.000	-.493	
		Significance (2-tailed)	.	.123	
		df	0	9	
	temp	Correlation	-.493	1.000	
		Significance (2-tailed)	.123	.	
		df	9	0	

a. Cells contain zero-order (Pearson) correlations.

المحاضرة 13

اختبار الفروض الإحصائية اللامعلمية (الجزء الأول)

الطرق الإحصائية اللامعلمية Nonparametric Methods

تتطلب معظم التحليلات تحديد بعض الافتراضات أو الشروط حول المجتمع أو المجتمعات التي اختيرت منها العينة أو العينات. **ففي كثير من الحالات يتم افتراض أن المجتمعات موضع الدراسة تتصف بالتالي:**

- افتراض أن المجتمعات موضع الدراسة تتبع توزيعاً طبيعياً
- افتراض أن تباينات هذه المجتمعات معلومة
- افتراض أن تباينات هذه المجتمعات غير معلومة ولكنها متساوية
- افتراض أن العينات المختارة مستقلة

وحيث أنه توجد مواقف أو حالات كثيرة يكون من الصعب التأكد من تحقق هذه الافتراضات، أو يكون هناك شك في تحققها، وحيث أننا نواجه في كثير من الأحيان بيانات واقعية يصعب فيها التعرف على صيغة التوزيع الاحتمالي الذي تتبعه. **لذلك فقد طور الإحصائيين أساليباً وطرقاً إحصائية بديلة وهذه الطرق تتصف بالتالي:**

- لا تتطلب افتراضات كثيرة
- لا تتطلب معرفة صيغة التوزيع الاحتمالي للمجتمعات التي تختار منها العينات

ومن هنا نشأت الطرق اللامعلمية

وهذه الطرق بالإضافة إلى أنه يمكن استخدامها تحت شروط وافتراضات عامة فإنها غالباً لا تحتاج إلى مجهود في العمليات الحسابية .كما انه يمكن التعامل معها لمتغيرات منفصلة ومتغيرات متصلة على السواء. ولهذه الأسباب أصبحت الطرق اللامعلمية مرغوبة بكثرة

الفرق بين الطرق الإحصائية المعلمية والطرق الإحصائية غير المعلمية:

سنعرض فيما يلي مقارنة بين الطرق الإحصائية المعلمية والطرق الإحصائية اللامعلمية ، وذلك من خلال الجدول التالي:

	الطرق الإحصائية المعلمية	الطرق الإحصائية اللامعلمية
1	يتطلب استخدام الطرق المعلمية الوفاء بافتراضات معينة حول التوزيع الأساسي للمجتمع، من حيث أن يكون توزيع المجتمع طبيعياً	لا يتطلب استخدام الطرق اللامعلمية ايه افتراضات أو معلومات حول خصائص التوزيع الأساسي للمجتمع (لذلك يطلق بعض الإحصائيين عليها إحصاءات التوزيعات الحرة)
2	لا يمكن استخدام الطرق الإحصائية المعلمية لمعالجة وتحليل البيانات في المواقف التجريبية التي يكون فيها حجم العينة صغير جداً، لأن الحجم الصغير للعينة يؤثر على خصائص التوزيع التكراري للعينة الصغيرة، فتبتعد بذلك عن اعتدالية التوزيع التكراري للمجتمع الأب	يمكن استخدام بعض الطرق اللامعلمية لمعالجة وتحليل البيانات في المواقف التجريبية التي يكون فيها حجم العينة صغير جداً
3	تكون الطرق المعلمية أكثر ملائمة لتحليل البيانات الفترية والبيانات النسبية فضلاً عن أنها يمكن استخدامها (مع بعض التحفظ) لتحليل يانات الرتيبة فقط وذلك بعد أن يتم تحويلها إلى بيانات فترية من خلال إعطاء كل رتبة درجات تتناسب مع قيمتها، مع مراعاة أن تكون البيانات في صورة رتب ذات درجات متتابعة مثل: موافق بشدة، موافق، غير موافق، غير موافق بشدة	تكون الطرق اللامعلمية عادة أكثر ملاءمة للاستخدام عندما تكون البيانات الخاصة بالبحث من النوعين الاسمي والرتبي فضلاً عن أنها يمكن استخدامها أحياناً عندما تكون البيانات فترية أو نسبية، وذلك بعد أن يتم تحويلها الى بيانات اسمية او رتبية (وهذا فيه إهدار لجزء كبير من البيانات التي نحللها، لان الطرق الإحصائية اللامعلمية لا تستخدم الأجزاء يسيراً من تلك البيانات) لذلك أوضح بعض الإحصائيين انه عندما نستخدم الطرق الإحصائية اللامعلمية للبيانات الفترية والنسبية فإنها لا تستخدم إلا للتقدير المبدئي، على أن يتلوه بعد ذلك استخدام الطرق المعلمية

4	تعتبر الطرق الإحصائية المعلمية بشكل عام أقوى من الطرق الإحصائية اللامعلمية، حيث أن الطرق المعلمية تميل إلى رفض الفرضية الصفرية اكثر من ميل الطرق اللامعلمية لرفض نفس الفرضية	بالرغم من تحرر الطرق الإحصائية اللامعلمية من الشروط والخصائص التي قد تعوق أحيانا استخدام الطرق الإحصائية المعلمية، إلا أن الطرق الإحصائية اللامعلمية بشكل عام أقل قوة من الطرق الإحصائية المعلمية
5	تعتمد الطرق الإحصائية المعلمية بشكل عام على الدرجات الأصلية والتي يتم تحليلها كما هي	تعتمد الطرق اللامعلمية في اغلب الأحيان على البيانات التي هي بشكل تكرارات أو رتب مما يؤدي إلى ضياع بعض المعلومات المفيدة
6	تعتبر الطرق الإحصائية المعلمية بصورة عامة أصعب في الاستخدام من الطرق الإحصائية اللامعلمية، وبالتالي فإن الوقت الذي يحتاجه الباحث لتحليل بياناته يكون أطول، مما يؤدي إلى تأخر الحصول على النتائج (نسبياً) والإفادة منها تطبيقياً	تعتبر الطرق الإحصائية اللامعلمية أفضل وسيلة للتقدير المبدئي السريع، فهي بصورة عامة اسهل استخداماً من الطرق المعلمية وبالتالي فإن الوقت الذي يحتاجه الباحث لتحليل بياناته يكون اقل، مما يؤدي إلى الإسراع في الحصول على النتائج والإفادة منها تطبيقياً
7	تعتبر الطرق الإحصائية المعلمية أكثر شيوعاً واستخداماً لدى المختصين في الإحصاء	تعتبر الطرق الإحصائية اللامعلمية أكثر شيوعاً واستخداماً لدي غير المتخصصين في الإحصاء
8	ينبغي أن يكون اختيار العينة من المجتمع بصورة عشوائية، وأن تكون إحصاءات العينة (مقاييس النزعة المركزية والتشتت) صورة مقربة للمعلمات الإحصائية للمجتمع	ليس هناك ضرورة أن يكون اختيار العينة من المجتمع بصورة عشوائية
9	تستخدم الطرق الإحصائية المعلمية لمعالجة وتحليل البيانات الكمية	تستخدم الطرق الإحصائية اللامعلمية لمعالجة وتحليل البيانات النوعية، و التي لا يمكن عادة استخدام أي طريقة إحصائية معلمية لتحليلها
10	من الأمثلة على الطرق الإحصائية المعلمية اختبار (Z) واختبار (t) واختبار (F) ... الخ.	من الأمثلة على الطرق الإحصائية اللامعلمية اختبار (كا^2) واختبار مان ويتني واختبار كروسكال واليز ... الخ.

طالما أن الاختبارات اللابارامترية (اللامعلمية) تتطلب هذا العدد القليل من الافتراضات حول البيانات، فلماذا لا نستخدمها في آل الحالات؟

إن الميزة السيئة للاختبارات اللابارامترية هي أنها غير جيدة عادة لإيجاد الفروقات عندما يكون هناك فروقات في المجتمع، وعندما تكون الافتراضات من أجل الاختبارات البارامترية محققة، بمعنى آخر الاختبارات اللابارامترية غير قوية كاختبارات تفترض توزيعا طبيعيا، الاختبارات البارامترية، ذلك بسبب أن الاختبارات اللابارامترية تتجاهل بعض المعلومات المتوفرة، فعلى سبيل المثال في اختبار ويلكوكسن نستبدل قيم البيانات برتبها.

بشكل عام، إذا كانت افتراضات اختبار بارامتري مقنعة فيجب أن نستخدم اختبارات بارامترية للتحليل لأنها أكثر قوة، وقد رأينا أن العديد من هذه الاختبارات يمكن أن تقوم بانتهاك الافتراض إلى حد معقول. أي أنها قوية robust ، الإجرائيات اللابارامترية أكثر نفعًا من أجل العينات الصغيرة، عندما يكون هناك ابتعاد ملموس عن الافتراضات المطلوبة، وهي أيضًا مفيدة عندما يكون هناك قيم حدودية، حيث أن الحالات المتطرفة لن تؤثر على النتائج بقدر التأثير الناتج في حال استخدمنا اختبارات معتمدة على إحصائية بسيطة كالمتوسط مثلا.

ماذا علينا أن نفعل عندما لا نكون متأكدين من اختيار الاختبار المناسب: بارامتري-لابارامتري؟

عندما تكون في حيرة وشك استخدم كليهما. إذا حصلت على نتيجة واحدة فلا شيء يدعو للقلق، أما إذا كانت النتائج التي حصلت عليها من الاختبار اللابارامتري تفتقد إلى مدلول بينما تلك التي حصلت عليها باستخدام الاختبار البارامتري أكثر دلالة حاول أن تظهر السبب، هل يوجد بين البيانات قيم متطرفة (أصغر أو أكبر من بقية القيم بشكل ملحوظ)، إذا كانت تلك الحالة موجودة فهذا يعني أن تلك القيم يمكن أن تؤثر على المتوسط وأن لها تأثير أكبر على نتائجك.

افحص تلك القيم بحرص لكي تتأكد من صحتها. إذا كانت المشكلة نابعة من كون قيم البيانات خارجة عن التوزيع الطبيعي انظر في إمكانية تحويل البيانات إلى شكل أكثر توافقًا مع افتراضات بارامترية. إذا حقق هذا التحويل نجاحًا تستطيع أن تستخدم واحدًا من الإجرائيات البارامترية القوية من أجل نتائجك

بعض الأمثلة للاختبارات الاحصائية اللامعلمية Some Examples of Nonparametric Methods

اختبار مان وتني Mann – Whitney U :

استخدامه:

يعتبر هذا الاختبار بديل لا معلمي للاختبار الخاص بالفرق بين متوسطي مجتمعين والمبني على أساس عينتين مستقلتين أي أن هذا الاختبار بديل لاختبار t لعينتين مستقلتين، بل أنه أفضل منه خاصة إذا كانت العينتان مختارتين من مجتمعين لا يتبعان توزيعاً طبيعياً.

ويعد هذا الاختبار أكثر الاختبارات اللابارامترية استخداماً فى البحوث عندما يكون المتغير التابع من المستوى الرتبى بدلاً من الدرجات الأصلية، كما يمكن استخدام هذا الاختبار إذا كانت المتغيرات من المستوى الفترى أو النسبى ولكنها لا تفى بشروط اختبار النسبة التائية مثل عدم إعتدالية التوزيع أو اختلاف التباين بين المجموعتين اختلافاً كبيراً.

نفرض أن لدينا عينتين مستقلتين الأولى حجمها n_1 والثانية حجمها n_2 تم اختيارهما من مجتمعين متصلين ومتماثلين الأول متوسطه μ_1 والثاني متوسطه μ_2

والمطلوب:

اختبار الفروض التالية:

الفرض العدمي: $H_0 : \mu_1 = \mu_2$

في حالة تساوي متوسطي المجتمعين يكون شكل الفرضية البديلة:

الفرض البديل: $H_A : \mu_1 \neq \mu_2$

في حالة عدم تساوي متوسطي المجتمعين (وجود اختلاف معنوي بين متوسطي المجتمعين) يكون شكل الفرضية البديلة:

الفرض البديل :

Or $H_A : \mu_1 < \mu_2$

Or $H_A : \mu_1 > \mu_2$

خطوات الاختبار:

١. دمج مشاهدات العينتين كأنهما عينة واحدة
٢. ترتيب المشاهدات تصاعدياً أو تنازليا وكأنهم مجموعة واحدة
٣. حساب إحصاء الاختبار من خلال تطبيق المعادلات التالية:

$$U_1 = W_1 - [(n_1)(n_2)]+n_1(n_1+1)/2$$
$$U_2 = W_2 - [(n_1)(n_2)]+n_2(n_2+1)/2$$

حيث:

W_1 مجموع رتب العينة الأولى
W_2 مجموع رتب العينة الثانية
n_1 عدد أفراد العينة الأولى
n_2 عدد أفراد العينة الثانية
α مستوى المعنوية

فإن القرار يكون: $U > W_{(1-\alpha)}$

نرفض الفرض العدمي $H_0 : \mu_1 = \mu_2$ والذي يفترض تساوي المجتمعين

ونقبل الفرض البديل $H_A : \mu1 > \mu2$ والذي يفترض أن متوسط المجتمع الأول أكبر من متوسط المجتمع الثاني

حيث أن: $W_{(1-\alpha)} = n_1n_2 - W_\alpha$

أو يكون: $U < W_{(\alpha)}$

نرفض الفرض العدمي $H_0 : \mu_1 = \mu_2$ والذي يفترض تساوي المجتمعين

ونقبل الفرض البديل $H_A : \mu1 < \mu2$ والذي يفترض أن متوسط المجتمع الأول أصغر من متوسط المجتمع الثاني

حيث أن: $W_{(1-\alpha)} = n_1n_2 - W_\alpha$

أما إذا كانت:

$U > W_{(1-\alpha/2)}$ Or

$U < W_{\alpha/2}$

نرفض الفرض العدمي $H_0 : \mu_1 = \mu_2$ والذي يفترض تساوي المجتمعين

ونقبل الفرض البديل $H_A : \mu1 \neq \mu2$ والذي يفترض عدم تساوي متوسطي المجتمعين

حيث أن: $W_{(1-\alpha)} = n_1n_2 - W_{\alpha/2}$

حيث أن:

$W_{\alpha/2}$ تحسب عن طريق جداول خاصة باختبار مان وتني

وإذا كانت كل من n_1 و n_2 أكبر من 8 فإن U تتبع تقريباً توزيع طبيعي

متوسطه: $\mu_U = n_1n_2/2$

وتباينه: $\sigma_U^2 = n_1n_2(n_1+n_2+1)/12$

ويكون إحصاء الاختبار: $Z = (U- \mu_U)/ \sigma_U \sim N(0,1)$

حساب اختبار مان وتني Mann – Whitney U من خلال برنامج SPSS

مثال:

فيما يلى بيان بدرجات مجموعة من الطلاب فى مادة المحاسبة، فى كل من جامعة الملك فيصل وجامعة الدمام:

١. درجات مادة المحاسبة بكلية إدارة الأعمال جامعة الملك فيصل:

10	14	7	8	16
3	7	15	14	7

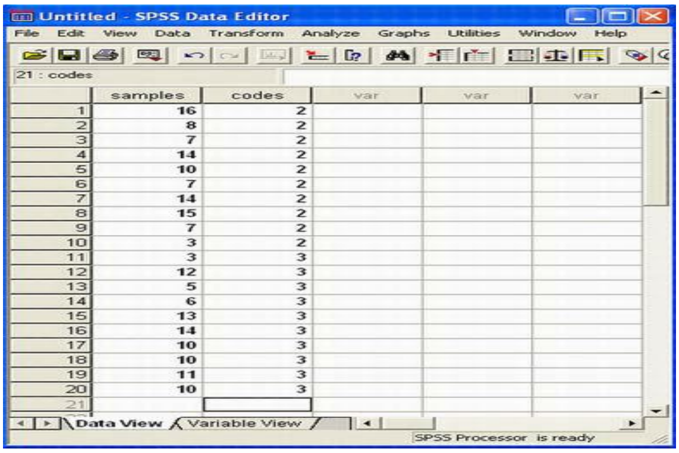
٢. درجات مادة المحاسبة بكلية إدارة الأعمال جامعة الدمام:

13	6	5	12	3
10	11	10	10	14

المطلوب:

بإستخدام إختبار مان – ويتنى: إختبر هل هناك إختلاف فى متوسط درجات مادة المحاسبة بين جامعة الملك فيصل وجامعة الدمام وذلك عند مستوى معنوية 5% .

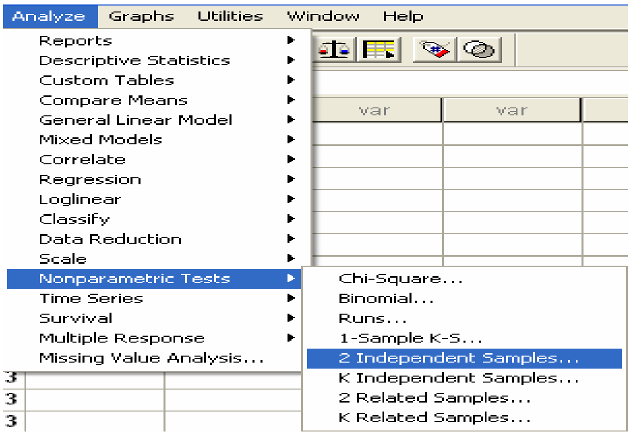
أولا: ندخل البيانات كالتالي:



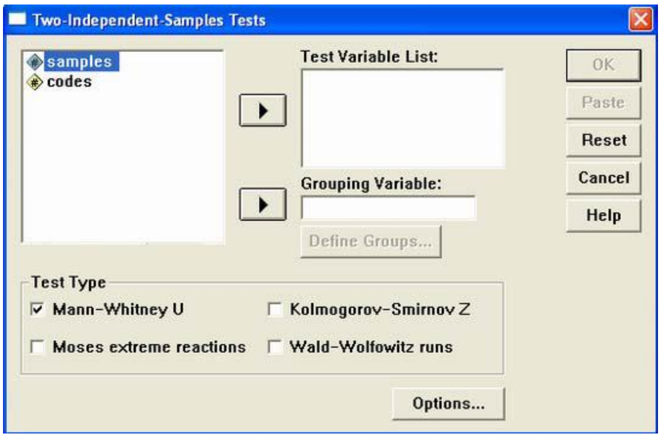
ملاحظة: فى هذا التدريب نحن بصدد إدخال بيانات لعينات مستقلة، لذا تم إدخال جميع المشاهدات فى عمود، ووالترميز الخاصة بالعينات فى عمود آخر وذلك من خلال إعطاء الرقم (2) لبيانات العينة الأولى و (3) لبيانات العينة الثانية.

ثانيا: خطوات تنفيذ الاختبار:

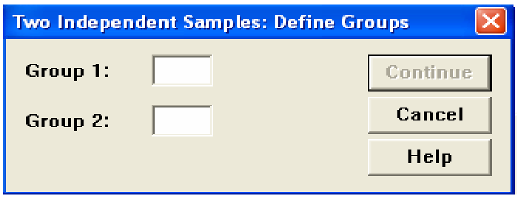
نضغط على قائمة Analyze ومن القائمة الفرعية لـ Nonparametric tests نختار 2 Independent Samples كما هو موضح بالشكل التالى:



سوف يظهر لنا المربع الحوارى التالى:



انقل المتغير Samples الى المربع الذى بعنوان Test Variable List ، ثم انقل متغير الترميز codes إلى المربع الذي بعنوان Grouping Variable، ثم بعد ذلك اضغط على Define Groups سوف يظهر لنا مربع حوارى جديد كما يلى:



- فى خانة [Group 1] اكتب الرمز الخاص بالعينة الاولى (2)، وفى خانة [Group 2] اكتب الرمز الخاص بالعينة الثانية (3)
- ثم اضغط Continue للعودة الى المربع الحوارى السابق
- ثم اضغط Ok سوف تظهر لك نافذة المخرجات الخاصة بهذا الإختبار

Ranks

	CODES	N	Mean Rank	Sum of Ranks
SAMPLES	2	10	11.10	111.00
	3	10	9.90	99.00
	Total	20		

Test Statistics[b]

	SAMPLES
Mann-Whitney U	44.000
Wilcoxon W	99.000
Z	-.457
Asymp. Sig. (2-tailed)	.648
Exact Sig. [2*(1-tailed Sig.)]	.684[a]

يلاحظ من نتائج هذا الإختبار: أن قيمة P.Value تساوى 0.648 و هي أكبر من مستوى المعنوية 5% وبالتالى فاننا نقبل الفرض العدمي بأن متوسط درجات مادة المحاسبة فى كلية إدارة الأعمال جامعة الملك فيصل يساوى متوسط درجات مادة المحاسبة فى جامعة الدمام، أى أن الفروق بين الجامعتين غير معنوية.

إختبار ويلكوكسون Wilcoxon Test :

استخدامه:

ويسمى باختبار اشارات الرتب Sign –rank، ويستخدم هذا الاختبار فى تحديد ما إذا كان هناك اختلاف أو فروق بين عينتين مرتبطتين فيما يتعلق بمتغير تابع معين، ويعد بديلاً لابارامترياً لاختبار T لعينيتين مرتبطين، وتشتمل العينتان على نفس المجموعة من الأفراد يجرى عليهم قياس قبلى Pre test، وقياس بعدى Post test وفى مثل هذه الحالة يكون لكل فرد من أفراد العينة درجتان أحداهما تمثل درجته فى الاختبار القبلى والثانية تمثل درجته فى الاختبار البعدى. ويستخدم مع البيانات العددية فقط دون الاسمية

حتى نحسب اختبار ويلكوكسن يجب اولا أن نجد الفرق بين القيمتين من أجل كل زوج ومن ثم من أجل كافة الحالات التي يكون عندها الفرق غير معدوم، نرتب الفروقات بشكل تصاعدي متجاهلين إشارة الفروقات، ذلك يعني بأن نسند إلى الفرق الصغير في القيمة المطلقة الرتبة 1 ونسند إلى الفرق الصغير التالي الرتبة 2 وهكذا، أما في حالة الفروقات المتساوية (الحالات المتعادلة) نسند رتبة المتوسط إلى تلك الحالات.

نفرض أن لدينا عينتين مترابطتين (غير مستقلتين)

والمطلوب:

اختبار الفروض التالية:

الفرض العدمي: $H_0 : \mu_1 = \mu_2$

في حالة تساوي متوسطي العينتين يكون شكل الفرضية البديلة:

الفرض البديل : $H_A : \mu_1 \neq \mu_2$

في حالة عدم تساوي متوسطي العينتين (وجود اختلاف معنوي بين متوسطي العينتين) يكون شكل الفرضية البديلة:

الفرض البديل :

Or $H_A : \mu_1 < \mu_2$
Or $H_A : \mu_1 > \mu_2$

خطوات الاختبار:

١. إيجاد الرتب مع ملاحظة فى عمود رتب $|D_1|$ يعطى أقل فرق؛ الرتبة رقم واحد اى يتم ترتيب الفروق تصاعديا
٢. احسب الفروق بين المجموعتين
٣. افرز الفروق للرتب السالبة في عمود ورتب للفروق الموجبة في عمود آخر.
٤. تحديد القيمة الأصغر في W_1، W_2 الرتب

نبحث عن القيمة الحرجة في جدول ويلكوكسون عند n (فلو كانت قيمة n = 5)، ومستوى الدلالة 0.05 نجد أن القيمة الحرجة=1

ولان الفرض ذو ذيلين إذن نضرب قيمة W من الجدول × 2 = 2

ولو أفترضنا أن القيمة المحسوبة لاختبار ويلكوكسن = (3)

إذن ليست هناك دلالة إحصائية، ونقبل الفرض الصفري

وهى فى ذلك تختلف عن اختبار T البارامترى الذى يقبل الفرض الصفرى عندما تكون قيمة T المحسوبة أكبر من T الجدولية.

حساب إختبار ويلكوكسون Wilcoxon Test من خلال برنامج SPSS

مثال:

تأثير ممارسة الرياضة على إنقاص الوزن:

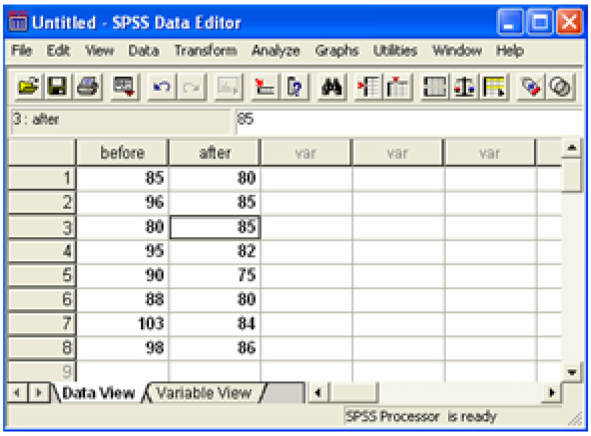
الوزن قبل ممارسة الرياضة	الوزن بعد ممارسة الرياضة
85	80
96	85
80	85
95	82
90	75
88	80
103	84
98	86

المطلوب:

إختبار هل هناك إختلاف معنوى فى الوزن بسبب ممارسة الرياضة، بإستخدام إختبار ويلكوكسون Wilcoxon عند مستوى معنوية 5% .

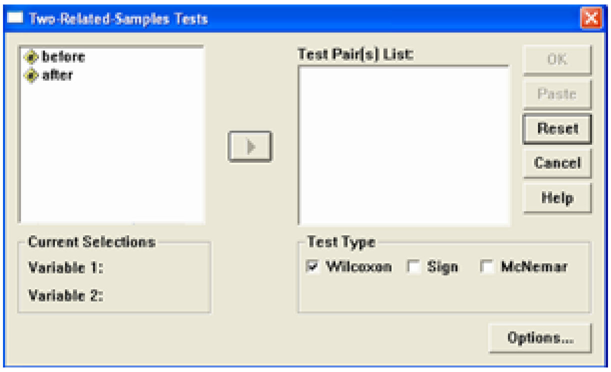
أولا: ندخل البيانات كالتالي:

حيث أننا بصدد عينات غير مستقلة، فإنه سيتم إدخال بيانات كل عينة فى عمود مستقل، كما يلى:



ثانيا: خطوات تنفيذ الاختبار:

نضغط على قائمة Analyze ومن القائمة الفرعية لـ Nonparametric tests نختار 2 Related Samples كما هو موضح بالشكل التالى:



اضغط بالماوس مرة واحدة على المتغير before ثم على المتغير after (لاحظ أنه قد تم تظليل المتغيرين معًا)، ثم قم بنقل هذين المتغيرين الى المربع الذي بعنوان Test Pair(s) List وذلك من خلال الضغط على السهم الصغير الموجود بين المربعين.

لاحظ فى نفس المربع الحوارى الذى أمامك: أن الإختيار الإفتراضى من جانب البرنامج هو اختبار ويلكوكسن، وهو الإختبار الذى نريده لذا سنتركة كما هو. إضغط Ok ستظهر لك نافذة المخرجات الخاصة بهذا الإختبار كالتالي:

Ranks

		N	Mean Rank	Sum of Ranks
AFTER - BEFORE	Negative Ranks	7[a]	4.93	34.50
	Positive Ranks	1[b]	1.50	1.50
	Ties	0[c]		
	Total	8		

Test Statistics[b]

	AFTER - BEFORE
Z	-2.313[a]
Asymp. Sig. (2-tailed)	.021

قام البرنامج بحساب الفروق في الوزن على أساس التالي:

الفرق = الوزن بعد ممارسة الرياضة – الوزن قبل ممارسة الرياضة

ويلاحظ أيضا: أن متوسط الرتب السالبة (4.93) أكبر من متوسط الرتب الموجبة (1.5)، وهذا معناه أن متوسط الوزن قبل ممارسة الرياضة أكبر من متوسط الوزن بعد ممارسة الرياضة (إذا فى غاية الأهمية أن نعرف الترتيب الذى إستخدمه البرنامج للعينتين)

ويلاحظ من نتائج هذا الاختبار أن قيمة P.Value تساوي 0.021 وهي أقل من مستوى المعنوية 5% وبالتالى فإننا نقبل الفرض البديل بأن متوسط الوزن قبل ممارسة الرياضة يختلف معنويًا عن متوسط الوزن بعد ممارسة الرياضة.

اختبار كروسكال واليس Kruskal-Wallis Test :

استخدامه:

يعتبر هذا الاختبار بديلاً لامعلميا لاختبار تحليل التباين في اتجاه واحد، وهو مبني على مجموع الرتب ويستعمل لاختبار الفروق بين ثلاث مجموعات أو أكثر في مثل الحالة الآتية :

نفرض أن لدينا k عينة عشوائية مستقلة الأولى حجمها n_1 والثانية حجمها n_2 وهكذا. أي أن العينة الأخيرة حجمها n_k وأن هذه العينات تم اختيارها من مجتمعات متصلة عددها k ومتوسطاتها هي $\mu_1, \mu_2, \ldots, \mu_k$ على التوالي.

والمطلوب اختبار فرض العدم:

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$

أي جميع متوسطات المجتمعات متساوية

الفرض البديل:

H_1 : ليست جميع متوسطات المجتمعات متساوية

وفي هذا الاختبار نقوم بدمج مشاهدات العينات في عينة واحدة وإعطاء رتب لهذه المشاهدات تصاعدياً، فإذا كانت R_i هي مجموع الرتب للمشاهدات التي تنتمي للعينة رقم i والتي عدد مفرداتها n_i فإن إحصاء الاختبار هو:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)$$

حيث: $n = \sum_{i=1}^{k} n_i$

وإذا كان فرض العدم صحيحاً وكانت $n_i \geq 5$ لجميع قيم i فإن الإحصاء H يتبع تقريباً توزيع مربع كاي بدرجات حرية k-1 أي أن: $H \approx \chi^2_{(k-1)}$

وعلى ذلك فإننا نرفض فرض العدم وهو أن جميع المتوسطات متساوية ونقبل الفرض البديل وهو أن المتوسطات ليست جميعها متساوية إذا كانت قيمة الإحصاء H أكبر من: $\chi^2_{(k-1,\alpha)}$

١. البيانات رتبية أو يمكن ترتيبها.
٢. لا يشترط أن تكون المجموعات متساوية العدد، فيمكن استخدامه مهما كان عدد أفراد العينة.
٣. إذا كان عدد المجموعات الفرعية للمتغير المستقل يزيد عن 3، وعدد الأفراد داخل أى مجموعة منها أكبر من 5، أى حجم العينة كلها أكبر من 15 نقارن القيم المحسوبة بقيم توزيع مربع كا.
٤. إذا كان عدد المجموعات الفرعية للمتغير المستقل = 3 ، وعدد الأفراد داخل أى مجموعة منها لا يتجاوز 5، أى حجم العينة كلها أقل أو = 15 نقارن القيم المحسوبة بالقيم الحرجة لاختبار كروسكال واليس.

حساب إختبار كروسكال واليس Kruskal-Wallis Test من خلال برنامج SPSS

مثال:

الجدول التالى يوضح درجات مجموعة من الطلاب فى مادة الإقتصاد فى ثلاث جامعات هى: جامعة الملك فيصل – جامعة الدمام – جامعة الملك سعود:

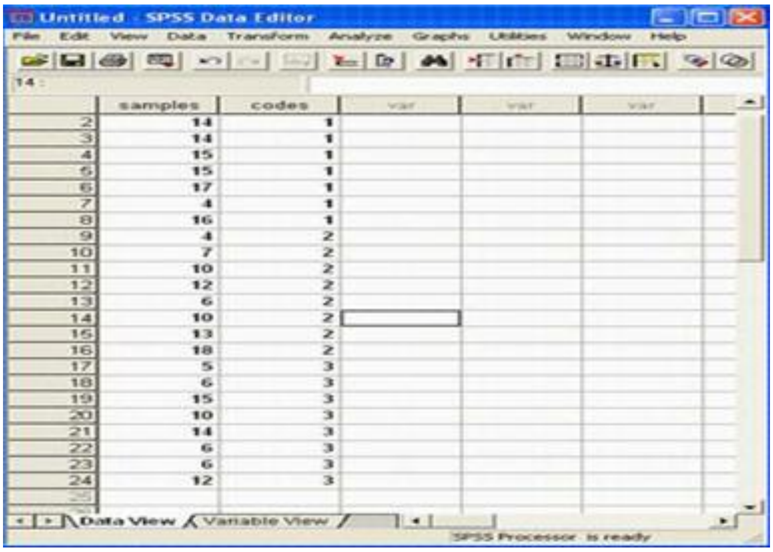
جامعة الملك فيصل	جامعة الدمام	جامعة الملك سعود
13	4	5
14	7	6
14	10	15
15	12	10
15	6	14
17	10	6
4	13	6
16	18	12

المطلوب:

دراسة مدى وجود إختلاف بين مستوى الطلاب فى الجامعات الثلاثة السابقة بإستخدام إختبار كروسكال- والس، وذلك عند مستوى معنوية 5%

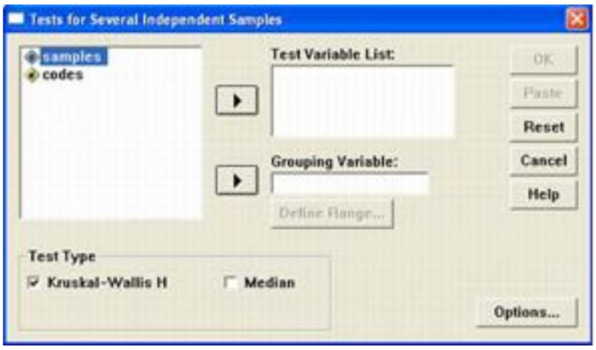
أولا: ندخل البيانات كالتالي:

حيث أننا بصدد ثلاث عينات مستقلة، لذا تم إدخال قيم المشاهدات فى عمود، والرموز الخاصة بالعينات فى عمود اخر، حيث تم إعطاء الرمز (1) لبيانات العينة الأولى، والرمز (2) لبيانات العينة الثانية، والرمز رقم (3) لبيانات العينة الثالثة كما يلي:



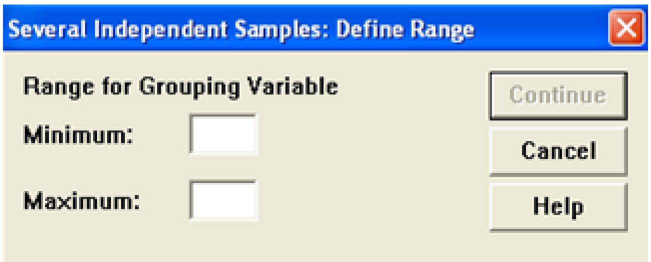
ثانيا: خطوات تنفيذ الاختبار:

نضغط على قائمة Analyze ومن القائمة الفرعية لـ Nonparametric tests نختار k independent Samples كما هو موضح بالشكل التالى:



انقل المتغير samples الى المربع الذى بعنوان Test Variable List ثم انقل متغير الاكواد codes الى المربع الصغير الذى بعنوان Grouping Variable (لاحظ أن الإختيار الإفتراضى من جانب البرنامج هو إختبار كروسكال – والس)

اضغط Define Groups سوف يظهر مربع حوارى جديد كما يلى:



فى خانة Minimum اكتب أصغر الرمز (1) ، وفى خانة Maximum اكتب أكبر الرمز (3) ، ثم اضغط Continue للعودة الى المربع الحوارى السابق.

ثم اضغط Ok سوف تظهر لك نافذة المخرجات الخاصة بهذا الإختبار كالتالي:

Ranks

	CODES	N	Mean Rank
SAMPLES	1	8	16.88
	2	8	10.75
	3	8	9.88
	Total	24	

Test Statistics[a,b]

	SAMPLES
Chi-Square	4.706
df	2
Asymp. Sig.	.095

يلاحظ من نتائج هذا الإختبار أن قيمة P.Value تساوى 0.095 وهى أكبر من مستوى المعنوية 5% ،

وبالتالى فاننا نقبل الفرض العدمى بأن متوسط درجات مادة الإقتصاد فى كلية إدارة الأعمال فى الجامعات الثلاثة متساوى، أي أن الفروق بين الجامعات الثلاثة غير معنوية.

ماذا لو كانت الفروق معنوية، بمعنى أن هناك إختلاف بين الجامعات الثلاثة فى مستوى الطلاب فى مادة الإقتصاد، ونريد تحديد مصدر الإختلاف. مصدر الإختلاف – فى الإختبارات اللامعلمية - لا يتم تحديده بإسلوب مباشر، لأنه فى هذه الإختبارات لايوجد إسلوب للإختبارات البعدية Post Hoc كما فى حالة الإختبارت المعلمية.

ولكن يتم تحديد مصدر الإختلاف بإسلوب غير مباشر، من خلال إجراء إختبار مان – ويتنى لكل عينتين من العينات الثلاثة مع بعضهما البعض، بمعنى نجري إختبار مان – ويتني للعينة (1) مع العينة (2) ، ثم للعينة (1) مع العينة (3) ، ثم للعينة (2) مع العينة (3) ، وبناء على نتيجة هذا الإختبار نحدد مصدر الإختلاف (ونود ان نشير هنا الى أننا إعتمدنا على إختبار مان – ويتني لأننا بصدد عينات مستقلة).

المحاضرة 14

اختبار الفروض الإحصائية اللامعلمية (الجزء الثاني)

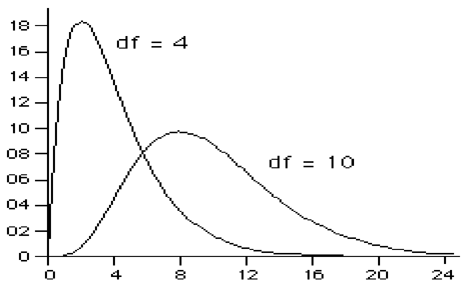
اختبارات الفروض باستخدام توزيع كاي تربيع (χ^2) Test Hypothesis Using Chi-Square Distribution

يعتبر توزيع كاي تربيع χ^2 من التوزيعات الإحتمالية الشائعة الاستخدام حيث توجد له تطبيقات عديدة بدرجة يمكن معها القول أنه يأتي في المرتبة الثانية للتوزيع المعتدل من حيث كثرة تطبيقاته

توزيع كاي تربيع χ^2 :

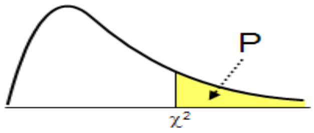
يعتمد توزيع χ^2 مثل توزيع t اعتمادا كاملا على درجات الحرية، وعلى الرغم من ذلك يوجد اختلاف رئيس بين التوزيعين حيث نجد أن توزيع t متماثل حول وسطه الحسابي ($\mu=0$)، بينما يعتبر توزيع χ^2 توزيعا ملتويا جهة اليمين (التواء موجب) وخاصة عندما تكون درجات الحرية صغيرة، **وكلما زادت درجات الحرية كلما قل التواء التوزيع واقترب من التماثل.**

شكل توزيع كاي تربيع χ^2



يبين الشكل السابق أن توزيع χ^2 بدرجات حرية 10 يعتبر تقريبا توزيعا متماثلا، وعندما تكون درجات الحرية كبيرة بدرجة كافية فإنه يمكن تقريب توزيع χ^2 باستخدام التوزيع المعتدل. وعلى أي حال فـإن توزيع χ^2 يعتبر توزيعا مستمرا ذو قمة واحدة

جدول توزيع χ^2



	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124
9	1.735	2.700	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.920	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.300	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32.000	34.267	37.146	39.252
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.790
18	6.265	8.231	22.760	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312
19	6.844	8.907	23.900	27.204	30.144	32.852	33.687	36.191	38.582	41.610	43.820
20	7.434	9.591	25.038	28.412	31.410	34.170	35.020	37.566	39.997	43.072	45.315
21	8.034	10.283	26.171	29.615	32.671	35.479	36.343	38.932	41.401	44.522	46.797
22	8.643	10.982	27.301	30.813	33.924	36.781	37.659	40.289	42.796	45.962	48.268
23	9.260	11.689	28.429	32.007	35.172	38.076	38.968	41.638	44.181	47.391	49.728
24	9.886	12.401	29.553	33.196	36.415	39.364	40.270	42.980	45.559	48.812	51.179
25	10.520	13.120	30.675	34.382	37.652	40.646	41.566	44.314	46.928	50.223	52.620
26	11.160	13.844	31.795	35.563	38.885	41.923	42.856	45.642	48.290	51.627	54.052
27	11.808	14.573	32.912	36.741	40.113	43.195	44.140	46.963	49.645	53.023	55.476
28	12.461	15.308	34.027	37.916	41.337	44.461	45.419	48.278	50.993	54.411	56.892
29	13.121	16.047	35.139	39.087	42.557	45.722	46.693	49.588	52.336	55.792	58.301
30	13.787	16.791	36.250	40.256	43.773	46.979	47.962	50.892	53.672	57.167	59.703
31	14.458	17.539	37.359	41.422	44.985	48.232	49.226	52.191	55.003	58.536	61.098

شكل مقطعي لجدول توزيع χ^2

P **DF**	**المساحة تحت المنحنى إلى يمين قيمة χ^2**						
	0.995	**.......**	**0.975**	**.......**	**0.05**	**.......**	**0.01**
. . .							
10	**2.156**		**3.247**		**18.307**		**23.209**
. . .							

فنلاحظ من الجدول السابق لـ χ^2 أن:

الصف العلوي لجدول توزيع χ^2 يحوي على الاحتمالات الشائعة الاستخدام والتي تمثل مساحة الطرف الأيمن للمنحنى، ويحوي كل صف من الصفوف التي تلي الصف العلوي على القيم المختلفة المناظرة لتوزيع معين من توزيعات χ^2 بدرجات حرية محددة، وتوجد درجات الحرية بالعمود الأول بيسار الجدول، وبالتالي فإن القيمة الموجودة أمام درجات حرية معينة وتحت احتمال محدد هي قيمة χ^2 المجدولة لـ كاي تربيع بدرجات الحرية المحددة والتي تكون المساحة تحت المنحنى على يمينها مساوية للإحتمال المحدد.

فمثلا نجد أن قيم χ^2 بدرجات حرية 10 والتي تساوي المساحة على يمينها 0.05, 0.01 من المساحة الكلية هي 18.307, 23.209 على التوالي.

وقيم χ^2 جميعها موجبة أي أنها أكبر من أو تساوي الصفر. ويرمز لـ χ^2 بالرمز $\chi^2_{(v,\alpha)}$ حيث يمثل الدليل السفلي الأول درجات الحرية بينما يمثل الدليل السفلي الثاني المساحة على يمين % القيمة.

فمثلا $\chi^2_{(10,0.05)}$ تمثل قيمة χ^2 بدرجات حرية 10 والتي تساوي المساحة على يمينها 0.05 وبالنظر للجدول نجد أن $\chi^2_{(10,0.05)} = 18.307$

استخدامات توزيع χ^2

اختبار تباين المجتمع

يستخدم توزيع χ^2 في إجراء العديد من الاختبارات الإحصائية مثل:

الاختبارات المتعلقة بتباين مجتمع ما (وذلك لاختبار المشاكل التي تتطلب اختبار تشتت مجتمع ما)، ويتم ذلك من خلال استخدام المعادلة التالية: $\chi^2 = \frac{(n-1)S^2}{\sigma^2}$

ويفترض في هذا الاختبار أن العينة مسحوبة من مجتمع معتدل وذلك من خلال مقارنة قيمة χ^2 المحسوبة من المعادلة بالقيمة الحرجة لـ χ^2 والمستخرجة من جداول χ^2 .

مثال:

إذا عرف أن تباين قوة مقاومة الكسر للكابلات التي تنتجها إحدى الشركات لاتزيد عن 40000 ، وتستخدم الشركة الآن طريقة إنتاج جديدة يعتقد أنها ستزيد من تباين قوة مقاومة الكابلات للكسر، سحبت عينة عشوائية من عشرة كابلات فوجد تباينها يساوي 50000 .

بافتراض أن قوة مقاومة الكسر للكابلات تتبع التوزيع المعتدل، اختبر الفرض القائل بوجود زيادة معنوية في التباين عند مستوى معنوية α = 0.01

الحل:

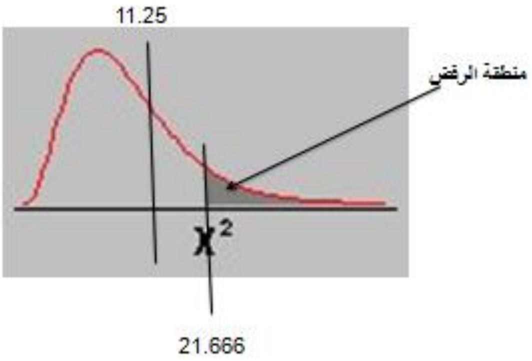
- وضع فرض العدم والفرض البديل.
  صياغة الفرضية الصفرية كالتالي: $H_0 : \sigma^2 \le 40000$
  في حين تفترض الفرضية البديلة التالي: $H_A : \sigma^2 > 40000$
- تحديد مستوى الدلالة (α): وهي 0.01 .
- درجات الحرية = 9 ، فإن قيمة χ^2 المجدولة هي 21.666

لذا فإن قاعدة القرار هي أن يتم رفض الفرضية الصفرية H_0 عندما تكون $\chi^2 \ge 21.666$

وحيث أن قيمة χ^2 لاختبار تباين المجتمع يتم حسابها كالتالي:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{(10-1)50000}{40000} = \frac{(9)50000}{40000} = \frac{450000}{40000} = 11.25$$

وحيث أن قيمة χ^2 المحسوبة أقل من قيمة χ^2 المجدولة، فإننا بالتالي نقبل الفرضية الصفرية H_0 عند مستوى دلالة 0.01 وبالتالي يمكننا القول أن بيانات العينة تدل على أن الزيادة الظاهرة في التباين ليست معنوية عند مستوى الدلالة المحدد ، والشكل التالي يوضح ذلك.



اختبار مربع كاي لجودة التوفيق Testing of Goodness of Fit

ويهتم هذا النوع من الاختبارات الإحصائية باختبار ما إذا كانت مشاهدات عينة تم اختيارها من مجتمع له توزيع احتمالي معين أو نظرية معينة.

ويستخدم هذا الاختبار عندما تكون **البيانات اسمية** أو **على شكل تكرارات** ويقصد بجودة التوفيق هنا دراسة مدى تشابه تكرارات العينة والتي تسمى عادة بالتكرارات الملاحظة Observed مع التكرارات المتوقعة Expected للمتغير موضوع الدراسة فى المجتمع الأصلى.

ويستخدم اختبار كا٢ كطريقة إحصائية للمقارنة بين التكرارين الملاحظ والمتوقع. فإذا كانت العينة ممثلة للمجتمع فى تكراراتها ومتطابقة معه فإن قيمة كا٢ تكون عادة صفراً وتزداد هذه القيمة لتصبح أكثر من صفر كلما كان هناك فرق بين تكرارات العينة (الملاحظة) وبين تكرارات التوزيع النظرى للمجتمع (المتوقعة).

الفروض الإحصائية:

H_0 : مجموعة المشاهدات التى تم اختيارها تتبع توزيع احتمالى معين أو نظرية معينة.

H_A : مجموعة المشاهدات التى تم اختيارها لاتتفق مع هذا التوزيع أو نظرية معينة.

احصاء الاختبار: $\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$

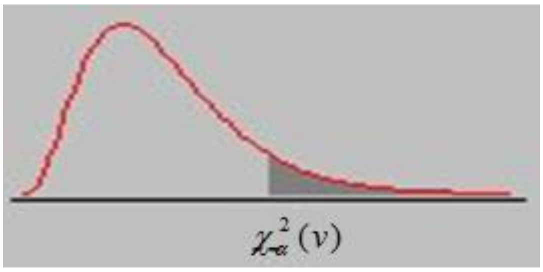
حيث: O_i تمثل التكرار المشاهد للنتيجة رقم i

E_i تمثل التكرار المتوقع المناظر للنتيجة رقم i حيث : $E_i = np_i$

حيث $n = \sum_i O_i$ والقيمة P_i نحصل عليها من التوزيع الإحتمالي أو النظرية المعطاة في فرضية العدم.

ويجب أن يكون التكرار المتوقع في أية خلية لا يقل عن 5 حتى يتم حساب إحصائي الاختبار χ^2 بشكل صحيح.

مناطق الرفض والقبول: سوف نستخدم جدول مربع كاي χ^2 لتعيين القيمة المجدولة (الحرجة) $\chi_\alpha^2(v)$ حيث $v = k^* - 1$ ، و k^* هي عدد المشاهدات النهائية بعد عملية الدمج إن وجدت.



القرار:

نقبل فرض العدم إذا كانت (قيمة كاي تربيع المحسوبة أصغر من القيمة المجدولة) أي أن: $\chi_0^2 < \chi_\alpha^2(v)$

ونرفض فرض العدم إذا كانت (قيمة كاي تربيع المحسوبة أكبر من القيمة المجدولة) أي أن: $\chi_0^2 > \chi_\alpha^2(v)$

مثال:

اختار أحد الباحثين عينة حجمها n=800 شخصا من أحد القبائل، وكان توزيعهم حسب فصيلة الدم كالتالي:

فصيلة الدم	A	B	AB	O
عدد الأشخاص (التكرار المشاهد)	200	150	100	350

هل يتفق هذا التوزيع مع توزيع أفراد قبيلة أخرى كان توزيع فصيلة دمهم حسب النسب التالية:

فصيلة الدم	A	B	AB	O
النسب المئوية للأشخاص	25%	15%	15%	45%

استخدم مستوى معنوية α = 0.05

الحل:

الفروض الإحصائية:

H_0 : توزيع فصيلة الدم في العينة يتفق مع التوزيع المناظر للقبيلة الأخرى.

H_A : توزيع فصيلة الدم في العينة لا يتفق مع التوزيع المناظر للقبيلة الأخرى.

مستوى المعنوية: α = 0.05

احصاء الأختبار: $\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$

حيث O_i تمثل التكرار المشاهد للنتيجة رقم i ،

E_i تمثل التكرار المتوقع المناظر للنتيجة رقم i حيث $E_i = np_i$

$$E_1 = np_1 = 800(0.25) = 200$$
$$E_2 = np_2 = 800(0.15) = 120$$
$$E_3 = np_3 = 800(0.15) = 120$$
$$E_4 = np_4 = 800(0.45) = 360$$

ونلاحظ أن جميع المشاهدات المتوقعة أكبر من % وأيضا حجم العينة، لذا يمكن تعيين احصاء الاختبار كاي تربيع لاختبار هذه البيانات، وبكتابة كلا من المشاهدات والقيم المتوقعة معا في جدول واحد كالتالي:

فصيلة الدم	A	B	AB	O
عدد الأشخاص (التكرار المشاهد)	200	150	100	350
التكرار المتوقع	200	120	120	360

وبتطبيق معادلة كاي تربيع للحصول على قيمة كاي المحسوبة ويتم ذلك كالتالي:

$$\chi_0^2=\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i}=\left[\frac{(200-200)^2}{200}+\frac{(150-120)^2}{120}+\frac{(100-120)^2}{120}+\frac{(350-360)^2}{360}\right]=11.11$$

مناطق الرفض والقبول:

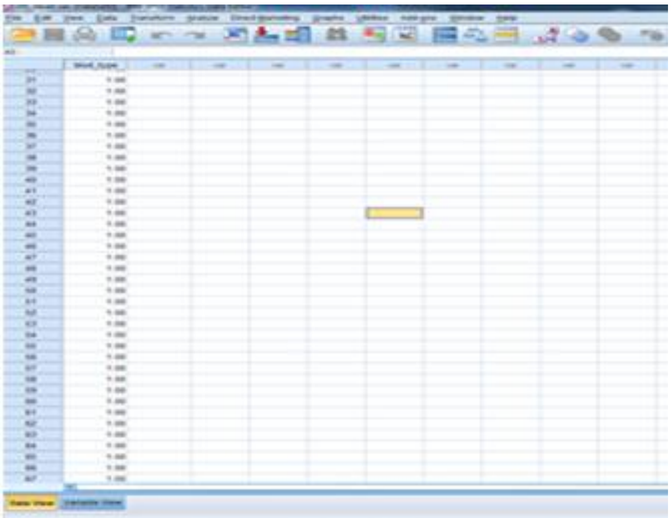
عند استخدام جدول مربع كاي لتعيين القيمة المجدولة لـ $\chi_\alpha^2(v)$ حيث v=4-1=3 ، وبالتالي فإن $\chi_{0.05}^2(3)=7.815$

القرار:

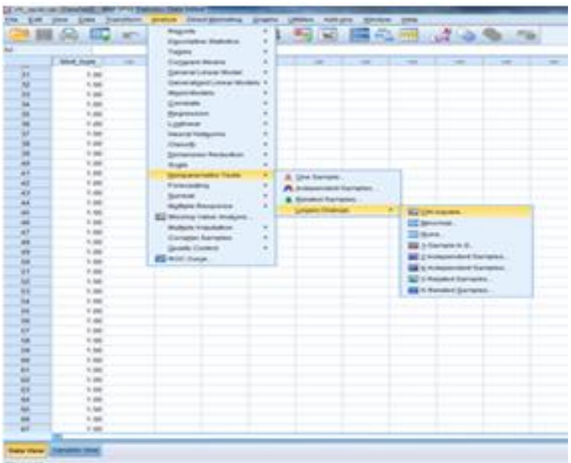
ونلاحظ أن $\chi_0^2>\chi_\alpha^2(v)$ لذا سوف نرفض فرضية العدم وبالتالي فإن توزيع فصيلة الدم في القبيلتين مختلف

حساب اختبار مربع كاي (كا2) لجودة التوافيق Testing of Goodness of Fit -Chi Squre من خلال برنامج SPSS

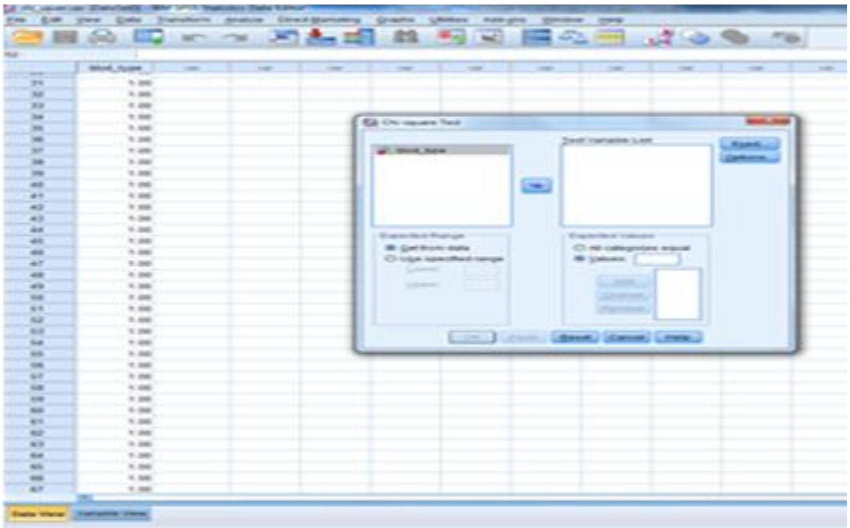
قم بإدخال البيانات على شكل تكرارات في عمود واحد كالتالي:



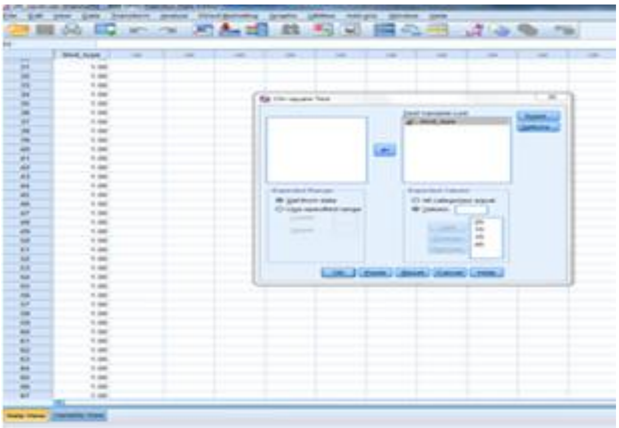
بعد إدخال البيانات قم باختيار Analyze ثم Nnparametric Tests ثم Legacy dialogs ثم ... Chi-Squre



سيفتح لك صندوق حوار كالتالي:



قم بنقل المتغير موضع البحث إلى الصندوق المعنون Test Variable list وقم بعد ذلك بتحديد التكرار المتوقع من خلال إدخال الاعداد أو النسب المحددة للتكرار المتوقع بشكل متتابع حسب ترتيب القيم الملاحظة



بعد ذلك إنقر على زر OK فيقوم البرنامج بحساب قيمة Chi_Squre والجداول المحددة لذلك كالتالي:

يظهر لنا من خلال المخرجات أن قيمة كا^{2} المحسوبة 11.11 ودرجات الحرية 3 ومستوى الدلالة .Sig = 0.011 وهذه القيمة تعني أن قيمة كا^{2} دالة إحصائيا أي توجد فروق ذات دلالة إحصائية بين المجموعتين

وهذه النتيجة ستدفعنا إلى أن نرفض فرضية العدم وبالتالي فإن توزيع فصيلة الدم في القبيلتين مختلف

اختبار مربع كاي للاستقلالية (الإعتمادية) Testing of Independence

كاي تربيع للاستقلالية (Chi-Square test of independency) هو اختبار بسيط يقوم به الباحث لمعرفة ما إذا كان هناك علاقة بين شيئين أو متغيرين. يجرى هذا الاختبار عن طريقة مقارنة قيمة يحددها الباحث مسبقا تعرف بمستوى المعنوية (الفا) بالقيمة المسماة p-Value تحسب من البيانات التوفرة، حيث سيتضح عن طريق المقارنة بين القيمتين ما إذا كانت هنالك علاقة بين الاثنين أم لا

فرضية العدم (Null hypothesis): لا توجد أي علاقة بين المتغيرين ويرمز لهذه الفرضية H0 والذي يتم افتراض صحته عند القيام بالاختبار.

عند القيام بالاختبار لمتغيرين، تكتب هذه الفرضية بهذه الطريقة: **V_1 مستقل عن V_2** ، حيث V_1 و V_2 تمثل المتغيرين تحت الدراسة. ويمكن كتابة فرض العدم الإحصائي بالشكل التالي:

H_0: V_1 is independent of V_2

الفرض البديل (Alternative hypothesis): توجد علاقة بين المتغيرين تحت الدراسة ويرمز لهذه الفرضية H_A وتكتب الطريقة التالية: V_1 **غير مستقل** أو يتبع لـ V_2 ، حيث V_1 و V_2 المتغيرين تحت الدراسة. ويمكن كتابة الفرض البديل بالشكل التالي:

H_A: V_1 is dependent on V_2

مستوى المعنوية (Level of Significance) الفـا:

عند إجراء إختبار كاي تربيع فإن على الباحث اختيار قيمة تسمى Level of Significance أو مستوى المعنوية (الفا) وهذه القيمة يمكن القول بأنها تمثل احتمال الوقوع في خطأ في الاختبار يسمى الخطأ من النوع الأول وهو رفض فرض العدم H_0مع أنه صحيح. بمعنى أن يستنتج الباحث بناء على البيانات المتوفرة أن هنالك علاقة بين المتغيرين مع أنه لا توجد علاقة وهو استنتاج خاطئ.

هذه القيمة التي يحددها الباحث يقوم بمقارنتها بقيمة تسمى p-value والتي يمكن حسابها يدويا أو باستخدام أحد البرامج الإحصائية وذلك من البيانات التي جمعها الباحث.

غالبا في الأبحاث ما يتم استخدام قيمة الفا أو Level of Significance على أنها 0،01 أو 0،05، و الاختيار يرجع للباحث ومدى مجال الخطأ الذي يود أن يسمح به، حيث في حالة إختيار الفا = 0،01 فإن نتيجة الاختبار تكون أدق.

المختبر الإحصائي:

التكرار المتوقع: $\dfrac{\text{مجموع الصف} \times \text{مجموع العمود}}{\text{حجم العينة}}$

نكرر تطبيق هذه المعادلة لجميع الصفوف والأعمدة لكلا المتغيرين

تحديد درجات الحرية: درجات الحرية = (عدد الصفوف – 1) × (عدد الأعمدة – 1)

تحديد قيمة كا^2 المجدولة: يتم بعد ذلك تحديد قيمة كا^2 المجدولة من خلال الرجوع إلى جدول كا^2 عند درجة حرية محددة وفقا لمعطيات الدراسة

القرار: نقارن كا^2 المحسوبة بالجدولية، فعندما تكون قيمة كا^2 المحسوبة **أكبر** من قيمة كا^2 المجدولة فإننا **نرفض** الفرضية الصفرية أو فرض العدم والتي تنص على أنه لا توجد أي علاقة بين المتغيرين ونقبل الفرض البديل والتي تثبت وجود علاقة بين المتغيرين تحت الدراسة.

أما إذا كانت قيمة كا^2 المحسوبة **أقل** من قيمة كا^2 المجدولة فإننا **نقبل** الفرضية الصفرية أو فرض العدم

حساب اختبار مربع كاي (كا^2) للإستقلالية Test of Independence -Chi Squre من خلال برنامج SPSS

مثال:

في دراسة للعلاقة بين التقدير الذي يحصل عليه الطالب في الجامعة وجنسه أخذت عينة من نتائج الطلاب الذكور و الاناث وكانت كما يلي:

أولا: الاناث

راسب	راسب	راسب	راسب	جيد جدا	ممتاز	مقبول	ممتاز
جيد	جيد جدا	جيد جدا	جيد	مقبول	مقبول	مقبول	راسب
مقبول	راسب	مقبول	مقبول	مقبول	راسب	جيد جدا	جيد جدا
جيد	جيد	ممتاز	جيد جدا	ممتاز	جيد	جيد	جيد
					جيد جدا	ممتاز	جيد

ثانيا: الذكور

راسب	جيد	جيد	جيد	راسب	جيد جدا	راسب	جيد جدا
جيد جدا	جيد	راسب	راسب	راسب	راسب	راسب	مقبول
مقبول	ممتاز	ممتاز	راسب	راسب	مقبول	مقبول	ممتاز
ممتاز	جيد	جيد	مقبول	راسب	راسب	جيد	جيد
			جيد جدا	ممتاز	جيد	جيد جدا	ممتاز

والمطلوب:

هل توجد علاقة بين تقدير الطالب وجنسه عند مستوى الدلالة $\alpha = 0.05$ ؟

الحل:

الفرضية الصفرية: تقدير الطالب لا يعتمد على جنسه (متغير الجنس والتقدير مستقلان)

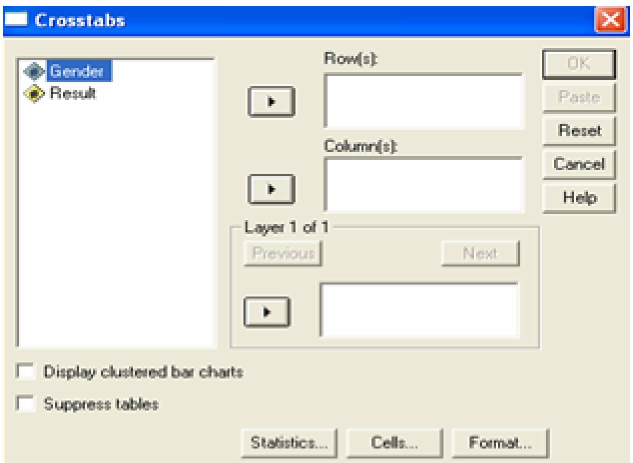
الفرضية البديلة: تقدير الطالب يعتمد على جنسه (توجد علاقة بين جنس الطالب وتقديره)

ثم نقوم بتعريف متغيرين نوعيين هما (*Result*) و (*Gender*) في شاشة تعريف المتغيرات بحيث يكون كود متغير (*Result*) هو (0= راسب، 1=مقبول، 2=جيد، 3 جيد جدًا، 4=ممتاز) وكود المتغير (*Gender*) هو (1=ذكر، 2=انثى)

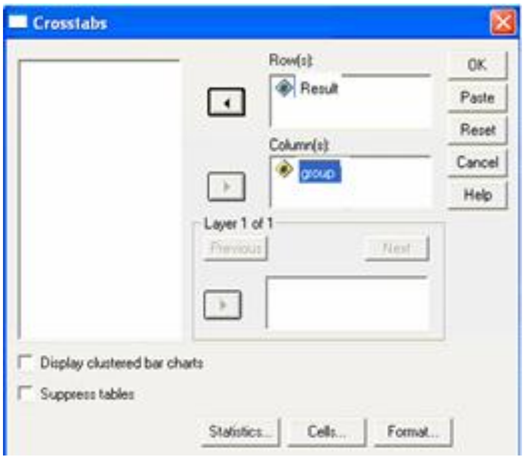
ندخل البيانات كما في الشكل التالي:

من قائمة التحليل Analyze نختار القائمة الفرعية للإحصاءات الوصفية Descriptive Statistics ومن ثم نختار الأمر Cross tabs كما في الشكل التالي

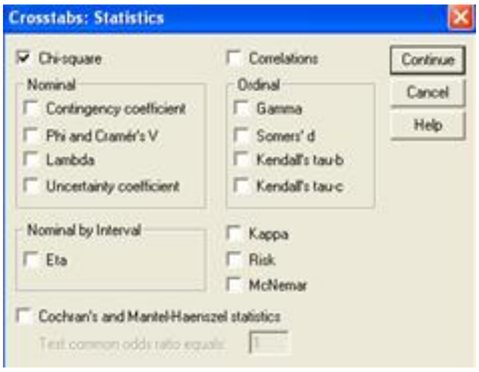
يظهر المربع الحواري التالي:



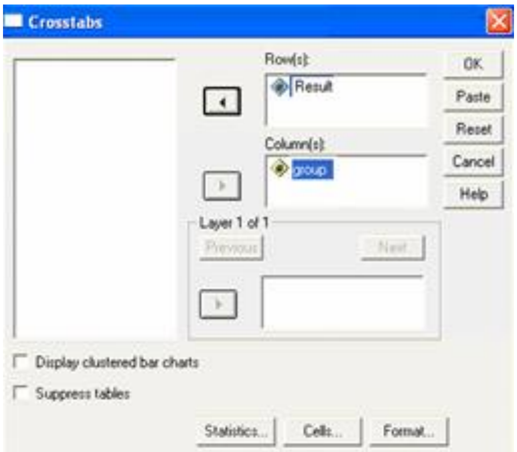
ننقل المتغير Result لخانة الصفوف Rows والمتغير Gender لخانة الأعمدة Columns باستخدام الأسهم .



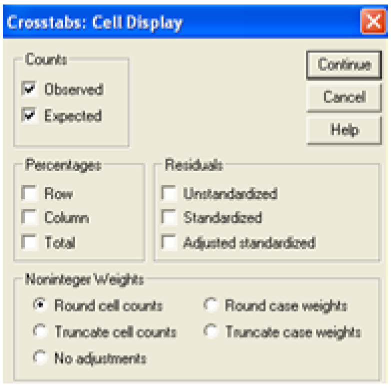
ومن ثم نضغط على Statistics للحصول على المربع الحواري التالي:



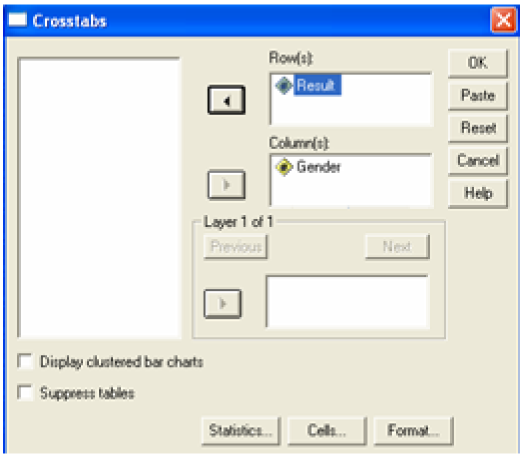
نضع علامة على خانة اختبار مربع كاي Chi-Square لحساب اختبار الاستقلالية ومن ثم نضغط على Continue للعودة للمربع الحواري السابق:



لاظهار جدول التوقعات نضغط على زر Cell ليظهر المربع الحواري التالي:



نختار الخيار Expected جدول توقعات ظهور البيانات ومن ثم نضغط Continue للعودة للمربع الحواري السابق.



نضغط على Ok للحصول على النتائج.

تتكون نتائج الأمر Cross tabulati من ثلاثة جداول:

<u>الأول</u> يصف حجم العينات المدخلة ونسب البيانات المفقودة كالتالي:

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
group * Result	72	100.0%	0	.0%	72	100.0%

الجدول <u>الثاني</u> يبين جدول توزيع العينة حسب المتغيرين والقيم المتوقعة حسب اختبار الاستقلالية كالتالي:

عدد الذكور الراسبين

group * Result Crosstabulation

			Result		Total
			ذكر	انثى	
group	راسب	Count	12.00	7.00	19
		Expected Count	9.76	9.24	19.0
	مقبول	Count	5.00	8.00	13
		Expected Count	6.68	6.32	13.0
	جيد	Count	9.00	8.00	17
		Expected Count	8.74	8.26	17.0
	جيد جدا	Count	5.00	7.00	12
		Expected Count	6.17	5.83	12.0
	ممتاز	Count	6.00	5.00	11
		Expected Count	5.65	5.35	11.0
Total		Count	37.00	35.00	72
		Expected Count	37.00	35.00	72.0

توقع الذكور الراسبين

الجدول الثالث يبين نتيجة اختبار مربع كاي كالتالي:

قيمة الاختبار

درجة الحرية

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.437[a]	4	.656
Likelihood Ratio	2.459	4	.652
Linear-by-Linear Association	.298	1	.585
N of Valid Cases	72		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.35.

مستوى دلالة الاختبار

يبين الجدول الثالث السابق أن قيمة اختبار مربع كاي هي 2.437 بدرجة حرية مقدارها 4

يتبين لنا من الجدول أن أقل قيمة لمستوى الدلالة هي *Asymp. Sig. (2-sided) = 0.656* وهي اكبر من مستوى الدلالة $\alpha = 0.005$ وبالتالي لا نستطيع رفض الفرضية الصفرية أي أن تقدير الطالب لا يعتمد على جنسه.

اختبار كولمجروف سيمرنوف لجودة التوافيق Goodness of Fit Test - Kolmogorov-Smirnov :

استخدامه:

يستخدم هذا الاختبار لمعرفة إذا ما كانت العينة موضع الاهتمام تتبع توزيعاً احتمالياً معينا **ويستخدم عوضاً عن اختبار مربع كاي** عندما يكون مجموع التكرارات أقل من 30 أو يكون التكرار المتوقع لأي خلية أقل من خمسة وعملية ضم الخلايا تؤدي إلى فقد كثير من درجات الحرية مما يتعذر معه إجراء الاختبار أو أن تكون عملية الضم غير مناسبة. ويفضل استخدامه أيضاً في حالة كون التوزيع الاحتمالي لمتغير متصل.

ويستخدم لاختبار ما إذا كانت عينتان لهما نفس التوزيع ويعتمد إجراء هذا الاختبار على دالة الاحتمال التجميعي للتكرار المشاهد والمتوقع وبذلك يدور الفرض العدمي والبديل حول هاتين الدالتين وهو كالتالي :

$H_0 : F_n(X) = F_0(X)$

$H_A : F_n(X) \neq F_0(X)$

$F_n(X) < F_0(X)$

$F_n(X) > F_0(X)$

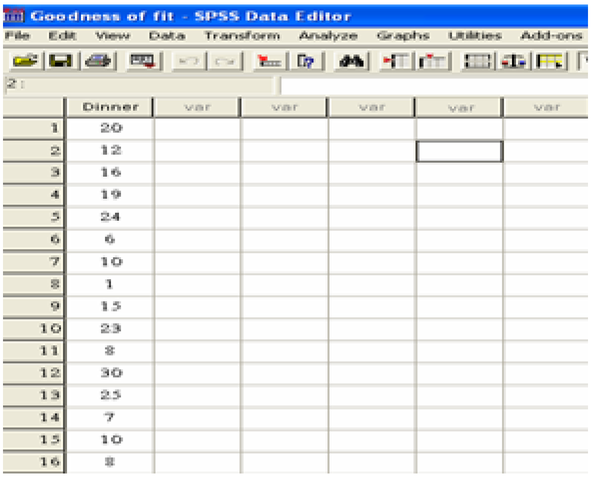
ويكون إحصاء الاختبار التالي :

$D = \max | F_n(X) - F_0(X) |$

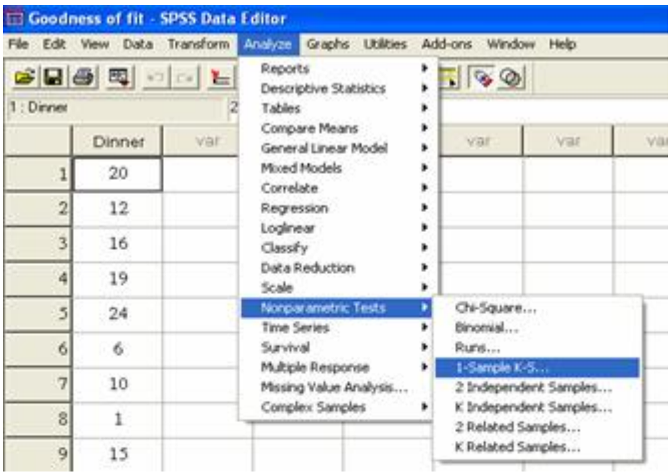
وتستخرج القيمة الجدولية لهذا الاختبار من جداول خاصة عند مستوى معنوية α عند درجات حرية تناظر مجموع التكرارات، ويرفض الفرض العدمي عندما تزيد القيمة المحسوبة عن القيمة الجدولية

حساب اختبار كولومجروف سيمرنوف لجودة التوافق Goodness of Fit - Kolmogorov-Smirnov Test من خلال برنامج SPSS

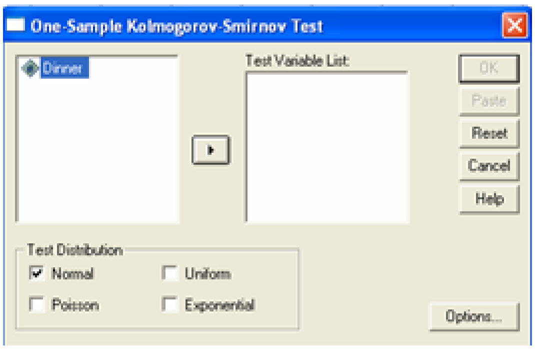
ندخل البيانات في متغير نسميه Dinner كما في الشكل التالي:



من قائمة التحليل Analyze نختار القائمة الفرعية الاحصاءات الغير بارامترية Non-Parametric Test ومن ثم نختار الأمر 1-Sample K-S



يظهر المربع الحواري التالي:



يمكنك المربع الحواري السابق من اختيار التوزيع الذي تريد اختباره هل هو توزيع طبيعي Normal أو بواسون Poisson أو منتظم Uniform أو أسي Exponential فنختار التوزيع الطبيعي كما في الشكل أعلاه ونضغط Ok للحصول على النتائج التالية:

NPar Tests

One-Sample Kolmogorov-Smirnov Test

		Dinner
N		50
Normal Parameters[a,b]	Mean	15.26
	Std. Deviation	6.782
Most Extreme Differences	Absolute	.081
	Positive	.081
	Negative	-.069
Kolmogorov-Smirnov Z		.573
Asymp. Sig. (2-tailed)		.898

a. Test distribution is Normal.

b. Calculated from data.

حجم العينة

متوسط البيانات

الانحراف المعياري للبيانات

أكبر فرق بين البيانات ودالة التوزيع الاحتمالية

قيمة اختبار جودة المطابقة

مستوى دلالة الاختبار

تبين النتائج أعلاه أن متوسط عدد الزبائن هو 15.26 بانحراف معياري قدره 6.782 وأن قيمة اختبار كولموجروف سميرنوف لجودة المطابقة هو 0.573

<u>القرار:</u>

يبين الجدول السابق أن قيمة مستوى دلالة الاختبار هي *Asymp. Sig. (2-tailed) = 0.898* وهي اكبر من مستوى دلالة الفرضية الصفرية $\alpha = 0.05$ وبالتالي نقبل الفرضية الصفرية، <u>أي أن البيانات تتبع التوزيع الطبيعي</u> وبالتالي نستنتج ان البيانات تتبع التوزيع الطبيعي بمتوسط قدره 15.26 وانحراف معياري 6.782 أي $X : N\ (15.26, 6.782)$

وإذا أردنا اختبار أن التوزيع يتبع توزيع بواسون نختار من الشاشة المخصصة لذلك توزيع بواسون وهكذا مع باقي التوزيعات.

تمنياتي لكم بالتوفيق والنجاح ☺ w r o o d