



التحليل الإحصائي

د.أحمد فرحان

إعداد :

Ghayda&dody-11

2014-2015

المحاضرة (١)

المجموعات

تعريف المجموعة :-

يمكن تعريف المجموعة على أنها عدد من العناصر بينها صفات مشتركة تكتب بين حاصرتين { } و تسمى بأحد الحروف الهجائية الكبيرة , C , B , A

و من الأمثلة على المجموعات $A = \{ 1,3,5,7,9,..... \}$

الأشياء التي تتكون منها المجموعة تسمى عناصر المجموعة و ترمز للعناصر بواسطة حروف صغير مثل:-

a , b , c ,

يستخدم الرمز e "ينتمي إلى" ليعين عناصر المجموعة فمثلاً إذا كان العنصر a من ضمن عناصر المجموعة A فإننا نقول أن a ينتمي إلى المجموعة A و يكتب بالصورة $a \in A$

أما إذا كان a ليس عنصراً من عناصر المجموعة A فإننا نقول أن العنصر a لا ينتمي إلى المجموعة A و يكتب على الصورة $a \notin A$

طريقة كتابة المجموعات :

طريقة العد (سرد العناصر) :-

يتم فيها وضع جميع عناصر المجموعة ، أو جزء منها ، بين قوسي المجموعة { } بحيث يفصل بين كل عنصرين بعلامة فاصلة " , " :-

مثال :-

$$A = \{ 1,5,10,15 \}$$

$$B = \{ a , b , c , d \}$$

$$C = \{ 1 , 2 , 3 \dots \}$$

(و هي مجموعة منتظمة تسير بنفس الشكل ١ ٢ ٣ ٤ وهكذا)

$$A = \{ 1, 2, 3, \dots, 100 \}$$

(و هي مجموعة مغلقة و للثن المساحة لا تكفي لكتابة من ١ إلى ١٠٠ و سوف نستخدم النقاط للتعبير عن بعض العناصر)

طريقة القاعدة (الصفة المميزة) :-

ويتم فيها وصف المجموعة بذكر صفة يمكن بواسطتها تحديد عناصرها ، أي الصفة التي تحدد ارتباط عناصر المجموعة ، فمثلاً :

$$A = \{x: \text{عدد زوجي}\}$$

$$B = \{x: \text{طالب بمقرر الاحصاء في الادارة}\}$$

$$C = \{x: \text{طالب بنظام التعليم عن بعد}\}$$

$$D = \{x: \text{عدد صحيح } -3 \leq x \leq 1\}$$

$$x = \{x: \text{عدد صحيح } 0 \leq x \leq 12\}$$

أنواع المجموعات :-

المجموعة الخالية :-

هي المجموعة التي لا تحتوي أي عنصر ويرمز لها بالرمز ϕ (فاي) أو $\{ \}$.

أمثلة :-

$$A = \{x: \text{عدد طبيعي زوجي و فردي}\}$$

$$B = \{x: \text{دولة عربية تقع في أمريكا الشمالية}\}$$

المجموعة المنتهية :-

المجموعة التي تكون عناصرها محدودة .

مثال :

المجموعات التالية مجموعات منتهية .

$$A = \{2, 4, 6, 8\}$$

$$B = \{1, 2, 3, \dots, 100\}$$

$$C = \{x, y, s, t, u\}$$

المجموعة غير المنتهية :-

المجموعة التي تكون عناصرها غير محدودة (وهي المجموعة التي لا يمكن تحديد عناصرها بشكل دقيق)

مثال :

المجموعات التالية مجموعات غير منتهية .

$$A = \{x: \text{عدد طبيعي فردي}\}$$

$$B = \{10, 20, 30, \dots\}$$

المجموعة الكلية :-

هي المجموعة التي تدرس جميع المجموعات باعتبارها مجموعات جزئية و يرمز لها بالرمز U .

المجموعة الجزئية :-

تكون المجموعة A جزئية من المجموعة B إذا كانت جميع عناصر A موجودة في B و تكتب على الصورة :-

$$A \subset B$$

أمثلة :-

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

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

تساوي المجموعات :-

تكون المجموعتان A و B متساويتان إذا كانت :-

$$A \subseteq B , B \subseteq A \implies A = B$$

أما المجموعتان المتكافئتان فهما المجموعتان اللتان تتساويان في عدد عناصرها و تكتب على الصورة $A \equiv B$

مثال :

أي المجموعات التالية متكافئة وأيها متساوية ؟

$$1- A = \{1, 5, 7, 9\} , B = \{9, 7, 5, 1\}$$

$$2- A = \{2, 5, 9\} , B = \{a, s, d\}$$

الحل

$$1 - A = B$$

$$2 - A \equiv B$$

العمليات على المجموعات :-

الاتحاد :-

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

مثال :-

إذا كان $A = \{1, 2, 3, 7\}$ و $B = \{2, 4, 6, 8\}$ أوجد $(A \cup B)$ ؟

الحل

$$(A \cup B) = \{1, 2, 3, 4, 6, 7, 8\}$$

التقاطع :-

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

مثال :-

$$\text{إذا كان } A = \{-1, 0, 1, 2, 3\} \text{ و } B = \{0, 2, 4, 6\} \text{ أوجد } A \cap B$$

الحل

$$(A \cap B) = \{0, 2\}$$

المكملة أو المتممة :-

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

مثال

$$\text{إذا كان } U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\} \text{ و } A = \{2, 4, 6, 8, 10\} \text{ أوجد}$$

الحل

$$\bar{A} = \{1, 3, 5, 7, 9\}$$

الفرق :-

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

مثال :-

1- $A \cup B$

2- $A \cap B$

3- $B - A$

4- \bar{A}

5- \bar{B}

6- $\bar{A} \cup \bar{B}$

7- $\bar{A} \cap \bar{B}$

8- $\bar{A} \cup A$

9- $\bar{A} \cap A$

$$\text{إذا كانت } A = \{1, 2, 3, x, y\} \text{ و } B = \{3, 4, 5, x, w\} \text{ أوجد } A - B$$

الحل

$$A - B = \{1, 2, y\}$$

مثال شامل :

$$\text{إذا كانت } A = \{1, 2, 3, x, y\} \text{ و } B = \{3, 4, 5, x, w\} :-$$

و المجموعة الكلية

$$U = \{1,2,3,4,5,w,x,y,z\}$$

فأوجد :-

الحل

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

$$2- A \cap B = \{3, X\}$$

$$3- B - A = \{4, 5, w\}$$

$$4- \bar{A} = \{4, 5, w, z\}$$

$$5- \bar{B} = \{1, 2, y, z\}$$

$$6- \bar{A} \cup \bar{B} = \{1, 2, 4, 5, y, w, z\}$$

$$7- \bar{A} \cap \bar{B} = \{z\}$$

$$8- \bar{A} \cup A = U$$

$$9- \bar{A} \cap A = \{ \}$$

أشكال فن

VIN Figures

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

1- المجموعة الكلية:

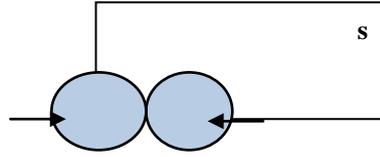
تمثل المجموعات الكلية بمستطيل ويرمز لها بالرمز S



S

٢- إتحاد الحوادث Events Union :

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



شكل فن لتمثيل إتحاد حادثتين A و B

$$(A \cup B)$$

وبشكل عام لأي n حادثة $A_1, A_2, A_3, \dots, A_n$ فإن إتحاد هذه الحوادث هو :

$$\bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup A_3 \dots \cup A_n$$

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

الإتحاد :- 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 :

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

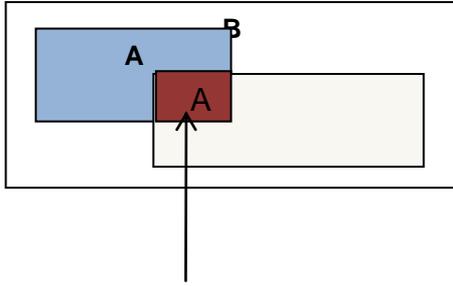
ويعني ذلك توزيع الإتحاد على التقاطع.

- وكذلك هناك خاصية التبديل والتي تعني أن

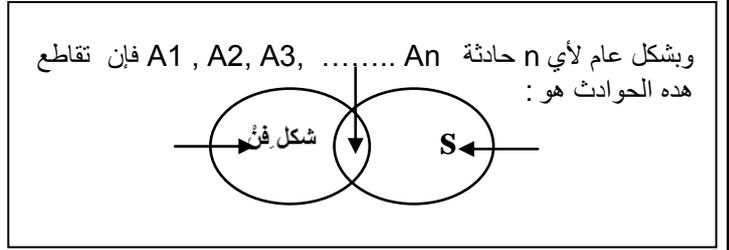
$$(A \cup B) = (B \cup A)$$

٣- تقاطع الحوادث Events Intersection :

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



(A ∩ B)



B

فإن تقاطع هذه الحوادث هو : A1 , A2, A3, An حادثة n وبشكل عام لأي

$$\bigcap_{i=1}^n A_i = A_1 \cap A_2 \cap A_3 \dots \dots \cap A_n$$

ويمكن القول أن $\bigcap_{i=1}^n A_i$ هو حدث يقع إذا وفقط وقعت كل الحوادث A_i على الأقل وهو ما يطلق عليه ضرب الحوادث **التقاطع :-** 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 C و B و A ثلاث حوادث فإن :

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

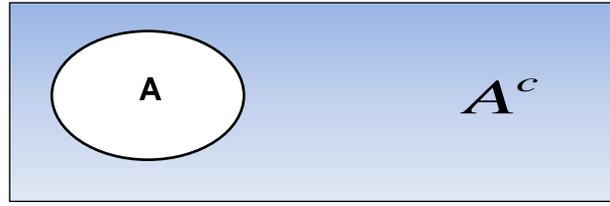
ويعني ذلك توزيع التقاطع على الإتحاد

- وكذلك هناك خاصية التبديل والتي تعني $(A \cap B) = (B \cap A)$

٤- الحادثة المتممة Complementary Event

فإن متممها هي الحادثة التي تتضمن كافة العناصر التي لا تنتمي A ويرمز لها بالرمز A^c أو \bar{A}

وهو حدث يتألف صر غير المنتمية إلى A وباستخدام أشكال فن فإن الجزء المظلل يمثل الحادثة المتممة من



$$\overline{A} \quad A^c$$

شكل فن لتمثيل مكملة حادثة A

مثال:

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

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

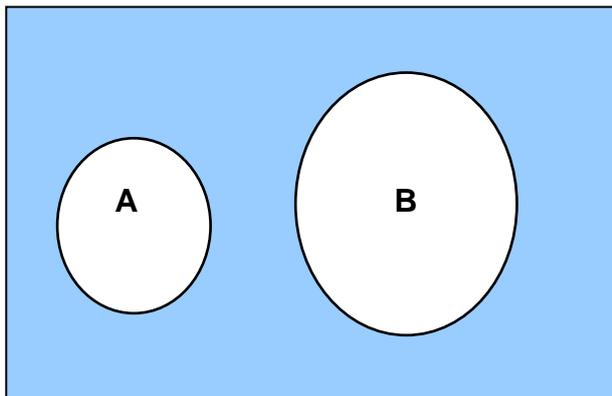
$$B = \{1, 2, 3, 6, 8, 11, 12, 14, 16\}$$

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

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

٥- الحوادث المتنافية Mutually Exclusive Events

الحادثتان A و B متنافيتان أو منفصلتان إذا كان تقاطعهما خاليا أي $A \cap B = \emptyset$ ويمكن القول أيضا أن $A \cap A^c = \emptyset$



$$A \cap B = \emptyset$$

$$A \cap A^c = \emptyset$$

بعض العلاقات المهمة :-

$$A \cup A^c = S$$

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

$$A \cap A^c = \phi$$

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

$$\overline{\overline{S}} = S$$

فإن: إذا كانت $A \subset B$

$$\overline{\overline{\phi}} = S$$

$$A = A \cap B$$

$$A \cup S = S$$

$$B = A \cup B$$

$$A \cap S = A$$

$$\overline{\overline{B}} = B$$

$$A \cap \phi = \phi$$

\cap = و
 \cup = أو

أمثلة وتمارين

يراد شراء ثلاثة أنواع من اللحوم من جزائر 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 :

$$\overline{A \cap B \cap C}$$

- 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

$$(A \cap \overline{B} \cap \overline{C}) \cup (\overline{A} \cap B \cap \overline{C}) \cup (\overline{A} \cap \overline{B} \cap C)$$

تمارين :-

١- وضح أي من هذه المجموعات هي مجموعة خالية أو مجموعة منتهية أو مجموعة غير منتهية :-

(a) $A = \{x : \text{عدد سالب و موجب}\}$

(b) $B = \{3, 6, 9, 12\}$

(c) $C = \{x : \text{دولة أوروبية تقع في شبة الجزيرة العربية}\}$

(d) $D = \{2, 4, 6, \dots, 100\}$

(e) $E = \{100, 200, 300, \dots\}$

(f) $F = \{w, e, r, t\}$

٢- إذا كانت $A = \{3, 5, 7\}$ و $B = \{1, 2, 3, 4, 5, 6, 7, 8\}$ فهل يمكن القول أن $A \subset B$ ؟

٣- أي المجموعات التالية متكافئة وأيها متساوية ؟

1- $A = \{5, 10, 15, 20\}$ ، $B = \{15, 10, 5, 20\}$

2- $A = \{20, 50, 70\}$ ، $B = \{k, d, u\}$

إذا كانت $U = \{0, 1, 2, 3, \dots, 10\}$ وكانت $A = \{2, 4, 6, 8, 10\}$ ، $B = \{1, 2, 3, 4, 5\}$ ، $C = \{6, 7, 8, 9, 10\}$ فجد ما يلي :-

1- $A \cup B$

5- $A \cap \bar{C}$

2- $A \cap C$

6- $A - (B \cap C)$

3- $\bar{A} \cap \bar{B}$

7- $(\bar{A} \cup B) - C$

4- $B \cup C$

8- $\overline{(B \cap \bar{C})}$

حل المثال :-

1- $A \cup B = \{1, 2, 3, 4, 5, 6, 8, 10\}$

2- $A \cap C = \{6, 8, 10\}$

3- $\bar{A} \cap \bar{B} = \overline{A \cup B} = \{0, 7, 9\}$

4- $B \cup C = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

5- $A \cap \bar{C} = A - C = \{2, 4\}$

$$6- A - (B \cap C) = B \cap C = \emptyset$$

$$A - (B \cap C) = A - (\emptyset) = A$$

$$7- (\bar{A} \cup B) - C$$

$$\bar{A} = \{0, 1, 3, 5, 7, 9\}$$

$$\bar{A} \cup B = \{0, 1, 2, 3, 4, 5, 7, 9\}$$

$$(\bar{A} \cup B) - C = \{0, 1, 2, 3, 4, 5\}$$

$$8- \overline{(B \cap C)} = \overline{(B \cup C)} = B \cup C$$

$$B \cup C = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

تمارين :-

٤- إذا كانت

$$A = \{8, 10, 12, r, m\} \text{ و } B = \{4, 6, 10, o, r\}$$

أوجد المجموعة الكلية

ثم أوجد :-

1- $A \cup B$

2- $A \cap B$

3- $B - A$

4- \bar{A}

5- \bar{B}

6- $\bar{A} \cup \bar{B}$

7- $\bar{A} \cap \bar{B}$

8- $\bar{A} \cup A$

9- $\bar{A} \cap A$

٥- 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) $3 \text{ — } A$

(ii) $3 \text{ — } B$

(iii) $x \text{ — } A$

(iv) $x \text{ — } B$

(v) $z \text{ — } A$

(vi) $z \text{ — } B$

(vii) $1 \text{ — } A$

(viii) $1 \text{ — } B$

المحاضرة (٢) نظرية الاحتمالات

تعريف الاحتمالات:

يمكن تعريف الاحتمالات بطرق عديدة غير أن أبسطها

معين "(Event)" هو مقياس لامكانية وقوع حدث

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 لكلمة مرادفة لبعض الكلمات الأخرى مثل: ممكن، غالباً، 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 تطوراً كبيراً وسريعاً وأصبح أساساً للعديد من الإحصاء وبحوث العمليات وغيرها.

١- التجربة العشوائية:-

هي تلك التجربة التي تكون جميع نتائجها معلومة مسبقاً ولكن لا يمكن التنبؤ بحدوث أي من هذه النتائج بصفة مؤكدة مثلاً:

رمي حجر نرد مرة واحدة يعتبر تجربة عشوائية، حيث نعلم جميع قيم نتائج التجربة وهي إما ١ أو ٢ أو ٣ أو ٤ أو ٥ أو ٦ 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

٢- الحادث والفراغ العيني:

فراغ العينة هو المجموعة الشاملة التي تحتوي على جميع النتائج الممكنة للتجربة العشوائية ويرمز له

Ω و يطلق عليه الحالات الممكنة

افتراض أننا نقوم بإجراء تجربة ما كرمي زهرة النرد مثلاً ونلاحظ كل النتائج الممكنة وهي ظهور أحد الأوجه الستة ١ أو ٢ أو ٣ أو ٤ أو ٥ أو ٦ ونفترض أننا مهتمون بظهور رقم فردي أي ١ أو ٣ أو ٥ 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 أو أكثر من الفراغ العيني .

الحادثة هي مجموعة جزئية من فراغ العينة وتمثل مجموعة النتائج التي تحقق الحدث وتسمى أيضا الحالات ، فمثلا الحصول على رقم زوجي في تجربة رمي زهرة النرد مرة واحدة تكون **Favorable Cases** المواتية الحادثة هي { ٢ ، ٤ ، ٦ } ، ويمكن أن تحتوي الحادثة على عنصر واحد أو أكثر

٣- أنواع الحوادث :-

أ- الحوادث المتنافية

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

ب- الحوادث المستقلة

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

ج- الحوادث الشاملة

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 مدخناً أو غير مدخن تعتبر هذه الحالات حوادث شاملة لأنه لابد للفردي أن يكون له صفة واحدة من هذه الصفات كذلك فإن الحصول على العدد ١ أو ٢ أو ٣ أو ٤ أو ٥ أو ٦ عند رمي حجر النرد تعتبر حوادث شاملة لأنه لابد من حدوث إحداها .

مثال

رمي حجر نرد مرد واحدة ، أحسب التالي:

- احتمال الحصول على رقم ٥
- احتمال الحصول على رقم زوجي
- احتمال الحصول على رقم أكبر من ٢
- احتمال الحصول على رقم أقل من ٧
- احتمال الحصول على رقم ٧

الحل:

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

$$P_{(A=5)} = \frac{1}{6}$$

$$P_{(A=2,4,6)} = \frac{3}{6}$$

$$P_{(A>2)} = \frac{4}{6}$$

$$P_{(A<7)} = \frac{6}{6} = 1$$

$$P_{(A=7)} = \frac{0}{6} = 0$$

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

مثال:

الجدول التالي يمثل توزيع عمال أحد المصانع حسب الحالة الاجتماعية للعامل والقسم الذي 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:

- أن يكون أعزبا
- أن يكون متزوجا
- أن يكون من القسم الأول
- أن يكون من القسم الأول أو الثاني
- أن يكون من القسم الأول وأعزب

الحل:

نفرض أن الحادثة **A** أن يكون العامل **أعزب** أي $A = \{ \text{أن يكون العامل أعزب} \}$ فيكون الاحتمال المطلوب

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

عدد العمال الكلي

نفرض أن الحادثة **B** أن يكون العامل **متزوج** أي $B = \{ \text{أن يكون العامل متزوج} \}$ فيكون الإحتمال المطلوب

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

عدد العمال الكلي

نفترض ان الحادثة **C** أن يكون العامل **القسم الأول** أي $C = \{ \text{أن يكون العامل من القسم الأول} \}$ فيكون

الإحتمال المطلوب

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

عدد العمال الكلي

نتفرض ان الحادثة **D** ان يكون العامل القسم الاول او الثاني **D** أي = { ان يكون العامل من القسم الاول او الثاني } فيكون الاحتمال المطلوب

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

عدد العمال الكلي

نتفرض ان الحادثة **E** ان يكون العامل من القسم الاول و أعزب أي ان **E** = { أن يكون العامل من القسم الاول و أعزب } فيكون الاحتمال المطلوب

$$P(E) = \frac{\text{عدد عمال القسم الاول و أعزب}}{\text{عدد العمال الكلي}} = 5/50 = 0.1$$

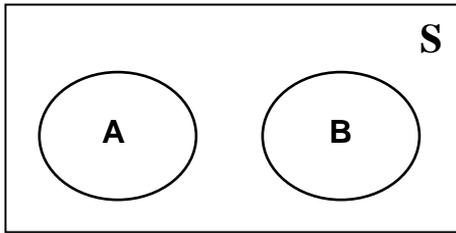
عدد العمال الكلي

جمع الاحتمالات

أ- في حالة كون الحوادث متنافية

إذا كانت الحوادث A_1, A_2, A_3, \dots حوادث متنافية بمعنى أن حدوث أحدها يؤدي 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



حوادث متنافية

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

$$\text{فإن } P(A \text{ أو } B) = P(A) + P(B)$$

يرمز أيضاً لاحتمال وقوع أحد الحادثين بالرمز

حوادث متنافية

$$P(A \cup B) = P(A) + P(B)$$

مثال:

رمي حجر نرد مرة واحدة ، احسب:

احتمال الحصول على رقم ٥ أو ٦

احتمال الحصول على رقم زوجي

الحل:

حيث أن الحصول على رقم ٥ أو ٦ حادثتان متنافيتان ، أي أن:

$A_1 = \{\text{الحصول على الرقم ٥}\}$ ، و $A_2 = \{\text{الحصول على الرقم ٦}\}$ فإن:

$$P(A_1 \cup A_2) = (1/6) + (1/6) = 1/3$$

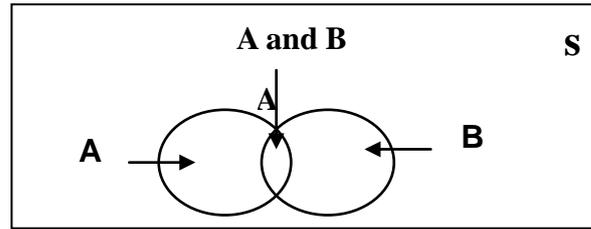
وحيث أن الحصول على رقم زوجي يعني الحصول على رقم ٢ أو رقم ٤ أو رقم ٦ وكلها حوادث متنافي، أي أن:

$A1 = \{\text{الحصول على الرقم ٢}\}$ ، و $A2 = \{\text{الحصول على الرقم ٤}\}$ ، و $A3 = \{\text{الحصول على الرقم ٦}\}$ فإن:

$$P(A1 \cup A2 \cup A3) = (1/6) + (1/6) + (1/6) = 1/2$$

ب- في حالة كون الحوادث غير متنافية

عند عدم اشتراط تنافي الحادثين A و B يكون المقصود بالحدث (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 للحدث A مضافاً إليها مجموع الحالات المواتية للحدث B ولكن يجب ملاحظة أن كل من الحالات المواتية للحدث A وتلك المواتية للحدث B تتضمن الحالات المواتية لوقوع A و B معاً ، وبهذا فإنه في حالة جمع $P(A)$ و $P(B)$ فإننا نجمع $P(A \cup B)$ مرتين ، لهذا لابد من طرح $P(A \cap B)$ مرة واحدة لنحصل على الاحتمال $P(A \cup B)$ وهذا هو :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

أو

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

مثال:-

الجدول التالي يمثل توزيع عمال أحد المصانع حسب الحالة الاجتماعية للعامل والقسم الذي 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

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

- احتمال أن يكون العامل من القسم الأول أو الثاني
- احتمال أن يكون العامل متزوجاً أو من القسم الأول
- احتمال أن يكون العامل من القسم الثالث أو أعزب

الحل:

نفرض أن الحادثة A_1 أن يكون العامل من القسم الأول أي أن $A_1 = \{\text{أن يكون العامل من القسم الأول}\}$
 نفرض أن الحادثة A_2 أن يكون العامل من القسم الثاني أي أن $A_2 = \{\text{أن يكون العامل من القسم الثاني}\}$

فيكون الاحتمال المطلوب:

$$P(A_1) = 12/50$$

$$P(A_2) = 22/50$$

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) = (12/50) + (22/50) = 34/50 = 0.68$$

نفرض أن الحادثة A_1 أن يكون العامل متزوجاً أي أن $A_1 = \{\text{أن يكون العامل متزوج}\}$

نفرض أن الحادثة A_2 أن يكون العامل من القسم الأول أي أن $A_2 = \{\text{أن يكون العامل من القسم الأول}\}$

فيكون الاحتمال المطلوب:

$$P(A_1) = 27/50$$

$$P(A_2) = 12/50$$

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = (27/50) + (12/50) - (7/50) = 32/50 = 0.64$$

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 الاحتمال 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 الحادتين A1 , A2 وكان P(A2) لا يساوي الصفر فإين الاحتمال الشرطي للحادث A1 بشرط وقوع الحادث A2 يعطي بالمعادلة التالية:

$$P(A1 | A2) = \frac{P(A1 \cap A2)}{P(A2)}$$

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

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

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 الذي 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:

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

الحل:

نفرض أن $A_1 = \{\text{أن يكون العامل من القسم الأول}\}$

$A_2 = \{\text{أن يكون العامل متزوج}\}$

$B_3 = \{\text{أن يكون العامل من القسم الثالث}\}$

$B_4 = \{\text{أن يكون العامل أعزب}\}$

فيكون بالتالي:

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

احتمال أن يكون من القسم الأول ومتزوج

احتمال أن يكون متزوج

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

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

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

احتمال أن يكون العامل أعزب بشرط أنه من القسم الثالث

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

إذا احتمال أن يكون العامل أعزب بشرط أنه من القسم الثالث هو 0.625

ضرب الاحتمالات

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

فمثلاً إذا كان لدينا صندوق به ١٠ كرات متماثلة منها ٦ بيضاء و ٤ 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 مكتملاً 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 وهذا ما يسمى سحب بإرجاع أو إحلال أو إعادة.

فإذا كان لدينا الحادثين المستقلين A_1 و A_2 فإن احتمال حدوثهما معا هو:

$$P(A_1 \cap A_2) = P(A_1) P(A_2)$$

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

$$P(A_1 \cap A_2 \cap A_3 \cap \dots \cap A_n) = P(A_1) P(A_2) P(A_3) \dots P(A_n)$$

مثال:

إذا رمينا قطعة نقود مرة واحدة ، احسب الاحتمالات التالية:

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

الحل:

نفرض أن: A_1 = {ظهور صورة في الرمية الأولى}

A_2 = {ظهور كتابة في الرمية الثانية}

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

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

$$P(B_1) = \frac{1}{2}, P(B_2) = \frac{1}{2}$$

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 الحوادث:

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 الذي يعمل به فإذا سحب عاملان من المصنع مع الإرجاع (أي إرجاع العامل الأول قبل سحب العامل الثاني) احسب :

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

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 للعاملين نفس الحالة الاجتماعية؟

٤. احتمال أن يكون العاملان من القسم نفسه؟

الحل:

١- 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 مستقلان (لأن السحب مع الإرجاع) فإن:

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

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

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

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

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

٤- 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:

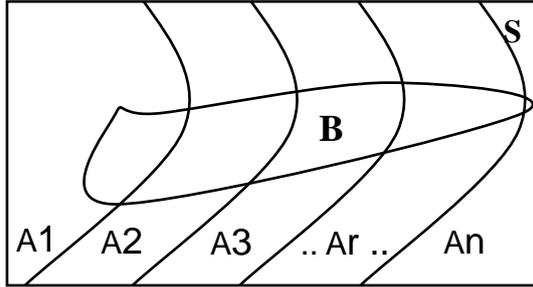
$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &= P(A_1A_2) + P(B_1B_2) + P(C_1C_2) \\ &= P(A_1 \times A_2) + P(B_1 \times B_2) + P(C_1 \times C_2) \\ &= \left[\frac{12}{50} \times \frac{12}{50} \right] + \left[\frac{22}{50} \times \frac{22}{50} \right] + \left[\frac{16}{50} \times \frac{16}{50}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

نظرية بايز (Bayes' Theorem)

إذا كانت A_1, A_2, \dots, A_n مجموعة أحداث متنافية وكانت احتمالات حدوثها

$P(A_1), P(A_2), \dots, P(A_n)$ وإذا كان هناك حدث B يحدث إذا حدث أي من الأحداث المتنافية أنظر الشكل بالأسفل ، فإن احتمال حدوث الحدث A_r بشرط حدوث B هو :

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



مثال:-

مصنع يقوم بإنتاج سلعة معينة به ثلاث آلات، تنتج الآلة الأولى ٢٠% من إجمالي إنتاج السلعة وتنتج الآلة الثانية نسبة ٣٥% والثالثة بنسبة ٤٥% ، فإذا كانت نسبة الإنتاج المعيب في الثلاث آلات على الترتيب هو ٢% و ٢.٥% و ٣% ، 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

$$P(A_1)=0.2 \quad \{إنتاج الآلة الأولى\}=A_1$$

$$P(A_2)=0.35 \quad \{إنتاج الآلة الثانية\}=A_2$$

$$P(A_3)=0.45 \quad \{إنتاج الآلة الثالثة\}=A_3$$

$$B = \{إنتاج سلعة معينة\}$$

فيكون بالتالي:

$$P(B|A_1)=0.02$$

$$P(B|A_2)=0.025$$

$$P(B | A_3) = 0.03$$

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

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{\sum_{i=1}^n P(A_i)P(B|A_i)} = \frac{0.2 \times 0.02}{(0.2 \times 0.02) + (0.35 \times 0.025) + (0.45 \times 0.03)} = 0.152$$

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

$$P(A_2|B) = \frac{P(A_2)P(B|A_2)}{\sum_{i=1}^n P(A_i)P(B|A_i)} = \frac{0.35 \times 0.025}{(0.2 \times 0.02) + (0.35 \times 0.025) + (0.45 \times 0.03)} = 0.333$$

مثال :-

مستشفى به أربعة أقسام، نسب عمال النظافة في هذه الأقسام هي ٣٠% ، ٤٠% ، ٢٠% ، ١٠% على التوالي، إذا كانت نسب العمال المدخنين بهذه الأقسام هي ١٥% ، ١٨% ، ١٢% ، ٩% على التوالي، اختير عامل عشوائياً 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:

$$P(A_1) = 0.3 \quad P(B | A_1) = 0.15 \quad \{A_1 = \text{أن يكون العامل من القسم الأول}\}$$

$$P(A_2) = 0.4 \quad P(B | A_2) = 0.18 \quad \{A_2 = \text{أن يكون العامل من القسم الثاني}\}$$

$$P(A_3) = 0.2 \quad P(B | A_3) = 0.12 \quad \{A_3 = \text{أن يكون العامل من القسم الثالث}\}$$

$$P(A_4) = 0.1 \quad P(B | A_4) = 0.09 \quad \{A_4 = \text{أن يكون العامل من القسم الرابع}\}$$

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

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{\sum_{i=1}^4 P(A_i)P(B|A_i)} = \frac{0.3 \times 0.15}{(0.3 \times 0.15) + (0.4 \times 0.18) + (0.2 \times 0.12) + (0.1 \times 0.09)} = 0.3$$

واحتمال أن يكون العامل من القسم الثاني إذا علم -بشروط - أنه مدخن :

$$P(A_2|B) = \frac{P(A_2)P(B|A_2)}{\sum_{i=1}^4 P(A_i)P(B|A_i)} = \frac{0.4 \times 0.18}{(0.3 \times 0.15) + (0.4 \times 0.18) + (0.2 \times 0.12) + (0.1 \times 0.09)} = 0.48$$

واحتمال أن لا يكون العامل من القسم الأول إذا علم -بشروط - أنه مدخن :

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

تمارين واجب :-

١- عرف المصطلحات التالية :-

(التجربة العشوائية - فراغ العينة - الحادث - الحوادث المتنافية - الحوادث المستقلة - الحوادث الشاملة) .

٢- الجدول التالي يمثل توزيع موظفي أحد الشركات حسب الحالة الاجتماعية للموظف والمستوى الإداري الذي يعمل به

الحالة الاجتماعية	أعزب	متزوج	المجموع
مستوى الإدارة الدنيا	10	14	24
مستوى الإدارة المتوسطة	16	28	44
مستوى الإدارة العليا	20	12	32
المجموع	46	54	100

أولاً :- اختر موظف من الجدول السابق بطريقة عشوائية، احسب الاحتمالات التالية:

- أن يكون أعزباً.
- أن يكون متزوجاً .
- أن يكون من مستوى الإدارة الدنيا.
- أن يكون من مستوى الإدارة الدنيا أو المتوسطة .
- أن يكون من مستوى الإدارة الدنيا وأعزب .

ثانياً : اختر موظف من الجدول السابق بطريقة عشوائية، احسب الاحتمالات التالية:

- احسب احتمال أن يكون موظفي الإدارة الدنيا بشرط أنه متزوج؟
- احتمال أن يكون الموظف أعزب بشرط أنه من موظفي الإدارة العليا ؟

ثالثاً : تم اختيار ٢ موظف من الجدول السابق بطريقة عشوائية، احسب الاحتمالات التالية:

- احتمال أن يكون الموظفين من الادارة الدنيا ؟
- احتمال أن يكون الموظفين متزوجان؟
- احتمال أن يكون للموظفين نفس الحالة الاجتماعية؟
- احتمال أن يكون الموظفين من القسم نفسه؟

٣- مصنع يقوم بإنتاج سلعة معينة به ثلاث آلات، تنتج الآلة الأولى ٤٠% من إجمالي إنتاج السلعة وتنتج الآلة الثانية نسبة ٢٥% والباقي من إنتاج الآلة الثالثة، فإذا كانت نسبة الإنتاج المعيب في الثلاث آلات على الترتيب هو ٤% و ٣% و ٥.٥% ، سحبت وحدة عشوائية من إنتاج المصنع فوجد أنها جيدة ، احسب الاحتمالات التالية:

- أن تكون القطعة الجيدة من إنتاج الآلة الأولى؟
- أن تكون القطعة الجيدة من إنتاج الآلة الثانية؟

المحاضرة (٣)

المتغيرات العشوائية

المتغير العشوائي Random Variable:

المتغير العشوائي هو الذي يأخذ قيمة 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:

- المتغيرات العشوائية المنفصلة Discrete Random Variables
- المتغيرات العشوائية المتصلة (المستمرة) Continuous Random Variables

أولاً : 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، 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، $Y: \{y=0,1,2,3,\dots\}$.
- 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.
- 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 التكرار النسبي للقيم التي يمكن أن يأخذها المتغير.

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

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	$f(x_n)$
Σ	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 .

والمطلوب:

- كون فراغ العينة.

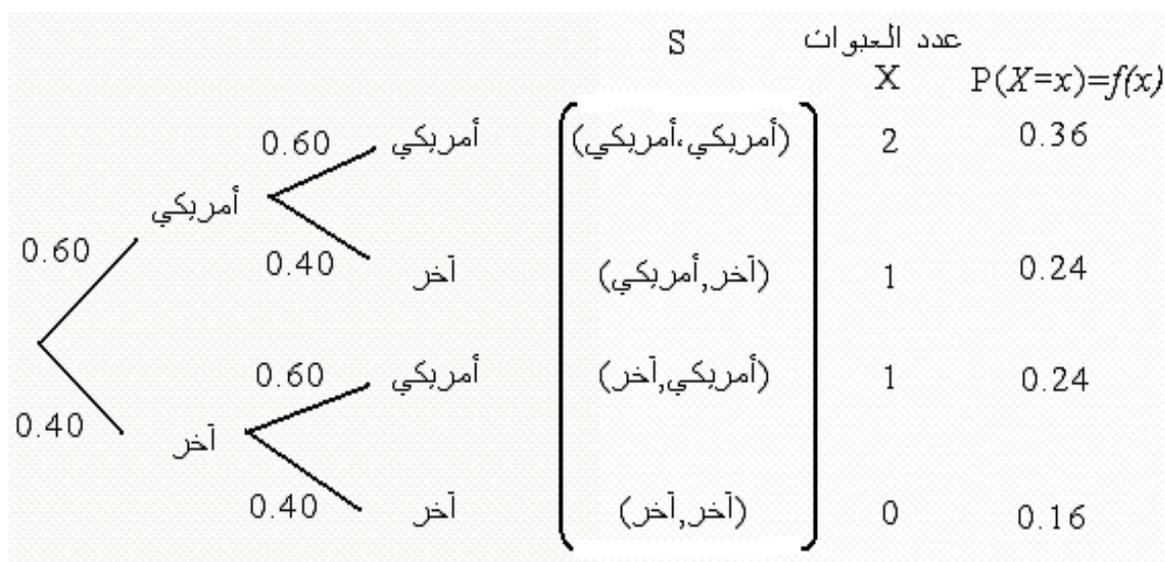
إذا عرف المتغير العشوائي بأنه عدد العبوات المشتراة من التفاح الأمريكي، فأوجد الآتي:

- 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



تابع الحل

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، أي 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)

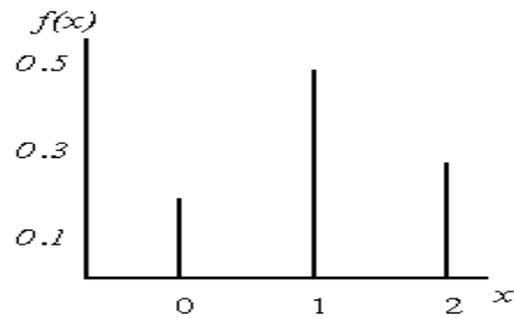
$x=2$ إذا كان العبوتين من النوع الأمريكي، أي إذا كانت نتيجة التجربة (أمريكي، أمريكي)

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:

جدول التوزيع الاحتمالي لعدد العبوات المشتراة من التفاح الأمريكي

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

رسم دالة الاحتمال $f(x)$:



الوسط الحسابي والتباين للمتغير العشوائي المنفصل:

يرمز للوسط الحسابي للمتغير العشوائي بالرمز (ميو)، ويحسب بتطبيق المعادلة التالية:

$$\mu = \sum x_i f(x_i) \quad (٣-٨)$$

وأما التباين ويرمز له بالرمز σ^2 (سيجما^٢)، فيحسب بتطبيق المعادلة التالية:

$$\begin{aligned}\sigma^2 &= \sum (x_i - \mu)^2 f(x_i) \\ &= \sum x_i^2 f(x_i) - \mu^2\end{aligned}\quad (٤-٨)$$

مثال:

في المثال السابق احسب ما يلي:

- الوسط الحسابي لعدد العبوات المشتراة من النوع الأمريكي.
- احسب الانحراف المعياري لعدد العبوات المشتراة من النوع الأمريكي.
- أوجد معامل الاختلاف النسبي .

الحل:

- الوسط الحسابي لعدد العبوات من النوع الأمريكي:
- لحساب الوسط الحسابي والانحراف المعياري يتم استخدام المعادلة الخاصة بذلك وهذا يتطلب تكوين جدول يشمل المجاميع التالية: $\sum x_i^2 f(x_i)$, $\sum x_i f(x_i)$ 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: $\mu = \sum x_i f(x_i) = 1.20$

ولحساب الانحراف المعياري يجب أولاً 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: $\sigma = \sqrt{\sigma^2} = \sqrt{0.48} = 0.693$

معامل الاختلاف النسبي هو:

$$C.V = \frac{\sigma}{\mu} \times 100 = \frac{0.693}{1.2} \times 100 = 57.7$$

ثانياً : المتغيرات العشوائية المستمرة Continuous Random Variables

المتغير العشوائي المستمر هو الذي يأخذ قيما متصلة، ويأخذ عدد لانهايني من القيم الممكنة له داخل مجاله، فإذا كان x متغير عشوائي مستمر، ويقع في المدى (a,b) أي $\{X = x: a < x < b\}$ فإن للمتغير X عدد لانهايني 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 المزروعة بالأعلاف في المملكة بالآلاف هكتار: $\{X = x: 1000 < x < 15000\}$
- فترة صلاحية حفظ الدجاج المبرد بالأيام: $\{X = x: 1 < x < 5\}$
- وزن الجسم بالكيلوجرام للأعمار من $(30-40)$: $\{X = x: 55 < x < 80\}$
- وهكذا الأمثلة على المتغير الكمي المستمر كثيرة.

الوسط الحسابي والتباين للمتغير العشوائي المستمر:

إذا كانت $f(x)$ هي دالة كثافة الاحتمال للمتغير العشوائي $a < x < b$ فإن معادلة الوسط والتباين يمكن كتابتها كما يلي:

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

(٨-١٣)

التوزيعات الاحتمالية للمتغيرات المنفصلة:-

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 مدي المتغير العشوائي X والذي يعبر عن عدد حالات النجاح هو: $X : \{x = 0, 1, 2, \dots, n\}$

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 في كل محاولة (النجاح) P ثابت ولا يتغير من محاولة لأخرى .

فبالتالي يمكن حساب الاحتمال من خلال المعادلة التالية :

$$P(X) = \frac{n!}{X!(n-X)!} P^X (1-P)^{n-X}$$

حيث $n! = 1, n, (n-1), (n-2), \dots, 3, 2, 1$ (" مضروب n ") $0! = 1$.

ويكون متوسط توزيع ذي الحدين $\mu = np$

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

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.

مثال:

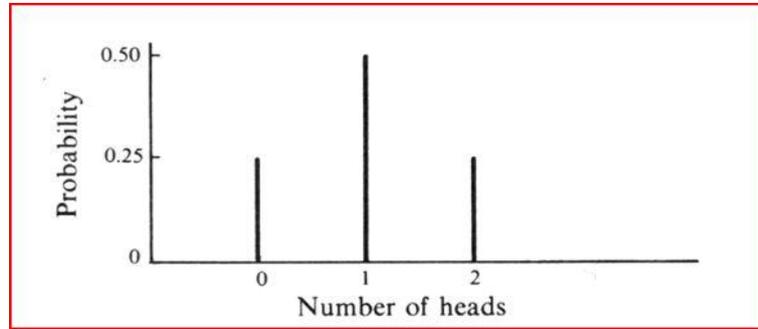
عند رمي عملة متوازنة مرتين فإن النواتج الممكنة هي HH, HT, TH, TT وعلى ذلك فإن :

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

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

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

ويمكن كذلك تمثيل ذلك من خلال الرسم التالي:



مثال :-

إذا كان احتمال نجاح الطالب في مقرر التحليل الإحصائي ٨٠% تم اختيار ٤ طلاب المطلوب :-

١. كون جدول توزيع ثنائي الحدين .
٢. أوجد احتمال نجاح ٣ طلاب .
٣. أوجد احتمال رسوب ٣ طلاب .
٤. أوجد احتمال نجاح طالبين على الأقل .
٥. القيمة المتوقعة (الوسط الحسابي) .
٦. الانحراف المعياري .

الحل :-

$$P = 0.80 , (1-P= 0.20) , n=4$$

١- جدول توزيع ثنائي الحدين :-

الناتج	الاحتمال	عدد الطلاب الراسبين	عدد الطلاب الناجحين
0.0016	$= 4C0 \times (0.80)^0 \times (0.20)^4$	4	0
0.0256	$= 4C1 \times (0.80)^1 \times (0.20)^3$	3	1
0.1536	$= 4C2 \times (0.80)^2 \times (0.20)^2$	2	2
0.4096	$= 4C3 \times (0.80)^3 \times (0.20)^1$	1	3
0.4096	$= 4C4 \times (0.80)^4 \times (0.20)^0$	0	4

٢- أجد إحتمال نجاح ٣ طلاب :-

$$P(3) = 0.4096$$

٣- أوجد إحتمال رسوب ٣ طلاب :-

$$P(1) = 0.0256$$

٤- أوجد إحتمال نجاح طالبين على الأقل :-

$$P(2)+P(3)+P(4) = 0.9728$$

٥- القيمة المتوقعة (الوسط الحسابي) =

$$\mu = n \times p = 4 \times 0.80 = 3.2$$

٦- الانحراف المعياري =

مثال :-

إذا كان إحتمال حياة شخص عند العمر ٣٠ هو ٦٠% تم إختيار ٥ أشخاص عند تمام العمر ٣٠ المطلوب :-

١. كون جدول توزيع ثنائي الحدين .

٢. أوجد إحتمال حياة ٤ أشخاص .

٣. أوجد احتمال وفاة ٣ أشخاص .

٤. أوجد احتمال حياة ٣ أشخاص على الأقل .

٥. القيمة المتوقعة (الوسط الحسابي) .

٦. الانحراف المعياري .

الحل :-

$$P = 0.60 , (1-P= 0.40) , n=5$$

١- جدول توزيع ثنائي الحدين :-

عدد الاحياء	عدد الوفيات	الاحتمال	الناتج
0	5	$= 5C0 \times (0.60)^0 \times (0.40)^5$	0.01024
1	4	$= 5C1 \times (0.60)^1 \times (0.40)^4$	0.0768
2	3	$= 5C2 \times (0.60)^2 \times (0.40)^3$	0.2304
3	2	$= 5C3 \times (0.60)^3 \times (0.40)^2$	0.3456
4	1	$= 5C4 \times (0.60)^4 \times (0.40)^1$	0.2592
5	0	$= 5C4 \times (0.60)^5 \times (0.40)^0$	0.07776

٢- أوجد احتمال حياة ٤ أشخاص :-

$$P(4) = 0.2592$$

٣- أوجد احتمال وفاة ٣ أشخاص :-

$$P(2) = 0.2304$$

٤- أوجد احتمال حياة ٣ أشخاص على الأقل :-

$$P= (p(3) + p(4) + p(5)) = 0.07776+0.2592+0.3456 = 0.68256$$

٥- القيمة المتوقعة (الوسط الحسابي) :-

$$\mu = n \times p = 5 \times 0.60 = 3$$

٦- الانحراف المعياري =

$\sigma =$

المحاضرة (٤)

تابع التوزيعات الإحتمالية

تابع توزيع ثنائي الحدين :-

مثال :-

إذا كان احتمال إصابة الهدف لشخص ما هو $\frac{1}{5}$ أتاحت له فرصة الرماية في 10 محاولات

1- ما هو احتمال إصابة الهدف مرتين على الأكثر

2- احتمال إصابة الهدف مرة واحدة

الحل

X متغير عشوائي يمثل عدد مرات النجاح في إصابة الهدف في 10 محاولات

$$n = 10, p = 1/5, q = 4/5 ; \quad x = 0, 1, 2, \dots, 10$$

1- احتمال إصابة الهدف مرتين على الأكثر:-

أي احتمال $x = 0$ or $x = 1$ or $x = 2$

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

$$= \binom{10}{0} (1/5)^0 (4/5)^{10} + \binom{10}{1} (1/5)^1 (4/5)^9$$

$$+ \binom{10}{2} (1/5)^2 (4/5)^8$$

$$= (0.8)^{10} + 2(0.8)^9 + 1.8(0.8)^8 = 0.6778$$

2- احتمال إصابة الهدف مرة واحدة

أي احتمال $x = 1$

$$P(X = 1) = \binom{10}{1} (1/5)^1 (4/5)^9$$

مثال :-

ألقيت عملة ثلاث مرات. فإذا كان X يمثل عدد ظهور الصور فأوجد التوزيع الاحتمالي وكذلك التوقع والتباين

الحل

احتمال النجاح (ظهور صورة) $p = 0.5$

احتمال الفشل (ظهور كتابة) $q = 0.5$

عدد الرميات المستقلة $n = 3$

X متغير عشوائي يمثل عدد الصور يأخذ القيم 0, 1, 2, 3

ويكون له توزيع ذي الحدين:

$$p(X = x) = \binom{3}{x} (0.5)^x (0.5)^{3-x}, \quad x = 0, 1, 2, 3$$

$$p(X = 0) = \binom{3}{0} (0.5)^0 (0.5)^{3-0} = \frac{3!}{0!(3-0)!} (1)(1/8) \\ = \frac{1}{8}$$

$$p(X = 1) = \binom{3}{1} (0.5)^1 (0.5)^{3-1} = \frac{3!}{1!(3-1)!} (1/2)(1/2)^2 \\ = \frac{3 \times 2 \times 1}{1 \times 2 \times 1} \times \frac{1}{2} \times \frac{1}{4} = \frac{3}{8}$$

$$p(X = 2) = \binom{3}{2} (0.5)^2 (0.5)^{3-2} = \frac{3!}{2!(3-2)!} (1/2)^2 (1/2)^1 \\ = \frac{3 \times 2 \times 1}{2 \times 1 \times 1} \times \frac{1}{2} \times \frac{1}{4} = \frac{3}{8}$$

$$p(X = 3) = \binom{3}{3} (0.5)^3 (0.5)^{3-3} = \frac{3!}{3!(3-3)!} (1/2)^3 (1/2)^0 \\ = \frac{3 \times 2 \times 1}{3 \times 2 \times 1 \times 0!} \times \frac{1}{8} \times 1 = \frac{1}{8}$$

$$E(X) = np = 3(0.5) = 1.5 \quad \text{التوقع}$$

$$\text{Var}(X) = npq = 3(0.5)(0.5) = 0.75 \quad \text{التباين}$$

مثال:-

وجد في إنتاج أحد المصانع أنه من بين 1000 وحدة إنتاج يوجد 150 وحدة معيبة. أخذت عينة بإرجاع مليونية من 5 وحدات، أوجد الاحتمالات التالية:

١- الوحدات المختارة كلها سليمة

٢- على الأكثر توجد واحدة معيبة

٣- على الأقل توجد وحدتان معيبتان

٤- القيمة المتوقعة و التباين للوحدات المعيبة .

الحل

احتمال النجاح (الحصول على وحدة معيبة) $p = 150/1000 = 0.15$

احتمال الفشل (عدم الحصول على وحدة معيبة) $q = 1-p = 1-0.15 = 0.85$

عدد المحاولات (عينة بإرجاع مكونة من 5 وحدات) $n = 5$

X متغير عشوائي يمثل عدد الوحدات المعيبة يأخذ القيم 0, 1, 2, 3, 4, 5

ويكون له توزيع ذي الحدين:

$$p(X=x) = \binom{5}{x} (0.15)^x (0.85)^{5-x}, \quad x=0,1,2,3,4,5$$

١- الوحدات كلها سليمة يعني أن $X = 0$

$$p(X=0) = \binom{5}{0} (0.15)^0 (0.85)^{5-0} = \frac{5!}{0!(5-0)!} (1)(0.85)^5 = 0.4437$$

٢- على الأكثر توجد وحدة معيبة يعني أن $X \leq 1$

$$P(X \leq 1) = p(X=0) + p(X=1)$$

$$\begin{aligned} p(X \leq 1) &= \binom{5}{0} (0.15)^0 (0.85)^5 + \binom{5}{1} (0.15)^1 (0.85)^4 \\ &= 0.4437 + \frac{5!}{1!5!} (0.15)(0.522) \\ &= 0.4437 + 5 \times 0.0783 = 0.4437 + 0.3915 = 0.8352 \end{aligned}$$

$$p(X=x) = \binom{5}{x} (0.15)^x (0.85)^{5-x}, \quad x=0,1,2,3,4,5$$

٣- على الأقل توجد وحدتان معيبتان يعني أن $X \geq 2$

$$P(X \geq 2) = 1 - p(X < 2)$$

$$= 1 - [p(X=0) + p(X=1)]$$

$$= 1 - 0.8325 = 0.1648$$

٤- القيمة المتوقعة و التباين للوحدات المعيبة .

$$0.75 = 5 \times 0.15 = n \cdot p$$

$$= n \times p \times (1 - p) = \text{التباين}$$

$$0.6375 = 5 \times 0.15 \times 0.85 =$$

ب - توزيع بواسون :-

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

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

$$, x = 0,1,2,\dots$$

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: $e = 2.718$ تقريباً، ويمكن حساب قيمتها باستخدام الآلة الحاسبة.

$x!$ = مضروب العدد x " ويساوي: $x! = x(x-1)(x-2)\dots 3 \times 2 \times 1$

μ = المتوسط

• يشترك توزيع بواسون من توزيع ذي الحدين عندما يكون :-

- عدد المحاولات n كبير جداً
- بينما يكون احتمال النجاح p صغير بحيث تبقى np قيمة ثابتة معتدلة
- يوصف متغيرات عشوائية متقطعة تعبر عن عدد كبير من الحوادث مثل:
 - عدد حوادث السيارات في الشهر داخل مدينة كبيرة
 - عدد الكرات الحمراء في عينة الدم
 - عدد الأخطاء المطبعية في الصفحات المختلفة للكتاب
 - عدد القطع التالفة في الإنتاج الكلي لسلسلة معينة
 - توزيع بواسون فإن X إذا كان للمتغير

التوقع $E(X) = \lambda$

التباين $Var(X) = \lambda$

مثال :-

في كمية كبيرة من القطع المصنعة، وكان معلوماً أن بها نسبة 0.3% من القطع المعيبة. أخذت منه عينة بإرجاع عشوائية حجمها 350 قطعة. احسب الاحتمالات الآتية:

(١) وجود قطعة معيبة

(٢) وجود قطعتان معيبتان

٣) عدم وجود أية قطع معيبة

٤) وجود على الأكثر وحدتان معيبتان

الحل

عملية سحب العينة تمثل سلسلة عددها $n=350$

واحتمال أن تكون القطعة معيبة (النجاح) $p=0.003$

واضح n كبيرة و p صغيرة

$$\lambda=np = 350(0.003) = 1.05$$

بفرض أن X يمثل عدد القطع المعيبة في العينة له توزيع بواسون

$$p(X = x) = e^{-\lambda} \frac{\lambda^x}{x!} = e^{-1.05} \frac{1.05^x}{x!}$$

$x = 0, 1, 2, \dots$

$$p(X = 1) = e^{-1.05} \frac{1.05^1}{1!} = (0.3499)(1.05) = 0.367 \quad \text{١) وجود قطعة معيبة في العينة}$$

$$p(X = 2) = e^{-1.05} \frac{1.05^2}{2!} = (0.3499)(0.55125) = 0.193 \quad \text{٢) وجود قطعتان معيبتان في العينة}$$

$$p(X = 0) = e^{-1.05} \frac{1.05^0}{0!} = 0.350 \quad \text{٣) عدم وجود أي قطع معيبة في العينة}$$

$$\text{٤) وجود على الأكثر وحدتان معيبتان} \quad P(X \leq 2) = p(X=0) + p(X=1) + p(X=2)$$

$$= 0.350 + 0.367 + 0.193$$

$$= 0.91$$

مثال :-

إذا كان عدد الأخطاء المطبعية في كتاب يتكون من 600 صفحة هو 50 خطأ فإذا كانت الأخطاء تتوزع توزيعاً عشوائياً، فما احتمال إذا اختيرت 10 صفحات عشوائياً أن لا تحتوي على أخطاء.

الحل

بفرض أن X يمثل عدد الأخطاء في كل صفحة وأن عدد المحاولات (الصفحات) تمثل سلسلة من محاولات

برنولي عددها $n = 10$

ونسبة الخطأ (النجاح) هي

$$p = \frac{50}{600} = 0.083$$

وعليه فإن :- $\lambda=np = 10(0.083) = 0.83$

وبالتالي فإن لـ X توزيع بواسون:

$$p(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} = e^{-0.83} \frac{0.83^k}{k!}$$

$$k = 0, 1, 2, 3, \dots$$

احتمال أن لا يوجد أخطاء يساوي

$$P(X=0) = \frac{e^{-0.83} 0.83^0}{0!} = 0.436$$

مثال:-

إذا كان من المعلوم أن عدد الوحدات التي تستهلكها الأسرة من سلعة معينة خلال الشهر تتبع توزيع بواسون بمتوسط 3 وحدات شهرياً، 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:

$$X : \{x = 0, 1, 2, 3, \dots\}$$

شكل دالة الاحتمال:

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

$$P(x) = \frac{e^{-\mu} \mu^x}{x!} = \frac{e^{-3} 3^x}{x!}, \quad x = 0, 1, 2, \dots$$

حساب الاحتمالات:

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

$$P(2) = \frac{e^{-3} 3^2}{2!} = \frac{0.0498(9)}{2 \times 1} = 0.22404$$

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

$$\begin{aligned} P(X \leq 3) &= p(3) + p(2) + p(1) + p(0) \\ &= \left[\frac{3^3}{3!} + \frac{3^2}{2!} + \frac{3^1}{1!} + \frac{3^0}{0!} \right] \left[\frac{0.0498}{1} \right] \\ &= [0.0498] \left(\frac{27}{6} + \frac{9}{2} + \frac{3}{1} + \frac{1}{1} \right) = 0.0498(13) = 0.6474 \end{aligned}$$

حساب الوسط الحسابي، والانحراف المعياري لعدد حالات الاستجابة:

الوسط الحسابي (μ) في حالة التوزيع بواسون 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 ، وهو:

$$C.V = \frac{\sigma}{\mu} \times 100 = \frac{1.732}{3} \times 100 = 57.7\%$$

تحديد شكل التوزيع:

دائماً توزيع بواسون موجب الالتواء

مثال:-

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

$$\begin{aligned} P(x) &= \frac{e^{-\mu} \mu^x}{x!} \\ &= \frac{e^{-5} 5^x}{x!} = , \quad x = 0,1,2,\dots \\ &= \frac{(25)(0.00674)}{(2)(1)} = 0.08425 \end{aligned}$$

التوزيع الإحصائي :-

و هو الشكل الذي تأخذه مجموعة البيانات، وشكل البيانات مهم جداً في تحليلها ووصفها وكخطوة تسبق قرار استخدام أي أسلوب إحصائي .

ويرتبط التوزيع الاحصائي عادة بنوعين من البيانات المتصلة والمنفصلة، ويناسب النوع المنفصل المقاييس الاسمية والرتبية ، وهناك بعض المقاييس المنفصلة ثنائية أي انه لا يوجد بها الا قيمتين، وهي لا تسمى 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 الاحتمالية 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 بشيء 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 أي فترة يمثلها مساحة التوزيع الاحتمالي (ويسمى أيضاً دالة الكثافة) داخل هذه الفترة، والمساحة الكلية تحت المنحنى (الاحتمال) تساوي 1

التوزيع الطبيعي

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

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

خصائص التوزيع الطبيعي:

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

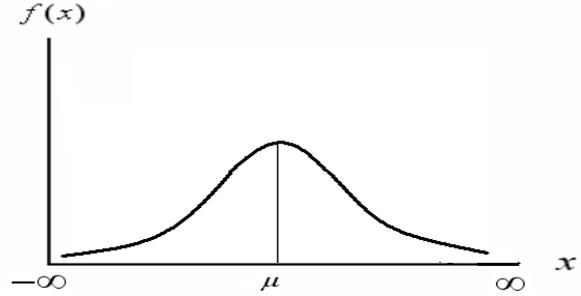
- توزيع جرسى أي يشبه الجرس.
- توزيع متصل
- توزيع متماثل حول الوسط
- الالتواء (الاطراف) والتفلطح (القمة) يساوي صفر.
- يحوي منوال ووسط ووسيط واحد وذات قيم متساوية بمعنى أن الجزء الذي على يمين الوسط مطابق للجزء الايسر
- الذيلين الايمن والايسر يقتربان من الخط الافقي ولكن لا تلامسه
- المساحة الكلية تحت المنحنى 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 أن لدينا جرس طويل مدبب، والقيمة الكبيرة لـ σ تعني أن الجرس قصير ومفرطح.

والشكل التالي يوضح ذلك:



والتوزيع الطبيعي وتطبيقاته الاحصائية ليس موضوعاً جديداً بل عرف منذ القرن السابع عشر الميلادي ومن ابرز الدراسات المعروفة تلك الدراسة البريطانية التي اخذت اطوال ٨٥٨٥ 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 : $\text{var}(x) = \sigma^2$

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 .

التوزيع الطبيعي القياسي (المعياري) :-

- 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:

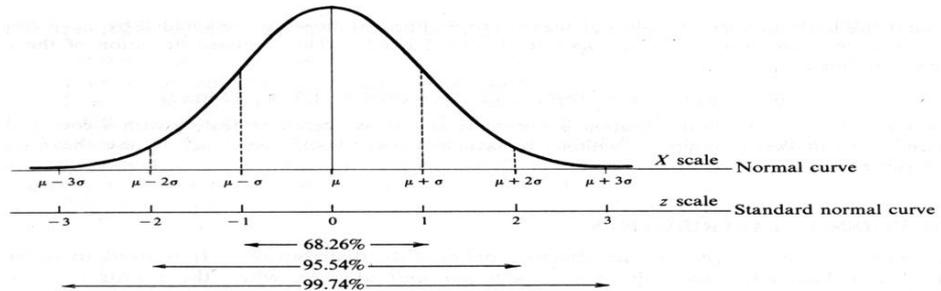


Fig. 3-4

مثال :-

تم دراسة متوسط طول الطالب في كلية إدارة الأعمال هو ١٨٠ سم و ذلك بانحراف معياري ١٠ سم تم اختيار أحد الطالب عشوائياً فإذا علمت أن هذه الظاهرة تتبع التوزيع الطبيعي فأوجد :-

١-احتمال أن ينحصر طول الطالب بين ١٧٠ سم و ١٩٠ سم $(p(170 < x < 190))$.

٢-احتمال أن ينحصر طول الطالب بين ١٦٠ سم و ٢٠٠ سم $(p(160 < x < 200))$.

٣-احتمال أن ينحصر طول الطالب بين ١٥٠ سم و ٢١٠ سم $(p(150 < x < 210))$.

٤- احتمال أن يكون طول الطالب أقل من ١٩٠ سم $(p(x < 190))$.

٥- احتمال أن يكون طول الطالب أكبر من ١٩٠ سم $(p(x > 190))$.

٦-احتمال أن يكون طول الطالب أكبر من ١٥٠ سم $(p(x > 150))$.

٧- احتمال أن يكون طول الطالب أقل من ١٦٠ سم $(p(x < 160))$.

الحل :-

١-احتمال أن ينحصر طول الطالب بين ١٧٠ سم و ١٩٠ سم $(p(170 < x < 190))$:-

$$-1 < z < 1$$

$$P = 68.26\%$$

٢-احتمال أن ينحصر طول الطالب بين ١٦٠ سم و ٢٠٠ سم $(p(160 < x < 200))$:-

$$\frac{160-180}{10} < z < \frac{200-180}{10} \quad -2 < z < 2 \quad P = 95.45\%$$

٣-احتمال أن ينحصر طول الطالب بين ١٥٠ سم و ٢١٠ سم $(p(150 < x < 210))$:-

$$-3 < z < 3$$

$$P = 99.74\%$$

٤- احتمال أن يكون طول الطالب أقل من ١٩٠ سم $(p(x < 190))$:-

$$z < 1$$

$$P = (0.6826/2) + 0.5 = 84.13\%$$

٥- احتمال أن يكون طول الطالب أكبر من ١٩٠ سم $(p(x > 190))$:-

$$z > 1$$

$$P = 0.5 - (0.6826/2) = 15.87\%$$

٦- احتمال أن يكون طول الطالب أكبر من ١٥٠ سم $(p(x > 150))$:-

$$z > -3$$

$$z < 3$$

$$P = (0.9974/2) + 0.5 = 99.87\%$$

٧- احتمال أن يكون طول الطالب أقل من ١٦٠ سم $(p(x < 160))$:-

$$z > -2$$

$$z > 2$$

$$P = 0.5 - (0.9545/2) = 0.02275 = 2.275 \%$$

Tables of the Normal Distribution



Probability Content from $-\infty$ 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

استخدامات التوزيع الطبيعي القياسي:-

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

مثال:

افترض أن إدارة المرور بالأحساء 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:

- نسبة السيارات التي تقل سرعتها عن 50 ميلا في الساعة .
- نسبة السيارات التي تزيد سرعتها عن 65 ميلا في الساعة.

- نسبة السيارات التي تكون سرعتها بين 60 ميلا و 70 ميلا في الساعة .
- عدد السيارات التي تكون سرعتها بين 60 ميلا و 77.45 ميلا من بين 10000 سيارة .

الحل :-

١- نسبة السيارات التي تقل سرعتها عن ٥٠ ميلا في الساعة :

$$P(X < 50) = P\left(\frac{x - \mu}{\sigma} < \frac{50 - 60}{\sqrt{25}}\right) = P(Z < -2) = 0.5 - (0.9545/2) = 0.02275$$

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

$$P(X > 65) = P\left(\frac{x - \mu}{\sigma} > \frac{65 - 60}{\sqrt{25}}\right) = P(Z > 1) = 1 - P(Z \leq 1) = 0.5 - (0.6826/2) = 0.1587$$

٣- نسبة السيارات التي تكون سرعتها بين 60 ميلا و 80 في الساعة :

$$\begin{aligned} P(60 \leq X \leq 77.45) &= P\left(\frac{60 - 60}{\sqrt{25}} \leq Z \leq \frac{77.45 - 60}{\sqrt{25}}\right) \\ &= P(0 \leq Z \leq 2) = P(Z \leq 2) - P(Z \leq 0) \\ &= (0.9545/2) = 0.4772 \end{aligned}$$

٤- عدد السيارات المتوقع سرعتها بين 60 ميلا و 80 ميلا من بين 10000 سيارة :

$$10000(0.47725) = 4772$$

ملاحظة مهمة :- الدكتور اعطا تنبيه انه هذه المحاضره مهمة جدا و من اهم المحاضرات ..

المحاضرة (٥)

توزيعات المعاينة

Sampling Distribution's

الاستدلال الإحصائي :-

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.

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

العينة العشوائية:-

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

فمثلاً نستعين بعينه مسحوبة من المجتمع لتقدير معالم هذا المجتمع مثل متوسطة أو تباينه أو غير ذلك أو إعطاء 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.

المجتمع Population

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

والمجتمع قد يكون مجموعة ما من البشر أو أشجار أنواع معينة من الفاكهة أو الحيوانات الزراعية أو إنتاج دولة ما لسلع معينة خلال فترة زمنية محددة... 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 (لانهاي) إذا كان لا يمكن حصر عدد أفراده مثل النجوم والكواكب أو الكائنات الحية بمياه المحيطات والأنهار

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

أولاً: أسلوب الحصر الشامل (census): وفيه تجمع البيانات عن كل مفردة من مفردات المجتمع، وهذا الأسلوب لا يتبع عادة إلا في حالة التعدادات التي تجريها الدول وتدعمها بإمكانيات ضخمة 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.

بعض مزايا أسلوب المعاينة:-

يتميز أسلوب المعاينة عن أسلوب الحصر الشامل بمزايا عديدة 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 قيمة محدودة بعد أن فقدت عنصر المطابقة مع واقع الظاهرة وتوزيعها الحالي في المجتمع.

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

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

أقسام العينات:-

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

١. العينات العشوائية:

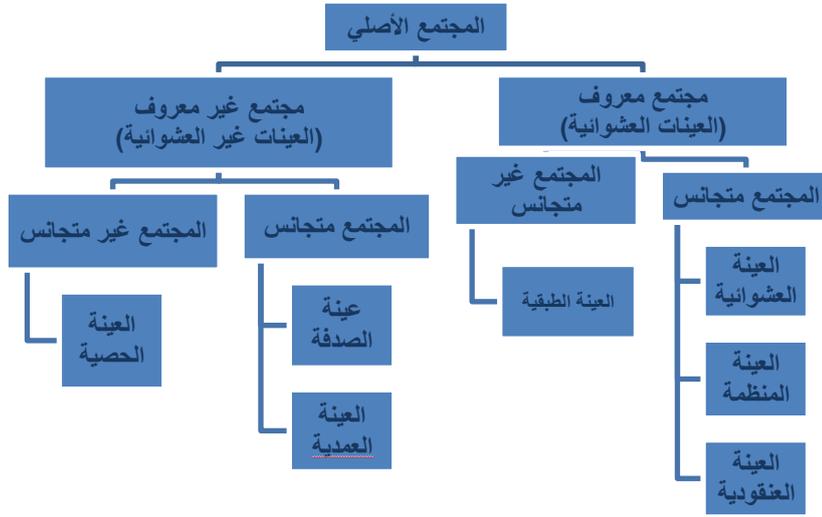
وهي تلك العينات التي يتم اختيار مفرداتها حسب خطه إحصائية لا يكون فيها للباحث أو لمفردات العينة دخل في اختيار أي مفردة فيها ، حيث يتم الاختيار باستخدام أساليب معينة تلعب الصدفة خلالها الدور الأول في اختيار المفردة ولكن بشرط أن يتحقق لجميع المفردات احتمال ثابت ومحدد للاختيار. والعينات العشوائية إذا ما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

٢. 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:

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

ب - العينات غير الاحتمالية:

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

أخطاء البيانات الإحصائية: -

تتعرض البيانات الإحصائية التي يتم جمعها إلى نوعين من الأخطاء:

١. **خطأ التمييز أو التحيز:** وهو ذلك الخطأ الناتج عن مصادر متعددة، منها أخطاء في تصميم البحث أو التجربة أو أخطاء فنية أثناء جمع البيانات أو خلال العمليات الحسابية التي تتم على البيانات المتجمعة. أخطاء التمييز تزداد بازدياد الفروق بين الإمكانات (المادية والفنية) اللازم توافرها لضمان أقصى درجة دقة ممكنة وبين الإمكانات الفعلية المتاحة للباحث.

٢. **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:

١ - خطأ التمييز أو التحيز:

إذا 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 التمييز أو التحيز**

أسباب خطأ التمييز أو التحيز:

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

٢ - 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 العشوائية**

كيف نقلل من خطأ المعاينة العشوائية:

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

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 نرسم لكل منها برموز تختلف عن رموز الأخرى، على سبيل المثال يرمز للمتوسط الحسابي للمجتمع بالرمز \bar{x} بينما يرمز للمتوسط الحسابي للعينة بالرمز \bar{x}_n ، أيضاً للانحراف المعياري للمجتمع بالرمز S بينما يرمز للانحراف المعياري للعينة بالرمز s وهكذا

S

الخطأ المعياري:

بمعرفة قيمة الانحراف المعياري لقيم العينة يمكن تقدير قيمة الخطأ المعياري في الانحراف المعياري للعينة باعتماد المعادلة الآتية:-

$$SE = \left[\frac{S}{\sqrt{n}} \right] \left[\sqrt{1 - \left(\frac{n}{N} \right)} \right]$$

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 مجهولاً، 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 (١٠٠) فتعتمد المعادلة أدناه:-

$$SE = \frac{S}{\sqrt{n^2}}$$

الخطأ المعياري = الانحراف المعياري لقيم العينة / جذر (حجم العينة)^٢

إن الانحراف المعياري للتوزيع النظري لمتوسطات العينات يقيس خطأ المعاينة ويسمى بالخطأ المعياري للمتوسط، ومن الضروري التذكّر دوماً 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 دائماً بصورة متماثلة **Normal Distribution**، والذي يمتاز رياضياً بالابتعاد بنسب ثابتة عن المتوسط مع كل درجة معيارية، وبالتالي تباينت متوسطات العينات المأخوذة منه فإنه يتوقع أن يقع متوسطه وباحتمالية قدرها كالتالي:
- مستوى ثقة إحصائية قدره (٦٨.٢٦%) أو باحتمالية قدرها (٠.٦٨٢٧) يقع متوسط مجتمع الدراسة بين قيمة متوسط متوسطات العينات و (+) و (-) درجة واحدة من الخطأ المعياري.
- مستوى ثقة إحصائية قدرها (٩٥%)، أو باحتمالية (٠.٩٥) يقع متوسط مجتمع الدراسة بين متوسط متوسطات العينات و (+) و (-) درجتان من الخطأ المعياري تقريبا.
- مستوى ثقة إحصائية قدرها (٩٩%) أو باحتمالية قدرها (٠.٩٩) يقع متوسط مجتمع الدراسة بين قيمة متوسط متوسطات العينات و (+) و (-) ثلاث درجات من قيمة الخطأ المعياري تقريبا.
- وتسمى هذه بمستويات الثقة Confidence Level ويعبر عنها بإشارة النسبة المئوية (%) بلن تكون التقديرات صحيحة أو باحتمالية (٠.٠١) أو (٠.٠٥) أن تكون خاطئة.

مثال :-

قام أحد الباحثين في مجال الزراعة بدراسة مائة مزرعة، فوجد أن متوسط مساحة المزرعة الواحدة (٥٣) هكتارا، وبانحراف معياري عن المتوسط بقيمة (٢٦) هكتارا

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

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 الجانبيين) وأصبحت عملية الاستدلال أكثر دقة. وللتوضيح نورد مثالا، إذا أريد معرفة نسبة طلبة قسم الإدارة إلى مجموع طلبة الكلية فإن عينة من عشرة طلبة قد لا تفي بالغرض، ولكن عينة من مائة طالب تفي بالغرض تماما. بعبارة أخرى، إن حجم العينة أساسي لإعطاء صورة عن مجتمع الدراسة وليس النسبة المنوية للعينة قياسا بحجم مجتمعها. فكلما ازداد حجم العينة ازدادت الثقة بتقديرات خصائص المجتمع وصغرت معه حدود الثقة.

(ب) توزيع (ت) للقيم :

من الضروري اخذ الحذر عندما يكون حجم العينة صغيرا اقل من (٣٠) وذلك لأنها تتطلب إجراءات خاصة عند التحليل. فعندما يكون حجم العينة أكثر من (٣٠) يتجه توزيع قيمها نحو التوزيع الطبيعي وبغض النظر عن التوزيع الحقيقي لقيم مجتمع الدراسة.

وبالنسبة للعينات الصغيرة الحجم فإن 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 (٣٠) فشكل توزيع (ت) للقيم لا يختلف كثيرا عن التوزيع الطبيعي إلا في الأعداد القليلة، وكلاهما متماثل النصفين لذا يعتمد كبدل له في القيم القليلة العدد.

ولتوزيع (ت) جداول للقيم الحرجة منظمة على شكل اسطر اعتمادا على درجة الحرية التي تقاس بـ(حجم العينة - ١). أما الأعمدة فتتمثل 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.

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 وبالدرجة 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 يعتمد البعض إلى تقليص حدود الثقة (المدى الذي يفترض أن يقع ضمنه المعدل المتوقع للمجتمع). إن إنقاص حدود الثقة من (٦%) إلى (٤%) يتطلب زيادة حجم العينة بنسبة (٢٥%)، 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

المحاضرة (٦)

التقدير الإحصائي

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)

١- التقدير بنقطة :-

التقدير بنقطة يعني 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 مرشحاً معيناً كتقدير لهذه النسبة في المجتمع نكون حصلنا على تقدير نقطة للنسبة في مجتمع الناخبين .

٢- التقدير بفترة:-

أما التقدير بفترة فنحصل من خلاله على مدى **Range** أو فترة تتحدد بحددين (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 أيضاً كسور السنوات، والأيام والشهور، والساعات.. الخ

تقدير الوسط الحسابي للمجتمع :-

أ - احسب الوسط الحسابي للعينة .

ب- احسب الخطأ المعياري للوسط والذي يساوي $\frac{\sigma}{\sqrt{n}}$

ج- أضرب الخطأ المعياري للوسط في معامل الثقة المناسب (أو الدرجة المعيارية) حسب درجة الثقة

المطلوبة أي احسب: $Z \frac{\sigma}{\sqrt{n}}$

د- 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.

$$\mu = \bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

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

مع ملاحظة أنه إذا كان الانحراف المعياري للمجتمع غير معروف - وهو غالباً ما يحدث في الواقع - فيمكن استخدام الانحراف المعياري للعينة S بدلاً منه طالما كان حجم العينة كبيراً بدرجة كافية وتصبح فترة **تقدير**

$$\hat{\mu} = \bar{X} \pm Z \frac{S}{\sqrt{n}} \quad \text{الوسط الحسابي للمجتمع كما يلي :}$$

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 أمراً صعباً من الناحية العملية نظراً لكبير 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 :-

$$\hat{\mu} = \bar{X} \pm Z \frac{S}{\sqrt{n}} \quad \text{بما أن فترة تقدير الوسط الحسابي للمجتمع هي :}$$

والمعلومات المعطاة هي :

حجم العينة n = 100

$$\bar{X} = 90$$

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 :

$$\begin{aligned}\hat{\mu} &= 90 \pm 1.96 \frac{25}{\sqrt{100}} \\ &= 90 \pm 1.96(2.5)\end{aligned}$$

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 معياري مقداره 60 وبالتالي فإن فترة تقدير الوسط الحسابي للمجتمع بدرجة ثقة 95 % هي : $\hat{\mu} = \bar{X} \pm 1.96S_{\bar{X}}$

$$\begin{aligned}&= \bar{X} \pm 1.96 \frac{S}{\sqrt{n}} \\ &= 100 \pm 1.96 \frac{60}{\sqrt{144}} \\ &= 100 \pm 1.96(5) \\ &= 100 \pm 9.8\end{aligned}$$

أي أن $\hat{\mu}$ تقع بين 90.2 , 109.8 بدرجة ثقة 95% . وكثيرا ما تستخدم أيضاً درجات الثقة 90 , 99% وهي 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 والاجتماعية... 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:

$$\hat{\mu} = \bar{X} \pm Z \frac{\sigma}{\sqrt{n}}$$

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 عليها من جدول التوزيع الطبيعي المعياري.

= 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 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 :

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

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 في تقدير متوسط عدد الدقائق التي 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 المعياري 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 : $n = \frac{Z^2 \cdot \sigma^2}{(e)^2}$

فإن حجم العينة مقرباً لأقرب عدد صحيح هو :

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

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

ومما سبق نستنتج أن :-

في 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 بحددين (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"

ولعل الاختلاف الأساسي بين توزيع t والتوزيع الطبيعي هو أن الانحراف المعياري للعينة هو المستخدم في الأول بدلاً 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 بـ درجات الحرية DEGREES OF FREEDOM .

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 لا بد وأن تكون (5) بالتالي نستطيع القول بأن درجة الحرية المتاحة لدى الباحث هي (2) أي $2 = 3 - 1$ أي أن درجات الحرية في هذه الحالة هي :

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 هي:

$$P\left(\bar{X} - t \frac{s}{\sqrt{n}} < \mu < \bar{X} + t \frac{s}{\sqrt{n}}\right) = 0.95 \quad (4.7)$$

حيث تشير t الى قيمة t التي تقع عندها 2.5% من المساحة الكلية للمنحنى عند كل طرف (عند درجات الحرية المستخدمة)، وتستخدم s/\sqrt{n} بدلا من σ/\sqrt{n}

شروط توزيع t :

ويمكن تحديد الشروط الثلاثة لاستخدام توزيع t كما يلي:

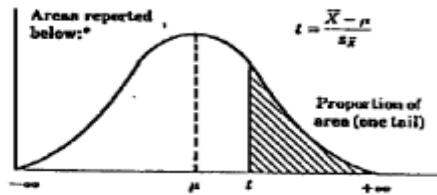
١. أن يكون المجتمع المسحوبة منه العينة له توزيع طبيعي
٢. والانحراف المعياري للمجتمع غير معروف (أو مجهول).
٣. والعينة صغيرة (حجمها أقل من 30 مفردة).

جدول توزيع t :

الجدول أدناه يعطي قيمة t

المقابلة للمساحة المظللة وقيمها ∞

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 one-tail	$t_{.50}$	$t_{.75}$	$t_{.90}$	$t_{.95}$	$t_{.99}$	$t_{.995}$	$t_{.9975}$	$t_{.999}$	$t_{.9995}$	$t_{.9999}$	$t_{.99995}$
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.025	0.01	0.005	0.001
df											
1	0.000	1.000	1.376	1.683	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.915	1.061	1.386	1.960	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	5.941	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	2.784	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	1.748	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	1.648	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%	99.8%	99.9%
		Confidence Level									

مثال :-

العينة أقل من ٣٠ و المجتمع يتبع التوزيع الطبيعي:

سحبت عينة عشوائية من $n=10$ بطارية فلاش متوسطها ٥ ساعات، والانحراف المعياري للعينة $s=1$ ساعة من خط إنتاج من المعروف أنه ينتج بطاريات عمرها موزع طبقاً للتوزيع الطبيعي.

المطلوب :

إيجاد فترة الـ 95% ثقة للمتوسط غير المعروف لعمر البطاريات في المجتمع كله.

الحل:

لإيجاد فترة الـ 95% ثقة للمتوسط غير المعروف لعمر البطاريات في المجتمع كله، فإننا نوجد أولاً قيمة t 0.025 و التي 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 التي سيتم التحصل عليها هي 2.262 إذن:

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

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

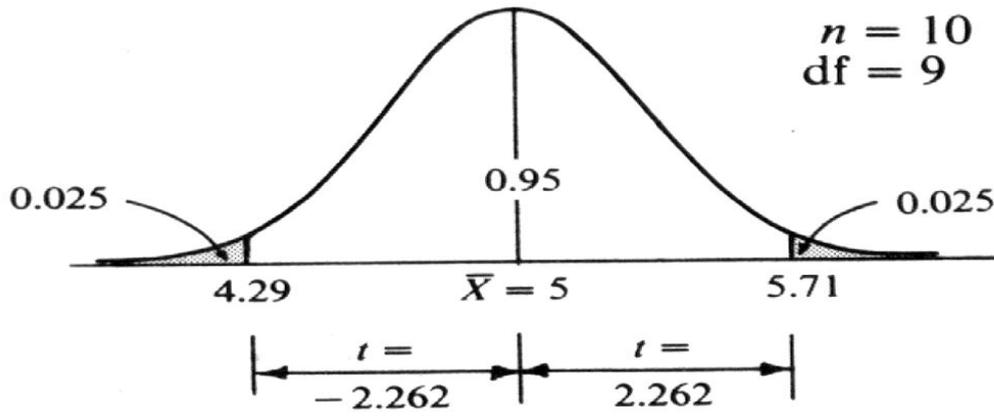


Fig. 4-4

تقدير فترة 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 غالباً 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} فتقدير النسبة في المجتمع تكون كما يلي:

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

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:

$$\hat{P} = \frac{60}{144} = 0.42$$

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 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 المناسب 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 $n = 144$

$$1 - \hat{P} = 1 - 0.42 = 0.58, \hat{P} = 0.42 \quad \text{والنسبة في العينة}$$

ومعامل الثقة $Z = 1.96$

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

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

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 (٧)

إختبار الفروض الإحصائية

اختبارات الفروض الإحصائية

Testing Statistical Hypotheses

المقصود بالفروض هنا الفروض الإحصائية **statistical hypotheses** بمعنى الفروض التي تتعلق بالمجتمع الإحصائي المسحوبة منه العينة، أو توزيع هذا المجتمع أو معالمه كالوسط الحسابي أو النسبة في المجتمع.

والفرض ما هو إلا تخمين أو استنتاج ذكي مبني على حيثيات معقولة أو منطقية ولكنه ليس مبنياً على حسابات دقيقة خاصة بالمجتمع لأنها نفترض أنه لا يمكن دراسة المجتمع بالكامل عن طريق الحصر الشامل بل نحاول استنتاج أو الاستدلال على مقاييس المجتمع باستخدام بيانات ونتائج العينة

فمثلاً : قد يفترض الباحث أن متوسط الدخل الشهري للفرد في دولة ما هو 200 ريال (بناءً على ما يراه من مستوى المعيشة في هذا البلد وأوضاعه الاقتصادية)، ويحتاج إلى اختبار علمي (إحصائي) لمعرفة مدى صحة هذا الفرض أو قد يفترض باحث آخر أن نسبة الناخبين في إحدى الدوائر الذين يؤيدون مرشحاً معيناً لا تقل عن 30% وهكذا... والمطلوب هو اختيار مدى صحة هذه الفروض. أي أن يصل الباحث إلى قرار إما بقبول الفرض أو عدم قبوله (أي رفضه) وذلك باحتمال معين. وقبل تناول كيفية إجراء الاختبارات الإحصائية نستعرض أولاً بعض المفاهيم والتعريفات الأساسية اللازمة لهذا الموضوع حتى تكون الصورة أكثر وضوحاً.

مفهوم الاختبارات الإحصائية :-

الفرض العدمي (أو الصفري) The Null Hypothesis

الفرض العدمي هو "الفرض الأساسي المراد اختباره". ويرمز له عادة بالرمز H_0 . هذا الفرض يأخذ - عادة - شكل معادلة أو مساواة. فمثلاً إذا كان الفرض العدمي المراد اختباره هو أن متوسط دخل الفرد في إحدى المناطق هو 200 ريال شهرياً فإن هذا الفرض يكتب بالرموز كما يلي : $H_0 : \mu = 200$

ويقرأ بالشكل التالي :

الفرض العدمي هو : أن متوسط دخل الفرد في المنطقة هو 200 ريال شهرياً.

وكمثال آخر : إذا كان الفرض المراد اختباره هو أن نسبة المؤيدين لبرنامج اقتصادي معين بين عمال أحد المصانع هي 30%، فإن هذا الفرض يكتب بالرموز كما يلي :

$$H_0 : P = 0.30$$

ويقرأ بالشكل التالي :

الفرض العدمي هو : أن نسبة المؤيدين للبرنامج الاقتصادي بين عمال المصنع هي 0.30

الفرض البديل : The Alternative Hypothesis

في اختبارات الفروض يتحتم وضع فرض آخر غير الفرض العدمي المراد اختباره يسمى الفرض البديل وهذا الفرض " هو الذي سيقبل في حالة رفض الفرض العدمي" أي لابد من تحدي فرض آخر بديل في الوقت الذي نحدد فيه الفرض العدمي، وبالتالي فإن الفرض البديل يعرف كما يلي :

"الفرض البديل هو الفرض الآخر الذي سيقبل في حالة رفض الفرض العدمي" ويرمز له عادة بالرمز : H1

والفرض البديل له أهمية كبيرة في قياس الظواهر الاجتماعية - كما سوف نرى - فهو الذي يحدد نوع الاختبار المستخدم لذلك فهو يأخذ أحد أشكال ثلاثة هي :-

أ- أن يأخذ شكل " لا يساوي " . وفي هذه الحالة نستخدم ما يسمى : اختبار الطرفين

فمثلاً : إذا كان الفرض العدمي هو أن متوسط الدخل الشهري لفئة معينة في المجتمع

هو 200 ريال . $H_0 : \mu = 200$

فإن الفرض البديل في هذه الحالة يأخذ الشكل التالي :

بمعنى أن متوسط دخل هذه الفئة من المجتمع " لا يساوي " 200 ريال شهرياً . $H_1 : \mu \neq 200$

بمعنى أن متوسط دخل هذه الفئة من المجتمع " لا يساوي " 200 ريال شهرياً .

ب- أو أن يأخذ شكل " أكبر من " . وفي هذه الحالة نستخدم ما يسمى " اختبار الطرف الأيمن " .

فمثلاً : قد يكون الفرض البديل كما يلي : $H_1 : \mu > 200$

أي أن متوسط الدخل لهذه الفئة من المجتمع أكبر من 200 ريال شهرياً .

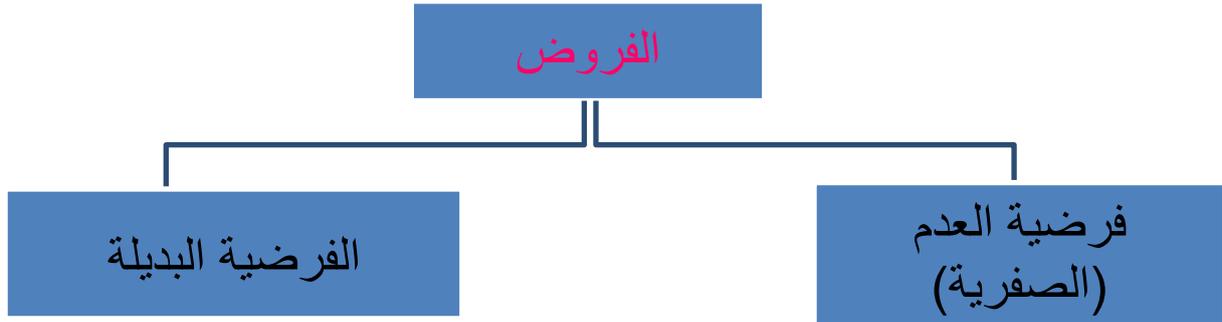
ج- وأخيراً قد يأخذ الفرض البديل شكل " أقل من " . وفي هذه الحالة نستخدم ما يسمى " اختبار الطرف الأيسر " .

فمثلاً : قد يكون الفرض البديل هو : $H_1 : \mu < 200$

أي أن متوسط الدخل لهذه الفئة من المجتمع أقل من 200 ريال شهرياً .

الخلاصة : الفروض الإحصائية :-

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



الخطأ في اتخاذ القرار :-

ففي حالة قبول الباحث لفرضه العدمي، فلا مجال للبحث في الفرض البديل، أما في حالة حدوث العكس بمعنى رفض الفرض العدمي فإنه يتحتم في هذه الحالة قبول الفرض البديل، على أنه من الجدير بالذكر أن الباحث هنا عرضة للوقوع في الخطأ عند اتخاذ قراره بقبول الفرض العدمي أو رفضه، فقد يرفض فرضاً هو في الواقع صحيح، وقد يقبل فرضاً هو في الواقع غير صحيح. لذلك فقد تم تصنيف هذه الأخطاء إلى نوعين هما :

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".

الخطأ من النوع الثاني : Type II error

وفي 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".

وقد يتساءل البعض عند مدى إمكانية تصغير الخطأين معاً ولكن لسوء الحظ لا يمكن تصغيرهما معاً إلى أدنى حد ممكن، ويبدو أن الطريقة الوحيدة المتاحة لذلك هي زيادة (أو تكبير) حجم العينة، الأمر الذي قد لا يكون ممكناً في كل الحالات. لذلك فإن الذي يحدث عادة هو تثبيت أحدهما كان يكون نسبة أو احتمال حدوث الخطأ من النوع الأول ومحاولة تصغير الآخر.

مستوى المعنوية : Level of Significance

والمقصود بمستوى المعنوية هو "احتمال حدوث الخطأ من النوع الأول". أو نسبة حدوثه "أي احتمال رفض الفرض العدمي بينما هو صحيح".

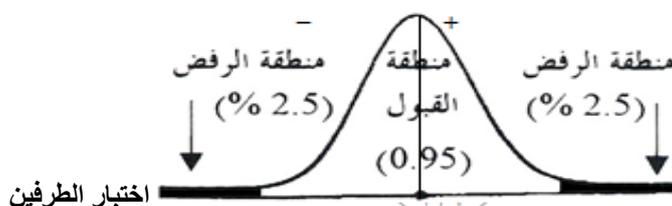
وعادة ما يرمز إلى مستوى المعنوية بالرمز اللاتيني ألفا وأشهر قيمتين لمستوى المعنوية هما 5%، 1%، ولكن ليس هناك ما يمنع من أن يأخذ قيمة أخرى.

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.

والفكرة الأساسية في اختبار الفرض هي تقسيم المساحة تحت المنحنى إلى منطقتين: أحدهما تسمى "منطقة القبول" أي منطقة قبول الفرض العدمي. والأخرى تسمى "منطقة الرفض"، أي منطقة رفض الفرض العدمي والتي تسمى أحياناً "بالمناطق الحرجة Critical region". والنقطة الجديرة بالملاحظة هنا هي أن منطقة القبول تمثل درجة الثقة، بينما تمثل منطقة الرفض مستوى المعنوية. وهناك ثلاث حالات مختلفة لمنطقتي القبول والرفض هي:

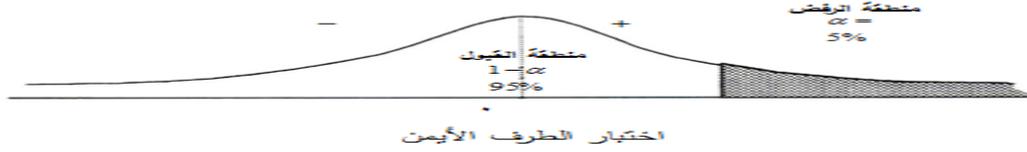
الأولى: إذا كان الفرض البديل يأخذ شكل "لا يساوي" كأن يكون الفرض في هذه الحالة هو أن متوسط دخل الفرد لا يساوي 200 ريال فإن منطقة الرفض تتكون موزعة على طرفي المنحنى بالتساوي، ويسمى الاختبار في هذه الحالة "اختبار الطرفين"، والذي يأخذ الشكل التالي (بافتراض أن $\alpha = 5\%$):

فالفرض العدمي هنا يعني أن متوسط دخل الفرد يساوي 200 ريال شهرياً، والفرض البديل في هذه الحالة هو بمعنى أن متوسط دخل الفرد لا يساوي 200 ريال شهرياً. حيث تمثل المنطقة البيضاء غير المظللة منطقة القبول والتي قد تساوي 95% وبالتالي فمناطق 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%.

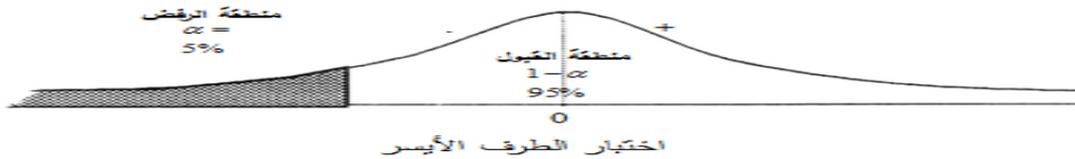
الثانية : إذا كان الفرض البديل يأخذ شكل "أكبر من" فإن منطقة الرفض تكون مركزة بالكامل في الطرف الأيمن للمنحنى. ويسمى الاختبار في هذه الحالة اختبار الطرف الأيمن. والذي يأخذ الشكل التالي أدناه :



فالفرض العدمي هنا نفس فرض المثال السابق، بينما الفرض البديل هو $H1: \mu > 200$

بمعنى أن متوسط دخل الفرد أكبر من 200 ريال 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 :



مع افتراض ثبات الفرض العدمي كما في المثال السابق، بينما الفرض البديل هو $H1: \mu < 200$ بمعنى أن متوسط دخل الفرد أقل من 200 ريال شهرياً، وبالتالي فإن مستوى المعنوية والذي يساوي مثلاً 5% مركز في الطرف الأيسر من المنحنى.

خطوات الاختبار الإحصائي :

(١) وضع الفرض العدمي H_0 ، 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 : $H_0: \mu = 20$

(٢) وضع 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 :

$$H1: \mu \neq 20$$

$$OR \mu > 20$$

$$OR \mu < 20$$

والفكرة الأساسية في اختبار الفرض هي تقسيم المساحة تحت المنحنى إلى منطقتين: أحدهما 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 **منطقة القبول تمثل درجة الثقة**، بينما تمثل **منطقة الرفض مستوى المعنوية**.

مثال (1) :-

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

الحل :-

١- الفرض العدمي : هو أن متوسط المجتمع يساوي 72 وبالرموز: $H_0: \mu = 72$

٢- الفرض البديل : هو أن المتوسط لا يساوي 72 وبالرموز: $H_1: \mu \neq 72$

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

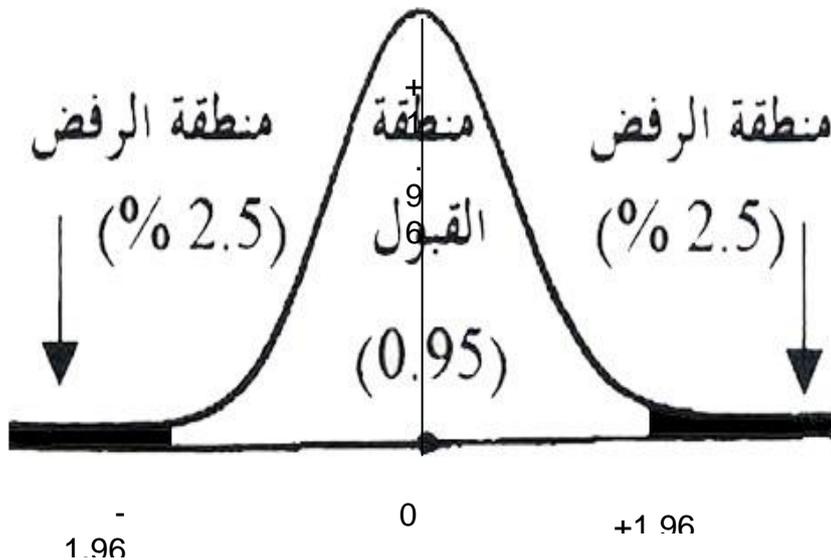
حيث $n = 49, \sigma = 14, \bar{X} = 75, \mu = 72$

وبالتعويض نحصل على: $Z_{\bar{X}} = \frac{75 - 72}{\frac{14}{\sqrt{49}}}$

$$Z_{\bar{X}} = \frac{3}{\frac{14}{7}} = \frac{3}{2} = 1.5$$

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 :

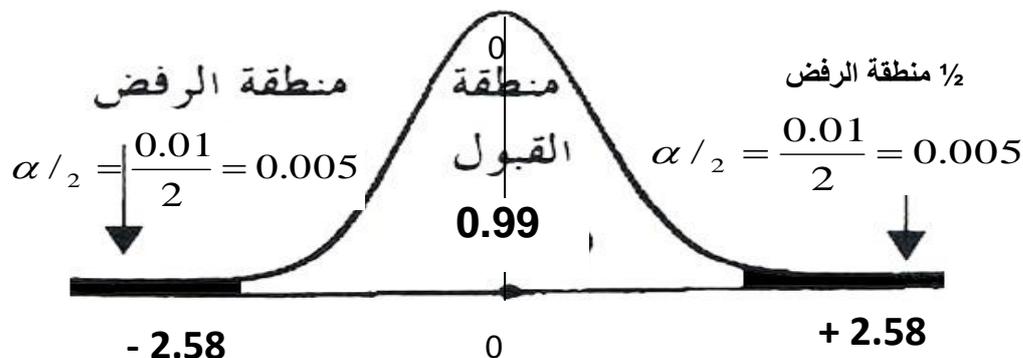


وقد حصلنا على حدود منطقتي القبول والرفض وذلك بقسمة درجة الثقة (المكاملة لمستوى المعنوية) والتي تساوي 0.95 على 2 فنحصل على 0.4750 وبالكشف في جدول التوزيع الطبيعي المعياري عن Z التي تقابل المساحة 0.4750 نجد أنها تساوي 1.96 وحيث أنها موزعة على طرفي المنحنى بمتساوي فنضع إشارة موجبة في النصف الأيمن، وإشارة 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 (٢): -

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:

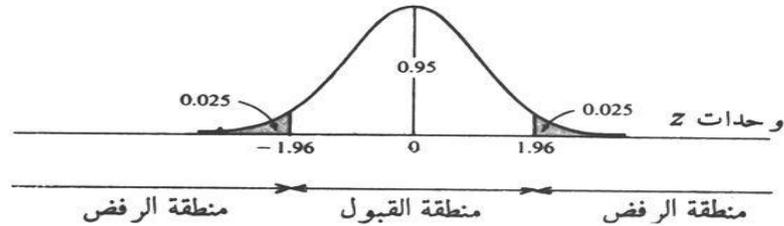
الحل :-

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

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

وحيث أن $n > 30$ ، فإذن 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 :

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



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 (3) :-

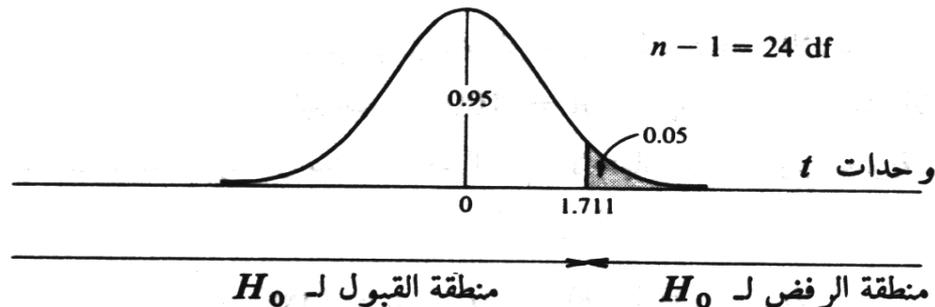
ترغب شركة أن تعرف بدرجة ثقة 95% ما إذا كان يمكنها الادعاء بأن صناديق الصابون المسحوق الذي
تبيعها تحتوي على أكثر من 500 جرام (حوالي 1.1 رطل) من الصابون. وتعرف الشركة من الخبرة الماضية
أن أوزان الصابون بالصناديق تتبع التوزيع الطبيعي وقد أخذت الشركة عينة عشوائية حجمها $n = 25$
ووجدت أن $X = 520$ جرام و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

الحل :-

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

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

وحيث أن التوزيع طبيعي ، $n < 30$ ، وكذلك σ غير معلومة، فعلى 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
وهي تقع داخل منطقة القبول، وتقبل H_0 أي $\mu = 500$ ، عند مستوى معنوية 5% (أو بدرجة ثقة 95%) .



مثال (٤) :-

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

الحل :-

١- الفرض العدمي : هو أن متوسط المجتمع يساوي 72 وبالرموز : $H_0 : \mu = 72$

٢- الفرض البديل : هو أن المتوسط لا يساوي 72 وبالرموز : $H_0 : \mu \neq 72$

٣- الإحصائية : بما أن العينة كبيرة فإن الإحصائية في حالة اختبار الوسط تأخذ الشكل التالي:

$$Z_{\bar{x}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

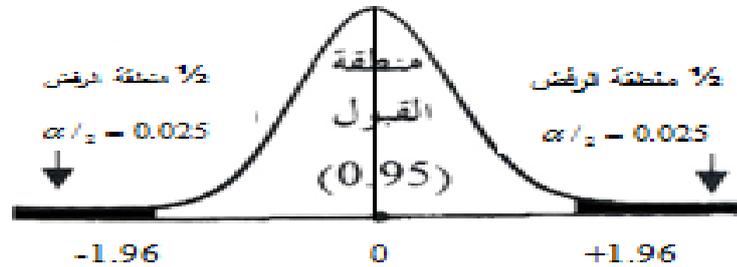
حيث أن :- $n = 49, \sigma = 14, \bar{X} = 75, \mu = 72$

وبالتعويض نحصل على :-

$$Z_{\bar{x}} = \frac{75 - 72}{\frac{14}{\sqrt{49}}}$$
$$Z_{\bar{x}} = \frac{3}{\frac{4}{7}} = \frac{3}{2} = 1.5$$

أي أن قيمة الاحصائي تساوي 1.5.

٤- حدود منطقتي القبول و الرفض : 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 :-

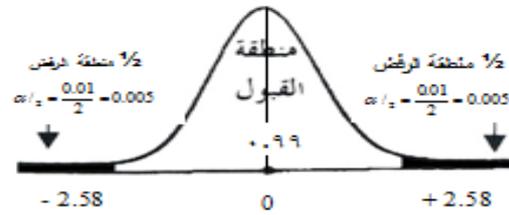


٥- المقارنة والقرار : وبمقارنة قيمة الإحصائية المحسوبة من الخطوة رقم 3 (والتي تساوي 1.5) بحدود منطقتي القبول والرفض (من الخطوة رقم 4) نجد أنها تقع في منطقة القبول لذلك فإن القرار هو :

قبول الفرض العدمي بأن متوسط دخول الأفراد الأسبوعية في هذه الدولة يساوي 72 دولاراً وذلك بمستوى معنوية 5%.

ملاحظة :

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 العدمي ولن يتغير بل يتأكد باستخدام مستوى معنوية 1%.

مثال (٥):-

يدعي أحد المرشحين في الانتخابات أنه سيحصل على نسبة 70% من أصوات الناخبين عندما تجري الانتخابات. ولاختبار هذا الادعاء تم اختيار عينة عشوائية من الناخبين حجمها 100 ناخب، ووجد أن نسبة من يؤيدون المرشح في العينة هي 60% اختبر مدى صحة ادعاء المرشح بأن النسبة في المجتمع هي 70% مقابل الفرض البديل أن النسبة أقل من 70% وذلك بمستوى معنوية 5%.

الحل :-

- الفرض العدمي هو أن النسبة في المجتمع (نسبة من يؤيدون المرشح في المجتمع) هي 0.70 أي أن الفرض العدمي هو أن الادعاء صحيح وأن المرشح سيحصل على النسبة التي ادعاها وهي 70% بالرموز $H_0 : P = 0.70$
- الفرض البديل والمنطقي : في هذه الحالة هو أن النسبة في المجتمع أقل من هذا الادعاء وبالرموز: $H_1 : P < 0.70$

٣ - الإحصائية : وتأخذ الإحصائية في حالة اختبار النسبة الشكل التالي :

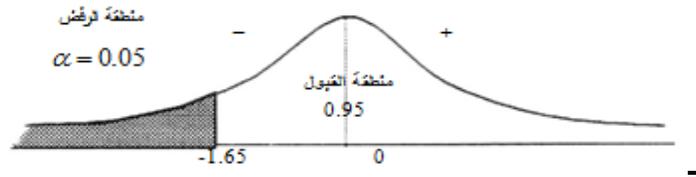
$$Z_{\hat{P}} = \frac{\hat{P} - P}{\sqrt{\frac{P(1-P)}{n}}}$$

حيث أن :- $n = 100, \hat{P} = 0.60, P = 0.70, 1 - p = 1 - 0.70 = 0.30$

$$\begin{aligned} Z_{\hat{P}} &= \frac{0.60 - 0.70}{\sqrt{\frac{0.70 \times 0.30}{100}}} \\ &= \frac{-0.10}{0.046} \\ Z_{\hat{P}} &= -2.17 \end{aligned}$$

أي أن قيمة الإحصائية تساوي -2.17

٤- حدود منطقتي القبول والرفض نحصل عليها من التوزيع الطبيعي المعياري، حيث مستوى المعنوية $\alpha = 5\%$ وبما أن الفرض البديل هو " أقل من " فنستخدم اختبار الطرف الأيسر.



٥- المقارنة والقرار : وبمقارنة قيمة الإحصائية التي حصلنا عليها في الخطوة رقم (٣) التي تساوي 2.17 - بحدود منطقتي القبول والرفض (من الخطوة رقم ٤) نجد أن قيمة الإحصائية تقع في منطقة الرفض لأن 2.17 - أصغر من 1.65 فبأن القرار هو :

رفض الفرض العدمي بادعاء المرشح بأن نسبة مؤيديه في المجتمع هي 70% وقبول الفرض البديل بأن النسبة أقل من 70% وذلك بمستوى معنوية 5% (أي أن احتمال الخطأ في هذا القرار لا يتعدى 5%).

مثال (٦):-

البيانات التالية تمثل نتائج عينتين عشوائيتين مستقلتين مسحوبتين من منطقتين لمقارنة متوسط عمر الناخب فيهما :

$$\bar{X}_1 = 35, \bar{X}_2 = 29, n_2 = 80, n_1 = 100$$

حيث

اختبر الفرض العدمي : أن متوسط عمر الناخب في المنطقة الأولى يساوي متوسط عمر الناخب في المنطقة الثانية بمستوى معنوية 5% مقابل الفرض البديل أنهما غير متساويين إذا علمت أن :

$$\sigma_1^2 = 60, \sigma_2^2 = 32$$

١- الفرض العدمي أن المتوسطين متساويين وبالرموز : $H_0: \mu_1 = \mu_2$

٢- الفرض البديل أن المتوسطين غير متساويين وبالرموز : $H_1: \mu_1 \neq \mu_2$

$$Z_{\bar{X}_1 - \bar{X}_2} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

٣- الإحصائية : تأخذ الشكل التالي :

وبالتعويض عن :- $n_1 = 100, n_2 = 80, \bar{X}_1 = 35, \bar{X}_2 = 29, \sigma_1^2 = 60, \sigma_2^2 = 32$

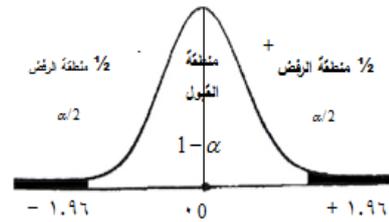
نحصل على :-

$$\begin{aligned}
 Z_{\bar{X}_1 - \bar{X}_2} &= \frac{35 - 29}{\sqrt{\frac{60}{100} + \frac{32}{80}}} \\
 &= \frac{60}{\sqrt{0.60 + 0.40}} \\
 &= \frac{6}{\sqrt{1}} = 6
 \end{aligned}$$

أي أن قيمة إحصائي الاختبار تساوي 6 .

٤- حدود منطقتي القبول والرفض التي نحصل عليها من جدول التوزيع الطبيعي Z لأن العينات كبيرة، والاختبار هو اختبار الطرفين (لأن الفرض البديل لا يساوي)

ومستوى المعنوية المطلوب هو 5% .



أي أن منطقة القبول تبدأ من -1.96 إلى +1.96 ومنطقة الرفض هي القيم التي أصغر من -1.96 - والتي أكبر من +1.96 .

٥- المقارنة والقرار ولما كانت قيمة الإحصائية (والتي تساوي) 6 تقع في منطقة الرفض فإن القرار هو رفض الفرض العدمي وقبول الفرض البديل بمستوى معنوية 5% أي أننا نرفض الفرض القائل بأن متوسط عمر الناخب في المنطقة الأولى يساوي متوسط عمر الناخب في المنطقة الثانية وذلك بمستوى معنوية 5% .

المحاضرة (٨)

اختبار الفروض الإحصائية المعلمية

انواع الاختبار (الفروض)

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 كبيراً ($n < 30$) 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:

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

ويعتبر ذلك مدخل ضروري لفهم اختبار t-test .

مثال على اختبار Z :

إذا كان متوسط استهلاك الفرد السعودي من الدجاج حسب تقارير وزارة الصحة هو (١٢) كيلوجرام بانحراف معياري (٦) كيلوجرامات لفترة السبعينات الميلادية. أجرى أحد الباحثين دراسة في عام ٢٠٠٣م من عينة قوامها (٤٩) فرداً ووجد أن متوسط الاستهلاك للفرد هو (١٤) كيلوجرام. هل تشير الدراسة الحالية أن متوسط الاستهلاك ارتفع عما عليه في السبعينات.

الحل:

(١) فرض العدم والفرض البديل.

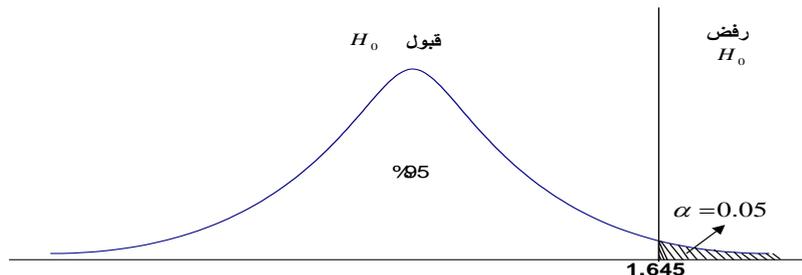
فرض العدم: $H_0: \mu=12$

الفرض البديل: $H_1: \mu>12$

(٢) مستوى الدلالة = (0.05):

(٣) إحصائية الاختبار (Z): $Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{14-12}{6/\sqrt{49}} = 2.33$

(٤) تحديد قيمة Z المعيارية من الجدول عند مستوى دلالة (٠.٠٥)، نحتاج لتحديد قيمة Z_{α} التي تقع على اليمين وتساوي ١.٦٤٥ (أنظر الشكل التالي):



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 اختلف بمستوى معنوي أو ذو دلالة عما عليه في سبعينات القرن الماضي.

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 اختبار (Z)، مما دفع كثيراً من الإحصائيين للبحث عن البديل المناسب.

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

مثال على اختبار t :

لو كانت لدينا عينة عشوائية تتكون من ٢٥٠ طالب وجد أن الوسط الحسابي لأطوال طلاب العينة ١٥٥.٩٥ سم، والانحراف المعياري = ٢.٩٤ سم، علماً بأن الوسط الحسابي لأطوال طلاب الجامعة يبلغ ١٥٨ 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 : $\alpha = 0.05$

منطقة الرفض : قيمة (ت) الجدولية عند مستوى دلالة $\alpha = 0.05$ ودرجات حرية ٢٤٩ = ١.٩٦٠

المختبر الإحصائي :

$$= 155.95 \text{ سم} ، = 250 \text{ طالب} ، = 2.94 \text{ سم}$$

$$= 158 \text{ سم}$$

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{155.95 - 158}{\frac{2.94}{\sqrt{250}}} = -11.006$$

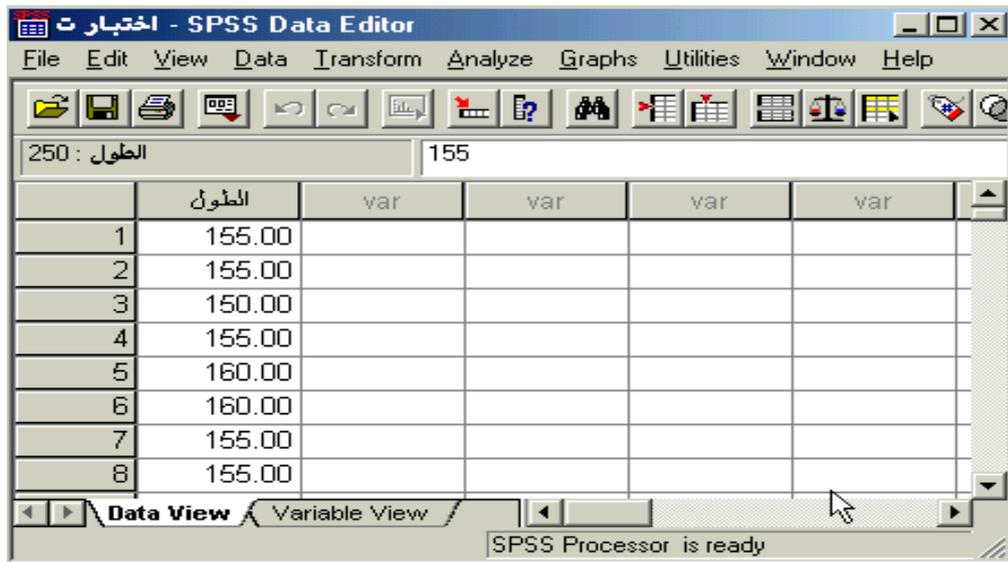
القرار :

٠.٠٠ قيمة ت المحسوبة (- ١١.٠٠٦) أكبر من قيمة ت الجدولة (١.٩٦) عند مستوى دلالة $\alpha = ٠.٠٥$.
∴ نرفض الفرضية الصفرية ونقبل البديلة .

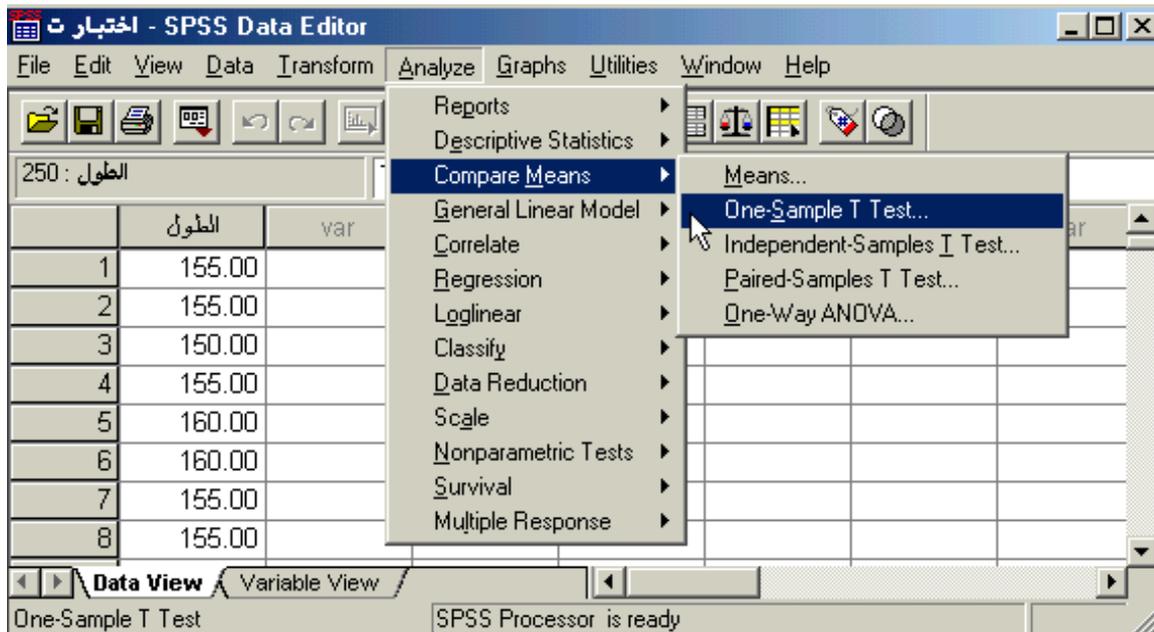
أي أنه توجد فروق ذات دلالة إحصائية بين الوسط الحسابي للعينة والوسط الحسابي لمجتمع البحث.

لغرض حساب قيمة (ت) لنفس المثال السابق من خلال استخدام برنامج ال-SPSS اتبع الخطوات التالية :

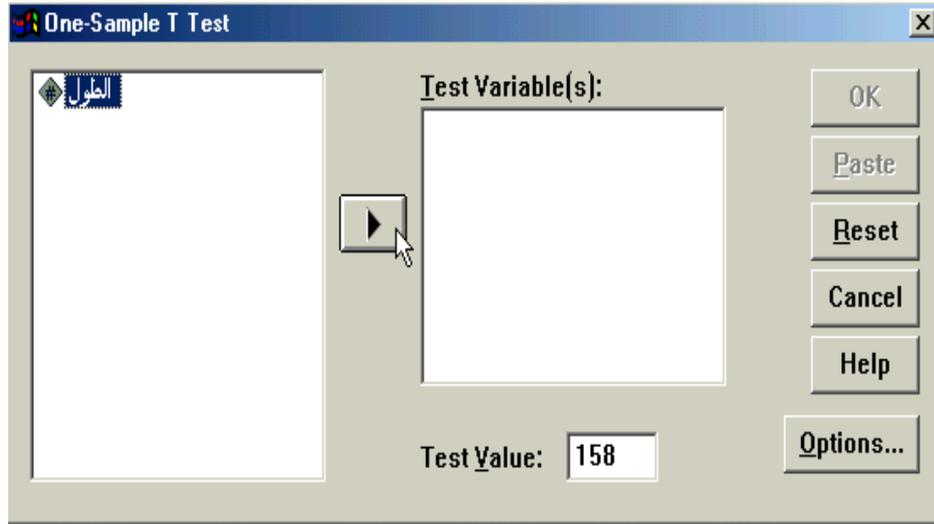
✓ قم بإدخال البيانات المراد تحليلها من خلال شاشة تحرير البيانات Data Editor بالطريقة المناسبة كالتالي :



✓ 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:



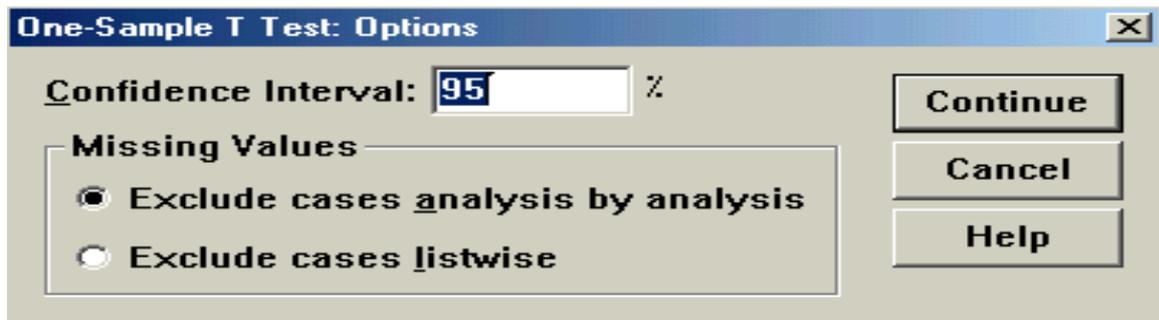
✓ بعد اختيار الأمر "اختبار (ت) لعينة واحدة" One-Sample T Test سوف يظهر لك صندوق الحوار التالي :



✓ من قائمة المتغيرات في الجهة اليسرى من صندوق الحوار انقر نقرا مزدوجا على المتغير "الطول" (أو انقر 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 ١٥٨ 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 ٩٥ %) ، وبعد الانتهاء من التعديل على هذا الصندوق الحواري انقر على زر "استمرار" Continue .



✓ انقر 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					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
الطول	-11.006	249	.000	-2.0480	-2.4145	-1.6815

يتضح من النتائج أن قيمة (ت) المحسوبة $t\text{-test} = -11.006$ ، ودرجات الحرية $df = 249$ ، وقيمة Sig. (2-tailed) = 0.000 ، وبما أن قيمة (2-tailed) Sig. (2-tailed) في الجدول (0.000) أصغر من قيمة $\alpha = 0.05$ فإننا بالتالي نرفض الفرضية الصفرية، أي أنه توجد فروق ذات دلالة إحصائية بين متوسط أطوال العينة ومتوسط أطوال طلاب الجامعة .

الاختبارات الاحصائية لعينتين مستقلتين Independent Samples t-test

مثال :-

أراد باحث أن يعرف أثر استخدام نظم مساندة القرارات على كفاءة القرارات التي تتخذها الإدارة بمساعدة تلك النظم، فوزع ٥٠ مديراً لمنشآت صناعية عشوائياً 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 بدلاً من الطريقة التقليدية فكانت النتائج كما يلي:

المجموعة الضابطة	المجموعة التجريبية
٢٥ =	٢٥ =
٦.٠ =	٧.٦٠ =
١.٧٨ =	٢.٢٧ =

فهل تدل هذه البيانات على أن أداء المجموعة التجريبية كان أفضل من أداء المجموعة الضابطة عند مستوى $\alpha = 0.05$ ؟

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 ($\mu_1 > \mu_2$)

مستوى الدلالة: $\alpha = 0.05$

منطقة الرفض: 0.05 قيمة مستوى الدلالة $\alpha = 0.05$ والاختبار بذيل واحد ، ودرجات الحرية = $25 + 25 - 2 = 48$ ، بذلك تكون قيمة (ت) الجدولية = 1.68

المختبر الإحصائي:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

ولتطبيق هذه العلاقة يلزمنا حساب قيمة الانحراف المعياري (S) من خلال العلاقة التالية:

$$S^2 = \frac{[(n_1 - 1)(S_1^2)] + [(n_2 - 1)(S_2^2)]}{(n_1 + n_2) - 2}$$

إذا التباين يساوي:

$$S^2 = \frac{[(25-1)(2.27)^2] + [(25-1)(1.78)^2]}{(25+25)-2} = 4.16$$

إذن الإنحراف المعياري يساوي :

$$S = \sqrt{S^2} = \sqrt{4.16} = 2.04$$

ثم نحسب قيمة (ت) من خلال تطبيق العلاقة التالية :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{7.60 - 6.0}{2.04 \sqrt{\frac{1}{25} + \frac{1}{25}}} = 2.77$$

القرار:

0.05 قيمة (t) المحسوبة (2.77) أكبر من قيمة (ت) الجدولية (1.68) عند مستوى دلالة $\alpha = 0.05$.

∴ نرفض الفرضية الصفرية ونقبل البديلة

أي أن المجموعة التي خضعت للتجربة يصبح أداءهم أفضل في عملية اتخاذ القرار من الذين لم يخضعوا للتجربة وذلك عند مستوى دلالة $\alpha = 0.05$.

الاختبارات الاحصائية لعينتين غير مستقلتين (العينات المرتبطة)

Paired Samples t-test

مثال :-

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، لابد على الباحث أن يتأكد من قيمة الارتباط بين الاختبارين والتي كانت $r = 0.46$ ، وقد كانت النتائج التي تم التوصل إليها كما يلي :

الاختبار القبلي	الاختبار البعدي
١٠٠ =	١٠٠ =
٥٤.٢٨ =	٥٨.٦٦ =
٤٩ =	٦٤ =

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 $\alpha = ٠.٠٥$ ؟

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 : $\alpha = ٠.٠٥$

منطقة الرفض : قيمة مستوى الدلالة $\alpha = ٠.٠٥$ والاختبار بذيولين، ودرجات الحرية $df = 100 - 1 = 99$ ، بذلك تكون قيمة (ت) الجدولية $= ١.٩٨٠$

المختبر الإحصائي :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} - 2r \left(\frac{S_1}{\sqrt{n_1}} \right) \left(\frac{S_2}{\sqrt{n_2}} \right)}}$$

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 (٥.٥٧) أكبر من قيمة (ت) الجدولة (١.٩٨٠) . عند مستوى دلالة $\alpha = ٠.٠٥$.

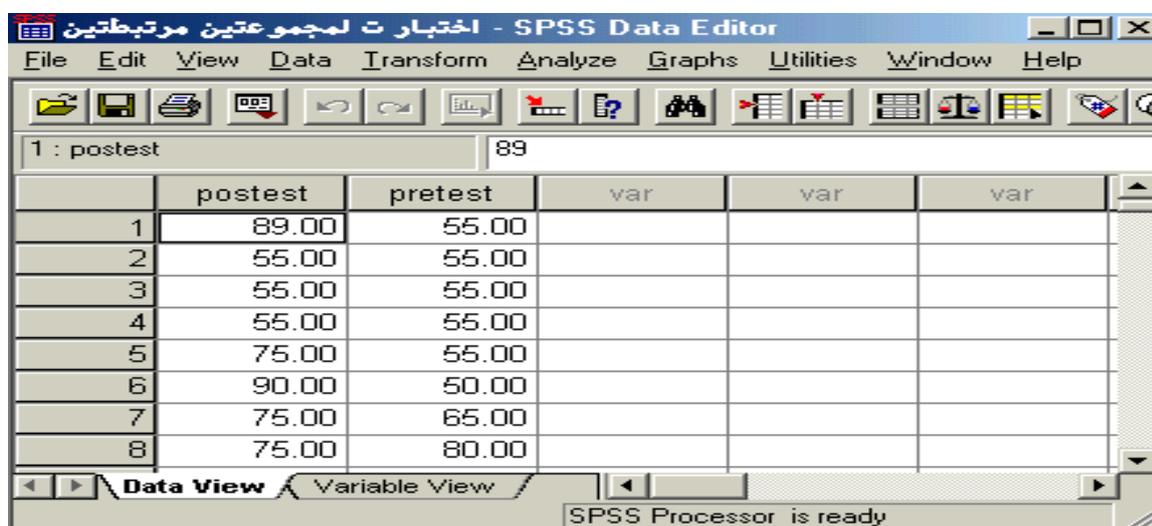
∴ نرفض الفرضية الصفرية ونقبل البديلة، أي أن للبرنامج التدريبي تأثير إيجابي على تحصيل الطلاب وأدائهم في الكلية وذلك عند مستوى دلالة α .

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 :

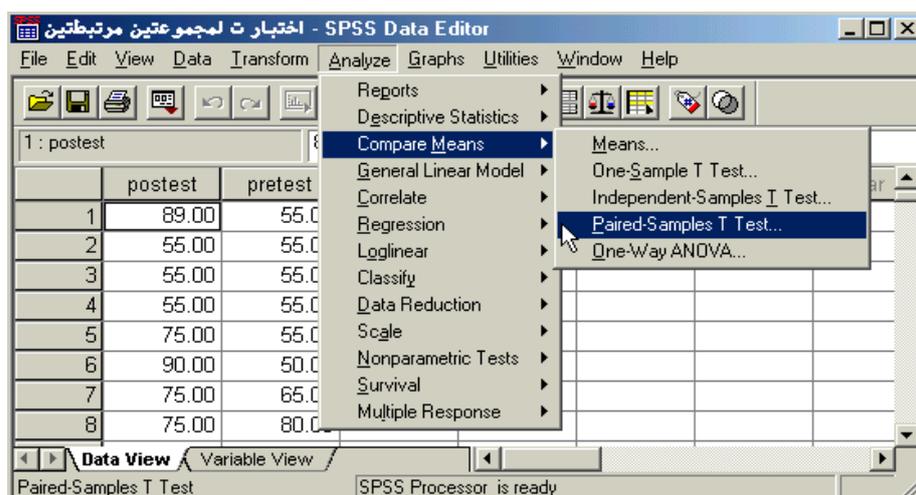
✓ قم بإدخال البيانات المراد تحليلها من خلال شاشة تحرير البيانات Data Editor بالطريقة المناسبة كالتالي :



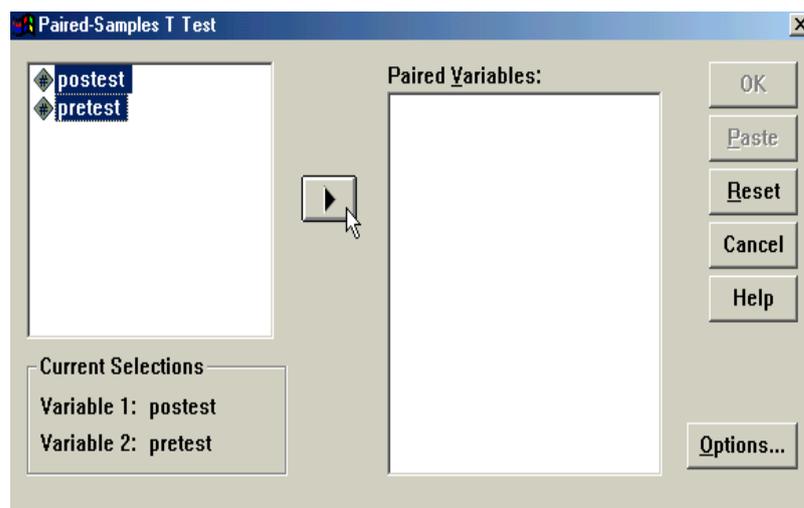
	posttest	pretest	var	var	var
1	89.00	55.00			
2	55.00	55.00			
3	55.00	55.00			
4	55.00	55.00			
5	75.00	55.00			
6	90.00	50.00			
7	75.00	65.00			
8	75.00	80.00			

لاحظ أنه تم إدخال البيانات بطريقة مختلفة عن ما تم اتباعه في حالة العينتين المستقلتين، هنا لا بد من إدخال بيانات كل متغير في عمود منفصل عن الآخر، وقد تم إعطاء كل متغير اسم مختلف عن الآخر posttest و pretest

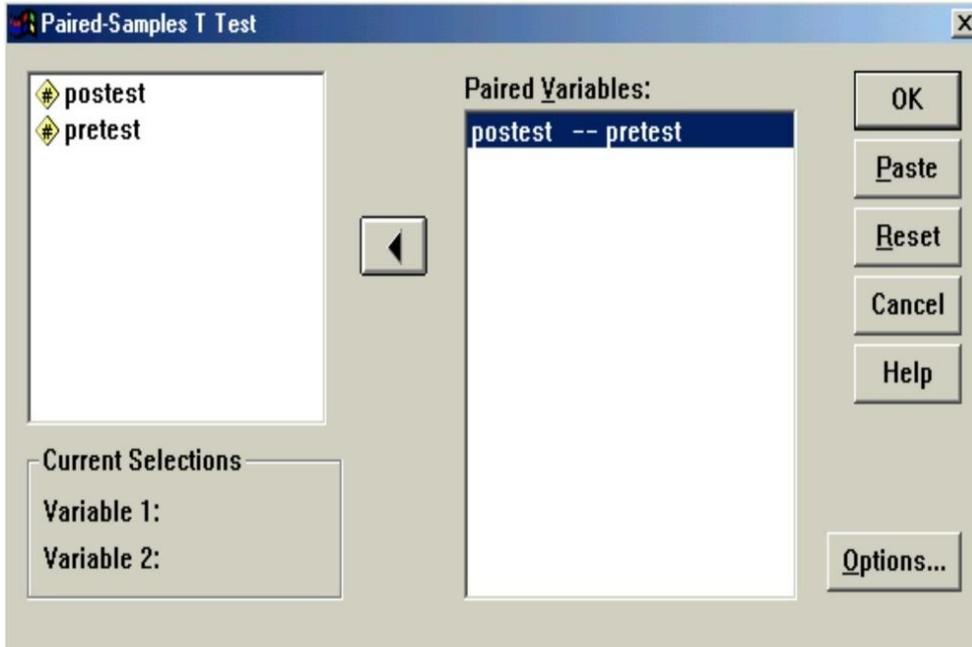
✓ من القائمة "تحليل" Analyze اختر الأمر "مقارنة المتوسطات" Compare Means فتظهر قائمة أوامر فرعية اختر منها "اختبار (ت) للعينات المرتبطة" Paired-Samples T-Test كالتالي :



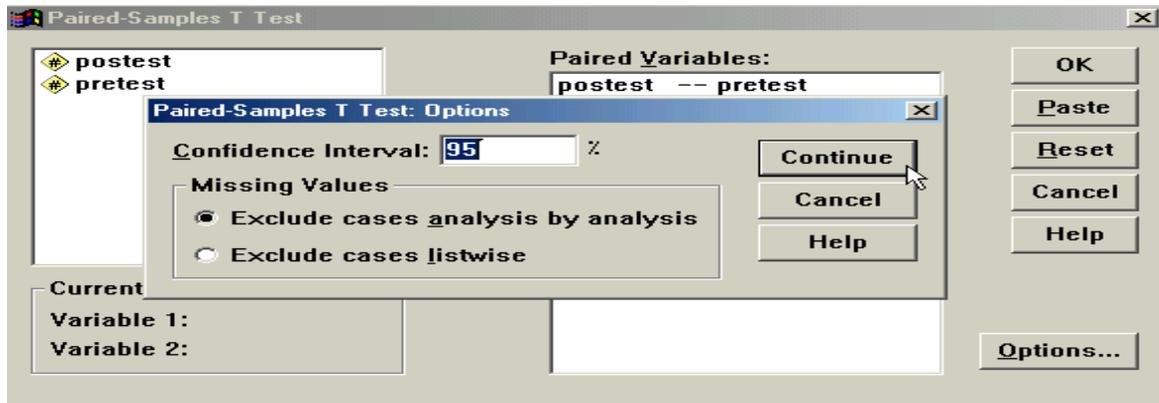
✓ بعد اختيار الأمر "اختبار (ت) للعينات المرتبطة" Paired-Samples T-Test سوف يظهر لك صندوق الحوار التالي:



✓ 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 أسفل قائمة المتغيرات)، ثم بعد ذلك انقر 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 ٩٥ %) ، 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	.000	

٨

نلاحظ أن برنامج الـ SPSS قام مباشرة بحساب الإحصاءات الأساسية للبيانات مثل المتوسط الحسابي للمتغير Posttest (٥٨.٦٦٠) والانحراف المعياري لنفس المتغير (٨.٠٠) ، أما المتغير Pretest فقد كان المتوسط الحسابي (٥٤.٢٨٠) والانحراف المعياري (٧.٠٠) . بالإضافة إلى ذلك تم حساب معامل ارتباط بيرسون للمتغيرات موضع الدراسة Paired Sample Correlation وقد كانت قيمته (٠.٤٥٨) .

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 $t\text{-test} = 5.575$ ، ودرجات الحرية $df = 99$ ، وقيمة (Sig. (2-tailed) = 0.000 ، وبما أن قيمة (Sig. (2-tailed) في الجدول (0.000) أصغر من قيمة $\alpha = 0.05$ 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 $\alpha = 0.05$.