ملخص التحليل الإحصائي

جامعة الملك فيصل – إدارة أعمال - المستوى الرابع

ابو جنى الجوني

<u>المجموعات .</u>

<u>1- المجموعة الخالية:</u>

ويرمز للمجموعة الخالية بالحرف اليوناني Ø" فاي" .

2- المجموعة المنتهية:

المجموعة التي تكون عناصرها محدودة.

3- المجموعة غير المنتهية:

المجموعة التي تكون عناصرها غير محدودة.

4- المجموعة الكلية:

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

5- المجموعة الجزئية:

فنقول عن مجموعة A أنها مجموعة جزئية من مجموعة B إذا كان كل عنصر ينتمي إلى A ينتمي إلى B

A ✚ B

6- تساوي المجموعات:

تكون المجموعتان A ،B **<u>متساويتان</u>** إذا كانت $A \times B, B \times A$ ✱ A ⌔ B

أما المجموعتان <u>المتكافئتان</u> فهما المجموعتان اللتان تتساويان في عدد عناصرهما وتكتب على الصورة A ◖ B

7- الاتحاد

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

8- التقاطع

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

<u>9- المكملة أو المتممة:</u>

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

10- الفرق

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

تعريف الاحتمالات:

يمكن تعريف الاحتمالات بطرق عديدة غير أن أبسطها

"هو مقياس لامكانية وقوع حدث (Event) معين"

1–التجربة العشوائية Random Experiment :

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

مثلا:

رمي حجر نرد مرة واحدة يعتبر تجربة عشوائية، حيث نعلم جميع قيم نتائج التجربة وهي إما 1 أو 2 أو 3 أو 4 أو 5 أو 6 ولكن لا يمكن تحديد أي الأرقام يظهر إلى الأعلى بصورة مؤكدة قبل إجراء التجربة.

- رمي عملة معدنية مرة واحدة أو عدد من المرات يعتبر تجربة عشوائية معروف جميع نتائجها قبل أن تبدأ التجربة، ولكن لا يمكن الجزم بظهور أي منها في رمية معينة.
- المشاركة في سباق الخيل لحصان معين يعتبر تجربة عشوائية فهو إما أن يفوز أو يخسر أو يتعادل.

2–الحادث والفراغ العيني:

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

الحادثة :

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

أوجد فراغ العينة في كل من التجارب العشوائية التالية:

1. رمي عملة معدنية واحدة.
2. رمي عملة معدنية مرتين.

الحل:

1– $\Omega = \{H , T\}$

2– $\Omega = \{(H,H), (H,T), (T,H), (T,T)\}$

مثال :

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

- الحصول على H (صورة) مرة واحدة

- الحصول على H (صورة) مرتين
- الحصول على H (صورة) ثلاث مرات
- عدم الحصول على H (صورة)

<u>الحل:</u>

- عند رمي عملة معدنية ثلاث مرات فيكون بالتالي **فراغ العينة :**

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

- ويمكن الحصول على الحادثة **H (صورة) لمرة واحدة** ونرمز لها بالرمز A1 كالتالي:

A1 ={HTT, THT, TTH}

ويمكن الحصول على الحادثة **H (صورة) لمرتين** ونرمز لها بالرمز A2 كالتالي:

A2 ={HHT, HTH, THH}

- ويمكن الحصول على الحادثة **H (صورة) ثلاث مرات** ونرمز لها بالرمز A3 كالتالي:

A3 ={HHH}

- ويمكن **عدم الحصول** على الحادثة **H (صورة)** ونرمز لها بالرمز A4 كالتالي:

A4 ={TTT}

<u>مثال :</u>

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

- A : الحصول على مجموع يساوي 7
- B : الفرق بين العددين الناتجين يساوي القيمة المطلقة 1
- C : الحصول على مجموع يساوي 9 على الأقل
- D : الحصول على 1 في الرمية الأولى
- E : الحصول على حاصل ضرب يساوي 6 على الأكثر
- F : الحصول على مجموع أقل من أو يساوي 2

<u>الحل :</u>

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

<u>أولا – الحصول على مجموع يساوي 7</u>

– **بطريقة سرد جميع العناصر بينهما فاصلة (طريقة نقاط العينة):**

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

– **بطريقة الصفة المميزة:**

A={ (x,y) : x + y = 7 }

<u>الفرق بين العددين الناتجين يساوي القيمة المطلقة 1</u>

الحادثة	التعبير بالكلمات عن الحوادث

– بطريقة سرد جميع العناصر بينهما فاصلة (طريقة نقاط العينة):

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

– بطريقة الصفة المميزة:

B={ (x,y) : x – y = | 1 |}

<u>الحصول على مجموع يساوي 9 على الأقل</u>

– بطريقة سرد جميع العناصر بينهما فاصلة (طريقة نقاط العينة):

C={(4,5), (4,6), (5,4), (5,5), (5,6), (6,3), (3,6). (6,4), (6,5), (6,6)}

– بطريقة الصفة المميزة:

C={ (x,y) : x + y ≥ 9 }

<u>الحصول على الرقم 1 في الرمية الأولى</u>

– بطريقة سرد جميع العناصر بينهما فاصلة (طريقة نقاط العينة):

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

– بطريقة الصفة المميزة:

D={ (x,y) : x = 1 }

<u>الحصول على حاصل ضرب يساوي 6 على الأكثر</u>

– بطريقة سرد جميع العناصر بينهما فاصلة (طريقة نقاط العينة):

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

– بطريقة الصفة المميزة:

E={ (x,y) : x * y ≤ 6 }

<u>الحصول على مجموع أقل من أو يساوي 2</u>

– بطريقة سرد جميع العناصر بينهما فاصلة (طريقة نقاط العينة):

F={(1,1)}

– بطريقة الصفة المميزة:

F={ (x,y) : x + y ≤ 2 }

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

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

H={(1,1), (2,1), (1,2), (3,1), (2,2), (1,3)}

I={(5,1), (1,5), (6,2), (2,6)}

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

K={(2,2), (2,4), (2,6), (4,2), (4,4), (4,6), (6,2), (6,4), (6,6)}

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

3–الحالات الممكنة (Possible Cases)

هي الحالات أو النتائج المختلفة التي يمكن أن تظهر نتيجة لإجراء تجربة معينة، فمثلاً عند رمي قطعة عملة تكون نتيجتها صورة أو كتابة ، بمعنى ان الحالات الممكنة لقطعة العملة 2 .

4–الحالات المواتية (Favorable Cases)

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

5–الحالات المتماثلة (Equally Likely Cases)

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

6–الحوادث المتنافية (Mutually Exclusive Events)

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

7–الحوادث المستقلة (Independent Events)

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

8–الحوادث الشاملة (Exhaustive Events)

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

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

التعريف النسبي للاحتمالات Rational Probability Definition :

حاصل قسمة عدد مرات حدوث الحادثة A مقسوما على عدد مرات حدوث التجربة .

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

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

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

التعريف التقليدي للاحتمالات Classical Probability Definition :

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

مثال:

رمي حجر نرد مرد واحدة ، أحسب التالي:

- احتمال الحصول على رقم 5
- احتمال الحصول على رقم زوجي
- احتمال الحصول على رقم أكبر من 2
- احتمال الحصول على رقم أقل من 7
- احتمال الحصول على رقم 7

الحل:

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

P(A=5)=1/6
P(A=2,4, 6)=3/6
P(A>2)=4/6
P(A<7)=6/6=1
P(A=7)=0/6=0

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

مثال:

الجدول التالي يمثل توزيع عمال أحد المصانع حسب الحالة الاجتماعية للعامل والقسم 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 = {أن يكون العامل أعزب} فيكون الإحتمال المطلوب:

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

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

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

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

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

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

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

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

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

أ- في حالة كون الحوادث متنافية

إذا كانت الحوادث A1 , A2 , A3 .... 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**

فإذا كان B ، A حادثين متنافيين فإن P (A أو B) = P(A) + P(B)

يرمز أيضاً لاحتمال وقوع أحد الحادثين بالرمز

حوادث متنافية

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

ب- في حالة كون الحوادث غير متنافية

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

P (A او B)= P (A) + P (B) – P (A و 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

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

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

-1

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

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

فيكون الاحتمال المطلوب:

$P(A1)=12/50$

$P(A2)=22/50$

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

-2

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

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

فيكون الإحتمال المطلوب:

$P(A1)=27/50$

$P(A2)=12/50$

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

3-

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

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

فيكون الاحتمال المطلوب:

$P(A1)=16/50$

$P(A2)=23/50$

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

- <u>الاحتمال الشرطي</u>

إذا كان لدينا الحادثين A1 , A2 وكان P: (A2) لا يساوي الصفر فأن الاحتمال الشرطي للحادث A1 بشرط وقوع الحادث A2 يعطي بالمعادلة التالية:

$P(A1|A2) =$

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

مثال:

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

الحل:

نفرض أن A1={نجاح الطالب في مقرر الإحصاء}

A2={نجاح الطالب في مقرر الرياضيات}

وبذلك يكون:

$P(A2)=0.64$

$P(A1 \cap A2)=0.32$

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

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

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

مثال:

الجدول التالي يمثل توزيع عمال أحد المصانع حسب الحالة الاجتماعية للعامل والقسم 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 A1={أن يكون العامل من القسم الأول}

A2={أن يكون العامل متزوج}

B3={أن يكون العامل من القسم الثالث}

B4={أن يكون العامل أعزب}

فيكون بالتالي:

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

احتمال أن يكون من القسم الأول ومتزوج

احتمال أن يكون متزوج

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

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

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

احتمال أن يكون العامل أعزب بشرط أنه من القسم الثالث

احتمال أن يكون من القسم الثالث

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

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

<u>ضرب الاحتمالات</u>

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

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

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

<u>مثال:</u>

إذا رمينا قطعة نقود مرة واحدة ، احسب الاحتمالات التالية:

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

الحل:

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

A2={ظهور كتابة في الرمية الثانية}

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

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

$$P(A_1 + A_2) = P(A_1) \cdot (P(A_2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

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

B2={ظهور صورة في الرمية الثانية}

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

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

$$P(B_1 + B_2) = P(B_1) \cdot (P(B_2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

مما سبق يمكن القول أن الحوادث المعطاة تكون مستقلة عندما تبقى الاحتمالات ثابتة .

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 :

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

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

1. احتمال أن يكون العاملان من القسم الأول؟
2. احتمال أن يكون العاملان متزوجان؟
3. احتمال أن يكون للعاملان نفس الحالة الاجتماعية؟
4. احتمال أن يكون العاملان من القسم نفسه؟

الحل:

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

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

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

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

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

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

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

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

التباديل:

مثال :

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

قاعدة:

$$P_r^n = \frac{n!}{(n-r)!}$$

مثال: رشح عشرة أشخاص لشغل الوظائف الأربع التالية: رئيس مجلس إدارة، نائب رئيس مجلس إدارة، مدير عام، ومدير إدارة موارد بشرية.

بين عدد الطرق الممكنة لشغل هذه الوظائف الأربع من بين هؤلاء المرشحين.

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

ويمكن حساب ذلك من خلال المعادلة السابقة:

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

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

قاعدة:

إذا كانت n=r فإن عدد التراتيب الممكنة لعدد n من المفردات مأخوذة مرة واحدة هو :

$$P_n^n = n!$$

مثال:

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

5!=(5)(4)(3)(2)(1)=120

مثال:

إذا كان لدينا الحروف التالية a, b, c

كم عدد المرات التي يمكننا أن نرتب هذه الحروف

abc, acb, bac, bca, cab, cba

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

<u>التوافيق :</u>

تهتم التوافيق بعدد الطرق الممكنة لاختيار r من المفردات من n من المفردات دون أخذ الترتيب في الإعتبار،

ويسمى ذلك بعدد التوفيقات الممكنة لأخذ r من المفردات من n من المفردات.

<u>قاعدة:</u>

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

$$C_r^n = \frac{n!}{r!}$$

مثال:

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

الحل:

هنا نجد أن r=3 و n=10 وبالتالي :

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

أي أنه توجد 120 طريقة لاختيار ثلاثة أفراد من بين عشرة مرشحين لعضوية مجلس الإدارة، لاحظ أن ترتيب اختيار المرشحين الثلاثة ليست له أهمية تذكر.

<u>الفرق بين التباديل والتوافيق:</u>

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

<u>مثال:</u>

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

الحل:

وحيث أن ترتيب اختيار الطلبة ليست له أهمية، فإننا سنستخدم التوافيق لتحديد عدد التوفيقات الممكنة (عدد الطرق الممكنة) لاختيار الطلبة وذلك كما يلي:

$$C_r^n = \frac{n!}{r!} = C_3^{20} = \frac{(20)(19)(18)}{(3)(2)(1)} = \frac{6480}{6} = 1140$$

<u>مثال:</u>

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

الحل:

وحيث أن الجوائز مختلفة لذا فإن ترتيب الاختيار ضروري، وبالتالي تستخدم التباديل لتحديد عدد الطرق الممكنة، لاحظ وجود 20 طريقة لمنح الجائزة الأولى، و19 طريقة لمنح الجائزة الثانية، و 18 طريقة لمنح الجائزة الثالثة، أي أن الطرق الكلية (عدد التبديلات الكلية) لمنح الجوائز الثلاث لثلاث طلاب من عشرين طالبا هو:

$$P_r^n = n! = P_3^{20} = 20(19)(18) = 6840$$

نظريــة بايـز (Bayes' Theorem)

مثال:

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

1- أن تكون القطعة المعيبة من إنتاج الآلة الأولى؟

2- أن تكون القطعة المعيبة من إنتاج الآلة الثانية؟

الحل:

نفرض أن

A1={إنتاج الآلة الأولى} P(A1)=0.2

A2={إنتاج الآلة الثانية} P(A2)=0.35

A3={إنتاج الآلة الثالثة} P(A3)=0.45

B= {إنتاج سلعة معينة}

فيكون بالتالي:

P(B│A1)=0.02

P(B│A2)=0.025

P(B│A3)=0.03

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

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

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

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

مثال:

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

1- أن يكون العامل من القسم الأول؟

2- أن يكون العامل من القسم الثاني؟

3- أن لا يكون العامل من القسم الأول؟

الحل:

نفرض أن

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

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

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

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

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

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

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

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

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

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

<u>المتغيرات العشوائية</u>

المتغير العشوائي Random Variable**:**

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

وينقسم المتغير العشوائي إلى قسمين هما:

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

<u>المتغيرات العشوائية المتقطعة (المنفصلة):</u>

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,

<u>التوزيع الاحتمالي للمتغير العشوائي المنفصل:</u>

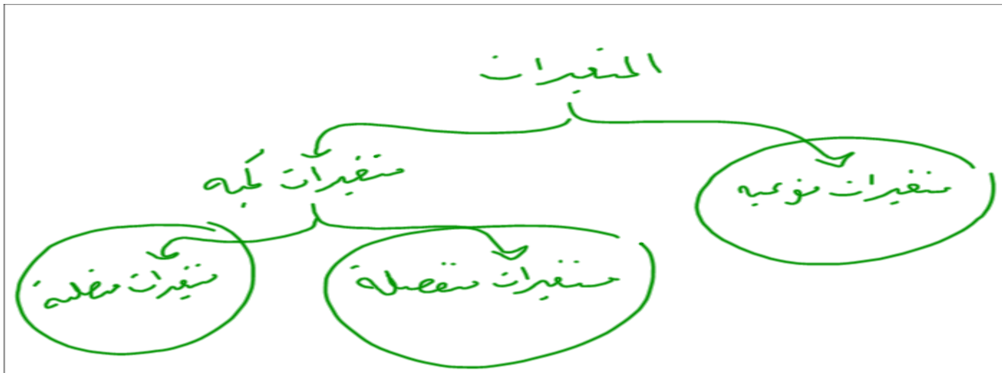
التوزيع الاحتمالي، هو الذي يبين احتمالات حدوث القيم التي يمكن يأخذها المتغير، والتي ترتبط باحتمالات النتائج الممكنة في فراغ العينة، وبمعنى آخر هو التكراري النسبي للقيم التي يمكن أن يأخذها المتغير.

<u>المتغيرات العشوائية المستمرة Continuous Random Variables</u>

المتغير العشوائي المستمر هو الذي يأخذ قيما متصلة، ويأخذ عدد لانهائي من القيم الممكنة له داخل مجاله .

<u>**التوزيع الاحتمالي للمتغير المستمر** Continuous Probability</u>

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



التوزيعات الاحتمالية للمتغيرات المنفصلة:

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

أ – توزيع ذي الحدين:

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

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

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

تحديد شكل التوزيع:

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

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

$p > 0.5$ **الاحتمال أكبر من 0.5**

ب– توزيع بواسون:

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

عندئذ :

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

<u>دائما توزيع بواسون موجب الالتواء</u>

يعنى **التوزيع الإحصائي** الشكل الذي تأخذه مجموعة البيانات، وشكل البيانات مهم جدا في تحليلها ووصفها وكخطوة تسبق قرار استخدام أي اسلوب احصائي.

ويرتبط التوزيع الاحصائي عادة بنوعين من البيانات المتصلة والمنفصلة، ويناسب النوع المنفصل المقاييس الاسمية والرتبية، وهناك بعض المقياس المنفصلة ثنائية أي انه لايوجد بها الا قيميتين، وهي لا تسمي توزيعات طبيعية وانما تسمى توزيعات ثنائية، ومن أهم مقاييس التوزيعات المنفصلة مقياس ذو الحدين وذلك عائد لان الاجابة على المقياس الاسمي اما نعم أو لا ، ولذلك غالبا ما يرمز لها في الحاسب بصفر (غياب الصفة) [ذكور – لا] أو 1 (وجود الصفة) [اناث – نعم].

أما التوزيعات الاحصائية المتصلة فهي ذات أهمية كبيرة في العلوم الإحصائية وذلك لأن اغلب الاختبارات الاحصائية تتعامل مع هذا النوع من البيانات.

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

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

التوزيع الطبيعي :

هو أفضل وأكثر التوزيعات الاحتمالية المتصلة استخداماً في النواحي التطبيقية، ومنها الاستدلال الإحصائي شاملا التقدير، واختبارات الفروض، كما أن معظم التوزيعات يمكن تقريبها إلى هذا التوزيع، وفيما يلي عرض لهذا التوزيع. ويكمن السبب في ذلك :

- أن توزيعات كثيرة لمتغيرات مثل الطول والوزن تتبع توزيعات طبيعية.
- النتيجة الرياضية التي تسمى بنظرية النهاية المركزية.

والتوزيع الطبيعي هو توزيع احتمالي متصل ويعد من أكثر التوزيعات استخداماً فى التحليل الإحصائى، وهو جرسي الشكل ومتماثل حول الوسط الحسابي، ويمتد إلى مالا نهاية فى الاتجاهين، ولكن معظم المساحة (الاحتمال) تتركز حول الوسط الحسابى.

خصائص التوزيع الطبيعي:

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

- توزيع جرسي أي يشبه الجرس.
- توزيع متصل
- توزيع متماثل حول الوسط
- الالتواء (الاطراف) والتفلطح (القمة) يساوي صفر.

- ومن أهم صفاته أن يتصف بمنوال ووسط ووسيط واحد وذات قيم متساوية بمعنى أن الجزء الذي على يمين الوسط مطابق للجزء الايسر
- الذيلين الايمن والايسر يقتربان من الخط الافقي ولكن لا تلامسه
- المساحة الكلية 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.
- القيمــة الصـغيرة لـ σ تعني أنــه لدينا جرس طويل مدبب، والقيمة الكبيرة لـ σ تعني أن الجرس قصير ومفرطح.

<u>ب-التوزيع الطبيعي القياسي المعياري :</u>

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

ويمكن تحويل أي توزيع طبيعي (بوحدات x) إلى توزيع طبيعي قياسي (بوحدات z) .

ولإيجاد الاحتمالات (المساحات) فى مسائل تحتوى على التوزيع الطبيعى ، فإننا نحول أولا قيم x إلى قيم z المناظرة لها ، من خلال المعادلة التالية:

$$Z = \frac{X - \mu}{\sigma}$$

- 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

<u>توزيع t ستيودنت :</u>

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

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

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

ولا شتقاق المتغير العشوائي t من المتغير العشوائي الاعتدالي، فإن ذلك يتطلب معرفة قيمة المتوسط µ للمتغير العشوائي الاعتدالي، بينما لانحتاج إلي معرفة انحرافه المعياري. $t= X-\mu/(s/n)$

خصائص توزيع t

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

حيث df هي درجة حرية المتغير العشوائي t .

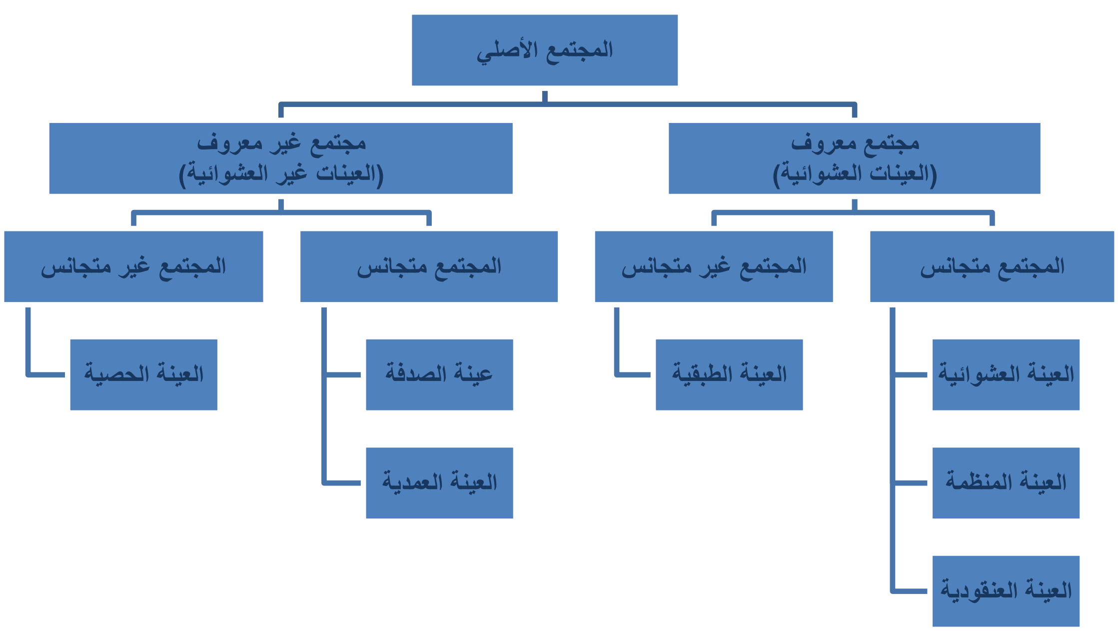
ويتبين من المعادلة السابقة للانحراف المعياري أنه كلما زادت درجات حرية المتغير العشوائي t بحيث تصل إلي 30 فأكثر، فإن الانحراف المعياري يقترب من الواحد الصحيح، وبصفة عامة فإن الإنحراف المعياريي لتوزيع t يساوي 1.035 أو أقل.

ويختلف الجدول الاحصائي t عن الجدول الاحصائي لتوزيع ذي الحدين , حيث يعطي جدول t نقاط القطع (القيم الحرجة) بينما يعطي جدول ذي الحدين الاحتمالات . مع زيادة حجم العينة ومن ثم درجات الحرية .

توزيعات المعاينة

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

أقسام العينات:



أ – العينات الاحتمالية:

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

ب – العينات غير الاحتمالية:

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

أخطاء البيانات الإحصائية:

تتعرض البيانات الإحصائية التي يتم جمعها إلى نوعين من الأخطاء:

1. **خطأ التميز أو التحيز:** وهو ذلك الخطأ الناتج عن مصادر متعددة، منها أخطاء في تصميم البحث أو التجربة أو أخطاء فنيه أثناء جمع البيانات أو خلال العمليات الحسابية التي تتم على البيانات المتجمعة..

أخطاء التميز تزداد بازدياد الفروق بين الإمكانيات (المادية والفنية) اللازم توافرها لضمان أقصى درجة دقة ممكنه وبين الإمكانيات الفعلية المتاحة للباحث.

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

أسباب خطأ التميز أوالتحيز:

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

كيفية التقليل من أخطاء التميز أوالتحيز:

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

2. **خطأ المعاينة العشوائية 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 .

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 **<u>وهذا الخطأ يسمى</u>** بخطأ المعاينة العشوائي .

كيف نقلل من خطأ المعاينة العشوائي:

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

تقويم عيِّنة الدراسة:

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

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

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

<u>المعالم والإحصاءات:</u>

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

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

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

وللتفرقة بين المعالم والإحصاءات يجب أن نرمز لكل منها برموز تختلف عن رموز الأخرى، على سبيل المثال :

يرمز للمتوسط الحسابي للمجتمع بالرمز μ

بينما يرمز للمتوسط الحسابي للعينة بالرمز $\bar{x}$

أيضاً للانحراف المعياري للمجتمع بالرمز σ

بينما يرمز للانحراف المعياري للعينة بالرمز S

<u>نظرية Central Limit Theorem</u>،

عند اخذ عينات بحجم كبير من أي مجتمع فان معدلات العينات ستتوزع بصورة طبيعية Normal، وان متوسط متوسطات العينات سيقترب من متوسط مجتمع الدراسة.

حساب متوسط قيم العينة ودرجة التباين Variance فيها وفق المعادلة:-

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

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

ونحسب الانحراف المعياري وهو الجذر التربيعي للتبيان :-

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

إن ارتفاع قيمة الانحراف المعياري يدل على التباين الكبير بين قيم العينة، ولهذا أهمية خاصة عند تحديد حجم العينة. **<u>فكلما كان التباين كبيرا في خصائص المجتمع كانت معدلات العينة بعيدة عن معدل مجتمعها،</u>** ولهذا فان قيمة الخطأ المعياري SE. Standard Error of Sample Standard Deviation في قيمة الانحراف المعياري للعينة ستكون كبيرة.

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

الخطأ المعياري:

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

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

وفق هذه المعادلة تؤخذ نسبة العينة إلى مجتمعها، وكلما كبرت هذه النسبة تحسن تمثيلها لمجتمع الدراسة، أما عندما يكون حجم مجتمع العينة مجهولا، حينها تعتمد المعادلة الآتية:–

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

أما عندما يكون حجم العينة أكثر من (100) فتعتمد المعادلة أدناه:–

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

ويقدم كريكوري تعديل بيسل عندما يكون حجم العينة اقل من (30):–

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

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

مستوى الثقة وحدودها:

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

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

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

- مستوى ثقة إحصائية قدرها (99%) أو باحتمالية قدرها (0.99) يقع متوسط مجتمع الدراسة بين قيمة متوسط متوسطات العينات و (+ و –) ثلاث درجات من قيمة الخطأ المعياري تقريبا.

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

وتحسب حدود الثقة لاعتمادها في تقدير متوسط مجتمع الدراسة كما يأتي :–

$$\mu = \bar{X} \pm Z \sigma_{\bar{x}}$$

(أ) التوزيع الطبيعي للقيم:

كلما كبر حجم العينة ازدادت دقة تمثيلها لمجتمعها واقترب توزيع القيم فيها من التوزيع الطبيعي (المتماثل الجانبين) وأصبحت عملية الاستدلال أكثر دقة.

فكلما ازداد حجم العينة ازدادت الثقة بتقديرات خصائص المجتمع وصغرت معه حدود الثقة.

وبتطبيق المعادلة المذكورة في أدناه يستطيع تحديد حجم العينة المناسبة لدراسته:

$$n = \frac{Z^2 \sigma^2}{e^2}$$

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

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

(ب) توزيع ت،T للقيم :

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

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

ودرجة الحرية تفضل على حجم العينة في الأحجام الصغيرة للعينة لأنها تقلل من الانحياز في تقدير خصائص مجتمع الدراسة.

العوامل المحددة لحجم العينـة:

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

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

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

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

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

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

توزيع المعاينة

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

توزيع المعاينة للمتوسط:

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

التقدير :

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

والمقصود بالتقدير هو تقدير معالم المجتمع الإحصائي (أو التوزيع الاحتمالي) والتي غالباً ما تكون مجهولة.

وهناك نوعان (أو أسلوبان) للتقدير:

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

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

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

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

وطالما نحن نتكلم عن تقدير متوسط المجتمع فإن أول ما نفكر فيه هو الوسط الحسابي للعينة،

وفترة التقدير (أو الثقة) للوسط الحسابي للمجتمع تأخذ الشكل التالي :

تقدير متوسط المجتمع = الوسط الحسابي للعينة ◈الخطأ المعياري للوسط

وبالرموز فإن : ④👎 $\overline{X}$ ◈1⑩² $\overline{x}$

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 بدلا من الانحراف المعياري للمجتمع في الثاني، وفيما عدا ذلك فالتوزيعان متماثلان وكلما زادت قيمة n كلما اقترب توزيع t من توزيع z ويعتمد توزيع t على ما يعرف بدرجات الحرية .

<u>**درجات الحرية :**</u>

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.

<u>**شروط توزيع t :**</u>

ويمكن تحديد الشروط الثلاثة لاستخدام توزيع t كما يلي:

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

فترة تقدير النسبة للمجتمع (فترة الثقة للنسبة) :

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

إختبار الفروض الإحصائية

هي عملية وضع قواعد تمكن من التوصل إلي قرار بقبول أو رفض خاصية أو بالمعني الإحصائي فرض عن معالم مجتمع واحد أو أكثر.

⑧ وبصفه عامه فان قبول او رفض اى قرار يجب ان يمر بعدة مراحل وهى:

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

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

الاستدلال الاحصائى يتم باستخدام عينة عشوائية مسحوبه من المجتمع وذلك لاستحالة التعامل مع المجتمع ككل.

اختبارات الفروض ترتكز اساسا على فكرة <u>ان هناك عينه اخرى غير مسحوبه من المجتمع المسحوب منه العينات فان الفرق بين الوسط الحسابى المحسوب من هذه العينه وبين المعلمه المجهوله قد يكون فرقا معنويا Significant غير راجع للصدفه</u>. وبذلك سميت هذه الاختبارات باسم اختبارات المعنويه Test of Significant

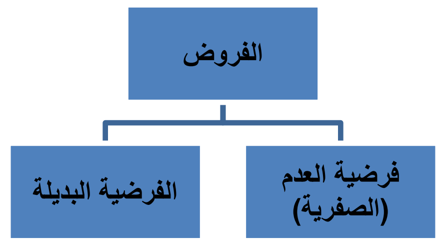
مفهوم الاختبارات الإحصائية المعلمية واللامعلمية :

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

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

<u>الفروض الإحصائية</u>

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



<u>الفرضية الصفرية (فرضية العدم) (Null Hypothesis) H_O :</u>

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

<u>الفرضية البديلة Alternative Hypothesis (H_a) :</u>

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

مستوى الدلالة الإحصائية ومنطقة الرفض :

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

اختبار الفرضيات من طرف واحد:

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

اختبار الفرضيات من طرفين:

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

وفي اختبار الفروض يمكن أن نرتكب نوعين من الخطأ:

<u>الخطأ من النوع الأول Type I error</u>: الخطأ من النوع الأول هو "رفض الفرض العدمي بينما هو صحيح ". أي أنه على الرغم من أن الفرض العدمي في الواقع صحيح وكان من الواجب قبوله فقد تم أخذ قرار خاطئ برفضه.

وباختصار شديد فإن **الخطأ من النوع الأول هو : " رفض فرض صحيح" ويرمز له بالرمز ⇨ .**

<u>الخطأ من النوع الثاني Type II error</u>: 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 ". ويرمز له بالرمز ⇦ .**

بمعنى :

1. فرضية صحيحة نتائج العينة تؤيد صحتها. (قبول صواب)
2. فرضية صحيحة نتائج العينة غير مؤيدة لصحتها. (رفض خطأ) وهذا يعطينا خطأ من النوع الأول ألفا (α) ويمكن أن يقلل برفع مستوى الدلالة .
3. فرضية خاطئة نتائج تؤيد صحتها (قبول خطأ) وهذا يعطينا خطأ من النوع الثاني بيتا (B) ويمكن أن يقلل بزيادة حجم العينة .
4. فرضية خاطئة نتائج غير مؤيدة صحتها (رفض صواب) .

<u>مستوى المعنوية :</u>

يعتبر مصطلح **"مستوى المعنوية"** واحدا من أهم المصطلحات المستخدمة في دراسة نظرية اختبارات الفروض. 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 قيما أخرى.

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

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

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

وتمسى ⇨ مستوى المعنوية، و ⇨− 1 مستوى الثقة للاختبار.

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

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

هناك ثلاث حالات مختلفة لمنطقتي القبول والرفض وهي :

الأولى: إذا كان الفرض البديل يأخذ شكل **"لا يساوي"** تكون موزعة على طرفي المنحنى بالتساوي، ويسمى الاختبار في هذه الحالة "اختبار الطرفين" .

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.

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.

خطوات إجراء الاختبار الإحصائي :

1) **صياغة الفرض الصفري** H_o **:** والذي يأخذ – عادة – شكل " يساوي"

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

"لا يساوي " **أو** "أكبر من" **أو** "أقل من"

3) **إحصائية الاختبار:** وهي الإحصائية التي يتم حسابها من بيانات العينة بافتراض أن الفرض العدمي صحيح.

ويتوقف شكل الإحصائية على العوامل التالية:

أ– توزيع المجتمع، وهل هو طبيعي أم لا، وهل تباينه معروف أم لا.

ب– حجم العينة، وهل هو كبير أم صغير.

جـ– الفرض العدمي المراد اختباره وهل هو عن الوسط أو النسبة أو التباين أو الارتباط... الخ.

4) **تحديد منطقتي القبول والرفض :**

وذلك بناءً على الجداول الإحصائية والتي تعتمد على:

أ- توزيع المعاينة (وهل هو طبيعي أو t أو...)

ب- والفرض البديل (وهل هو لا يساوي أو أكبر من أو أقل من أي هل يستخدم اختبار الطرفين أو الطرف الأيمن أو الأيسر).

ج- ومستوى المعنوية (وهل هو %1 أو %5 أو غير ذلك).

5) المقارنة والقرار :

أي أنه توجد طريقتين لاتخاذ القرار فى الاختبارات الاحصائية وهي:

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

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

اختبار الفروض الإحصائية المعلمية

المختبر الإحصائي هو متغير عشوائي ذو توزيع احتمالي يصف العلاقة بين القيم النظرية للمعلم والقيم المحسوبة من العينة.

أوجه الشبه والإختلاف بين توزيع t والتوزيع المعتدل المعياري:

توزيع t يعتمد على افتراض أن العينة مسحوبة من مجتمع توزيعه معتدل أو قريب من الاعتدال، لذلك يستخدم اختبار t لاختبار الفروض وإيجاد حدود الثقة لمتوسط المجتمع أو الفرق بين متوسطين عندما تكون إحجام العينات صغيرة وتباينات المجتمعات مجهولة.

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

انواع الاختبار (الفروض) فى حالة عينه واحده:

بفرض اننا سوف نرمز للمعلم المجهول بالرمز μ ونريد اختبار الفرض القائل ان قيمته تساوى μ_0 سيكون فرض العدم على الصورة التالية:

$$H_0 : \mu = \mu_0$$

وسيكون الفرض البديل فى حالة الاختبار ذو طرف واحد

$$H_a : \mu > \mu_0 \quad or \quad H_a : \mu < \mu_0$$

وسيكون الفرض البديل فى حالة الاختبار ذو طرفين

$$H_a : \mu \neq \mu_0$$

<u>انواع الاختبار (الفروض) في حالة عينتين:</u>

بفرض اننا سوف نرمز للمعلم المجهول بالرمز μ ونريد اختبار الفرض القائل ان قيمته متساويه فى المجتمعين

سيكون فرض العدم على الصورة التالية: $H_0 : \mu_1 = \mu_2$

وسيكون الفرض البديل فى حالة الاختبار ذو طرف واحد

$$H_a : \mu_1 > \mu_2 \;\; or \;\; H_a : \mu_1 < \mu_2$$

وسيكون الفرض البديل فى حالة الاختبار ذو طرفين

$$H_a : \mu_1 \neq \mu_2$$

بفرض اننا سوف نرمز للمعلم المجهول بالرمز μ ونريد اختبار الفرض القائل ان قيمته متساويه فى المجتمعات التى عددها r سيكون فرض العدم على الصورة التالية

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_r$$

وسيكون الفرض البديل H_a : at least two are different

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

<u>الاختبار(ت) لعينة واحدة :</u>

<u>اختبار t-test</u> : يستخدم هذا الصنف من اختبار (ت) للحكم على معنوية الفروق بين متوسط عينة ومتوسط المجتمع الذي سحبت منه.

<u>الاختبارات الاحصائية لعينتين مستقلتين:</u>

يستخدم هذا النوع من اختبار (ت) للحكم على **<u>معنوية الفروق</u>** بين متوسطي عينتين غير مرتبطتين.

ويستند هذا الاختبار إلى توفر عدد من الافتراضات وهذه الافتراضات هي :

- **مستوى القياس:** يشترط لاستخدام هذا الاختبار أن تكون البيانات فئوية (فترية) أو نسبية.
- **أن يكون حجم العينة صغيراً:** يقتضي هذا الافتراض أن يكون حجم العينة أقل من (30) وأكبر من (5) .
- **التوزيع الطبيعي.**

- **تجانس التباين في المجتمعين.**

- عندما تكون العينة الكبيرة الحجم منتمية للمجتمع ذي التباين الكبير، والعينة صغيرة الحجم منتمية للمجتمع ذي التباين الصغير، فإن احتمال ارتكاب الخطأ من النوع الأول تكون أقل من (α).
- أما عندما تكون العينة كبيرة الحجم منتمية للمجتمع ذي التباين الصغير والعينة صغيرة الحجم منتمية للمجتمع ذي التباين الكبير، فإن احتمال ارتكاب الخطأ من النوع الأول يزداد ولتفادي ذلك نستخدم اختبار ولتيش _ Weltch.

الاستقلالية: ويقتضي هذا الافتراض أن (n_1) من المشاهدات قد تم الحصول عليها عشوائيا من المجتمع الأول بشكل مستقل عن (n_2) من المشاهدات والتي تم الحصول عليها عشوائياً من المجتمع الثاني. وإن الاستقلالية هنا لا تعني استقلالية البيانات بين المجتمعين فقط، بل تعني استقلالية المشاهدات ضمن المجتمع الواحد.

<u>الاختبارات الاحصائية لعينتين غير مستقلتين (العينات المرتبطة):</u>

يستخدم هذا النوع للحكم على **<u>دلالة الفروق ومعنويتها</u>** بين متوسطي عينتين مرتبطتين .

- **<u>إختبار الفروض الإحصائية المعلمية</u>**

الاختبارات الاحصائية لأكثر من عينتين مستقلتين:

سمي توزيع F بهذا الاسم تخليدا للعالم رونالد فيشر R.A. Fisher الذي يعتبر أول من اشتق هذا التوزيع ووصفه وذلك في العشرينات والثلاثينات من القرن العشرين لذلك تعرف أحيانا بتحليل فيشر للتباين.

ويستخدم توزيع F أساسا لاختبار تساوي تبايني مجتمعين، ومن المثير للانتباه ملاحظة أن اختبار تساوي التباينين يستخدم لاختبار تساوي ثلاث متوسطات أو أكثر.

وتسمى طريقة الاستدلال الإحصائي عن تساوي ثلاث متوسطات أو أكثر بتحليل التباين Analysis of Variance .

وتوزيع F عبارة عن مجموعة من المنحنيات التكرارية يتميز كل منها عن الآخر برقمين لدرجات الحرية أحدهما يمثل درجة حرية للبسط والآخر درجة حرية للمقام .

وقيمة F هي قيمة توضح نسبة التباين Variance ratio لعينتين والرمز F إشارة إلي العالم Fisher الذي قام بعمل هذا الاختبار والمعروف باختبار F وقد قام العالم Snedecor بحساب جداول خاصة لتوزيع F

وفيها درجات الحرية التي في أعلى الجدول تخص البسط أما درجات الحرية على العمود الجانبي فتخص المقام.

وتوزيع F هو توزيع ملتو جهة اليمين بمعلمتين تتمثلان بدرجتي حرية (البسط ، المقام) وهما $k-1$ للبسط ، $n-k$ للمقام حيث n مجموع أحجام العينات و k هي عدد المجموعات موضع الدراسة.

أهم استخدامات توزيع F (ف) هي :

- تقدير فترة الثقة لـ σ_1^2/σ_2^2
- اختبار فرضيات حول تساوي تباينين أي : $H_0 : \sigma_1^2 = \sigma_2^2$
- اختبار فرضيات حول تساوي أكثر من متوسطين أي : $H_0 : \mu_1 = \mu_2 = \mu_3 = ...$

<u>تحليل التباين الأحادي:</u>

تحليل التباين هو عملية يقصد بها تقسيم مجموع مربعات الانحرافات عن المتوسط الحسابي إلي مكوناته إرجاع كل من هذه المكونات إلي مسبباتها . وطريقة تحليل التباين تفيد في مقارنة عدد من المعاملات يزيد عن أثنين كما تمتاز طريقة تحليل التباين بأنه يمكن فيها استعمال كل البيانات المأخوذة من التجربة في حساب قيمة واحدة للانحراف القياسي يمكن بها مقارنة المجموعات أو المعاملات التجريبية.

فهي مجموعة من النماذج الإحصائية (statistical model) مع إجرءات مرافقة لهذه النماذج تمكن من مقارنة المتوسطات لمجتمعات إحصائية مختلفة عن طريق تقسيم التباين variance الكلي الملاحظ بينهم إلى أجزاء مختلفة.

تتلخص طريقة تحليل التباين في:

1- حساب المجموع الكلي لمربعات انحرافات كل المفردات في التجربة عن المتوسط العام .

2- تقسيم هذا المجموع الكلي لمربعات الانحرافات Total Sum Squares الى مكوناته طبقا للمصادر المسببة لها والذي يختلف عددها طبقا للتصميم المستعمل في التجربة.

3- تقسم درجات الحرية الكلية طبقا للمصادر السابقة أيضا.

4- تدون النتائج في جدول يسمي جدول تحليل التباين ANOVA ترتب فيه مصادر الاختلافات حسب التصميم الإحصائي المستعمل ويسهل هذا الجدول عمل اختبار معنوية المعاملات.

<u>سبب الحاجة لاختبار تحليل التباين:</u>

- كلما ازداد عدد المتوسطات كلما ازداد عدد الأزواج التي يجب أن نجري عليها الاختبار ، وقد يصبح العدد كبيراً جداً .

- عند المقارنة بين كل زوج من الأوساط فإننا نستخدم فقط المعلومات عن المجموعتين المقارنتين ونهمل المعلومات المتوفرة في باقي المجموعات والتي تجعل المقارنة أقوى فيما لو استعملت .
- إن هذا الاستخدام المتعدد لاختبار (t) مثلاً يزيد من الخطر في ارتكاب الخطأ من النوع الأول (رفض الفرضية الصفرية عندما تكون صحيحة) .
- إن الباحث كثيرا ما يرغب في اختبار مشكلة أوسع من مجرد معرفة ما إذا كانت أزواج المتوسطات مختلفة عن بعضها البعض .

تحليل التباين الأحادي (مستوى واحد) :

هو طريقة لاختبار معنوية الفرق بين المتوسطات لعدة عينات بمقارنة واحدة، ويعرف أيضاً بطريقة تؤدي لتقسيم الاختلافات الكلية لمجموعة من المشاهدات التجريبية لعدة أجزاء للتعرف على مصدر الاختلاف بينها ولذا فالهدف هنا فحص تباين المجتمع لمعرفة مدى تساوى متوسطات المجتمع ولكن لا بد من تحقيق ثلاثة أمور قبل استخدامه وهي:

1) العينات عشوائية ومستقلة.

2) مجتمعات هذه العينات كلً لها توزيع طبيعي.

3) تساوي تباين المجتمعات التي أخذت منها العينات العشوائية المستقلة.

الافتراضات الاساسية لاختبار تحليل التباين:

يستند اختبار تحليل التباين إلى توفر عدد من الافتراضات، ومن هذه الافتراضات ما يلي :

1- مستوى القياس :

يشترط لاستخدام هذا الاختبار أن تكون البيانات فترية (فئوية) أو نسبية

2- حجم العينة :

يقتضي هذا الافتراض أن يكون حجم العينة كبيراً .

3- التوزيع الطبيعي للمجتمع الإحصائي :

يقتضي هذا الافتراض أن تكون المشاهدات في كل مجتمع من المجتمعات موزعة بشكل طبيعي، ولكن يرى الإحصائيون أن اختبار (F) لا يتأثر كثيراً بعدم توفر هذا الشرط وذلك عندما يكون حجم الخلية كبيراً والتوزيع ليس طبيعياً.

4- تجانس التباين :

أي أن يكون للمجتمعات في مستويات المعالجة المختلفة نفس التباين.

ولغرض حساب تحليل التباين الأحادي، **علينا اتباع الخطوات التالية:**

- نجمع قيم كل متغير للحصول على $\sum X$ لكل متغير .
- نربع كل درجة في كل متغير للحصول على X^2 لكل متغير .
- نجمع قيم مربع كل درجة للحصول على $\sum X^2$ لكل متغير .
- نربع مجموع كل متغير للحصول على $(\sum X)^2$ لكل متغير .
- نحسب متوسط كل متغير من خلال العلاقة : $\overline{X} = \frac{\sum x}{n}$
- نحسب مجموع المربعات الكلي Total Sum of Squares وذلك من خلال العلاقة التالية :

$$Total..SS = \sum X^2 - \frac{(\sum X)^2}{n}$$

حيث n تعني مجموع أعدد الأفراد في جميع المجموعات .

أو يمكن حساب مجموع المربعات الكلي من خلال العلاقة التالية :

Total SS = Between SS + Within SS

- نحسب مجموع المربعات بين المجموعات Between Sum of Squares وذلك من خلال العلاقة التالية :

$$Between..SS = \sum \frac{(\sum X_g)^2}{n_g} - \frac{(\sum X)^2}{n}$$

حيث n_g تعني عدد الأفراد في كل مجموعة .

n تعني مجموع أعدد الأفراد في جميع المجموعات .

$\sum \frac{(\sum X_g)^2}{n_g}$ تعني مربع مجموع قيم كل مجموعة مقسوما على عدد أفراد تلك المجموعة .

$\frac{(\sum X)^2}{n}$ تعني مربع مجموع قيم كل المجموعات مقسوما على مجموع أعداد الأفراد في جميع المجموعات .

- نحسب مجموع المربعات داخل المجموعات Within Sum of Squares وذلك من خلال العلاقة التالية

$$Within..SS = \sum \left[\sum X_g^2 - \frac{(\sum X_g)^2}{n_g} \right]$$

أو يمكن حساب مجموع المربعات داخل المجموعات من خلال العلاقة التالية :

Within SS = Total SS - Between SS

- نحسب درجات الحرية بين المجموعات Between groups degrees of freedom من خلال العلاقة التالية :

$K - 1$ حيث K تعني عدد المجموعات .

ودرجات الحرية داخل المجموعات Within groups degrees of freedom من خلال العلاقة التالية

$n - K$ حيث n تعني عدد الأفراد أو الاستجابات في المجموعات موضع الدراسة ، وK تعني عدد المجموعات.

ودرجات الحرية الكلية Total degrees of freedom من خلال العلاقة التالية :

$n - 1$ حيث n تعني عدد الأفراد أو الاستجابات في المجموعات موضع الدراسة .

- نحسب التباين بين المجموعات أو ما يسمى متوسط المربعات بين المجموعات Between mean square وذلك من خلال العلاقة التالية:

$$Beween..groups..mean..square = \frac{Between..SS}{K - 1}$$

- نحسب التباين داخل المجموعات أو ما يسمى متوسط المربعات داخل المجموعات Within mean square وذلك من خلال العلاقة التالية :

$$Within..groups..mean..square = \frac{Within..SS}{(n - K)}$$

- نحسب قيمة F من خلال العلاقة التالية :

$$F = \frac{Between..groups..mean..square}{Within..groups..mean..square}$$

- نقارن بعد ذلك قيمة F المحسوبة بقيمة F المجدولة لاتخاذ القرار المناسب اتجاه الفرضية موضع الدراسة

عند الرغبة في حساب المقارنات البعدية بين متوسطات المتغيرات موضع الدراسة والكشف عن مواقع الفروق وذلك في حالة كون قيمة F ذات دلالة إحصائية، وهناك العديد من الأساليب الإحصائية المعدة لهذا الغرض اشهرها :

1- **طريقة شيفيه Scheffe** وتستخدم هذه الطريقة في إجراء جميع المقارنات بين الأوساط وهي الطريقة المفضلة في حالة كون حجوم الخلايا غير متساوية أو عند الرغبة في إجراء مقارنات معقدة كأن نقارن ثلاث مجتمعات بمجتمع واحد، أو مجتمعين مقابل مجتمعين أو غيرها من مثل هذه المقارنات .

2- **طريقة توكي Tukey** وتستخدم هذه الطريقة لمقارنة جميع الأزواج الممكنة للأوساط موضع الدراسة سواء كانت حجوم الخلايا متساوية أو غير متساوية (في حالة عدم تساوي حجوم الخلايا يستخدم الوسط التوافقي لحجم الخلية) ، ويعتبر هذا الاختبار أدق من اختبار شيفيه Scheffe لمقارنة أزواج الأوساط.

3- **طريقة نيومن–كولز Newman-Keuls (S-N-K)** وتفيد هذه الطريقة في المقارنة بين أزواج الأوساط فقط، وهي تستند كما هي الحال في طريقة توكي على توزيع مدى ستيودنتايز Studentize range ، وهي طريقة جيدة وقوية للكشف عن الفروق بين الأوساط في حالة تساوي حجوم الخلايا أو عدم تساويها (في حالة عدم تساوي حجوم الخلايا يستخدم الوسط التوافقي لحجم الخلية كما هو الحال في اختبار توكي)، ويعتبر هذا الاختبار (نيومن–كولز) أدق الاختبارات البعدية للكشف عن الفروق بين أزواج الأوساط، يليه اختبار توكي ثم بعد ذلك اختبار شفيه

حساب المقارنات المتعددة Multiple Comparisons

هي طريقة لإجراء عدد من الاختبارات الأولية لتحديد الفروق المعنوية للمتوسطات حال رفض فرضية العدم.

حال رفض فرضية العدم عند مستوى معنوية فلا دليل على وجود فروق معنوية بين المتوسطات ولا نعرف أي منها يختلف عن متوسطات (العينات) ومن حيث قيمة F معنوية ففرق المشاهدات بين المتوسطات معنوي أيضاً وعلى العموم توجد عدة طرق إحصائية لعمل اختبار بهذا الخصوص مثل:

- ★ Least Significant difference,
- ★ Scheffe Test,
- ★ Student-Newman Keuls,
- ★ Tukey's Procdure,
- ★ Duncans New Multiple range

مقاييس العلاقة نعني بذلك تلك المقاييس التي تبين درجة العلاقة والارتباط بين متغيرين أو أكثر.

معامل الارتباط: هو تعبير يشير إلى المقياس الإحصائي الذي يدل على مقدار العلاقة بين المتغيرات سلبية كانت أم إيجابية، وتتراوح قيمته بين الارتباط الموجب التام (+1) وبين الارتباط السالب التام (−1) .

العلاقة الطردية بين المتغيرات: هو تعبير يشير إلى تزايد المتغيرين المستقل والتابع معا، فإذا كانت الإنتاجية مرتفعة، ومستوى الجودة مرتفع، يقال حينئذ أن بينهما ارتباط موجب، وأعلى درجة تمثله هي (+1) .

العلاقة العكسية بين المتغيرات: هو تعبير يشير إلى تزايد في متغير يقابله تناقص في المتغير الآخر، فإذا كانت الإنتاجية منخفضة ومستوى الجودة مرتفع، يقال حينئذ أن بينهما ارتباط سالب، وأعلى درجة تمثله هي (−1) .

طرق التعرف على العلاقة بين متغيرين وحسابها:

أولا: طريقة شكل الانتشار Scatter Diagram :

هناك وسيلة مبدئية يعرف الباحث من خلالها نوع الارتباط بين المتغيرين وما إذا كان الارتباط قوياً وضعيفاً أو منعدماً، وما إذا كانت العلاقة خطية أو غير خطية، موجبة أو سالبة. هذه الوسيلة هي " شكل الانتشار " والتي تصلح إذا كان المتغيران كميين. وجدير بالذكر أن هذه وسيلة مبدئية تساعد فقط في معرفة نوع الارتباط ولا تعتبر بديلاً عن الطرق الإحصائية التي سوف نتناولها بالتفصيل في هذه المحاضرة.

ثانيا: معامل الارتباط Correlation Coefficient :

يقاس الارتباط بين متغرين بمقياس إحصائي يسمى " معامل الارتباط " ويعكس هذا المقياس درجة أو قوة العلاقة بين المتغيرين واتجاه هذه العلاقة. وتنحصر قيمة معامل الارتباط بين + 1، − 1.

- فإذا كانت قيمة معامل الارتباط تساوي + 1 فمعنى ذلك أن الارتباط بين المتغيرين طردي تام، وهو أقوى أنواع الارتباط الطردي بين متغيرين.

- وإذا كانت قيمة معامل الارتباط تساوي – 1 فمعنى ذلك أن الارتباط بين المتغيرين عكسي تام، وهو أقوى أنواع الارتباط العكسي بين متغيرين.
- وإذا كانت قيمة معامل الارتباط تساوي صفر، فمعنى ذلك أنه لا يوجد ارتباط بين المتغيرين.
- وكلما اقتربت قيمة معامل الارتباط من + 1 أو – 1 كلما كان الارتباط قوياً، وكلما اقترب من الصفر كلما كان الارتباط ضعيفاً.

والخلاصة :

أنه كلما كانت العلاقة قوية بين المتغيرين كلما اقترب معامل الارتباط من + 1 أو – 1 فإذا وصلت قيمة المعامل إلى + 1 أو – 1 كان الارتباط تاماً بين المتغيرين. وأنه كلما كانت العلاقة ضعيفة بين المتغيرين كلما اقترب معامل الارتباط من الصفر، فإذا وصلت قيمة المعامل إلى الصفر كان الارتباط منعدماً بين المتغيرين. ومعنى ذلك أيضاً أنه لا يوجد ارتباط بين متغيرين تكون قيمة المعامل فيه أكبر من + 1 ولا أصغر من – 1.

معامل بيرسون للارتباط الخطي البسيط :

يفترض بيرسون Pearson أن المتغيرين كميان، وأن العلاقة بينهما خطية (أي تأخذ شكل خط مستقيم، ويرى بيرسون أن أفضل مقياس للارتباط بين متغيرين قد يختلفان في وحدات القياس و / أو في مستواهما العام هو عن طريق حساب انحرافات كل من المتغيرين عن وسطه الحسابي وقسمة هذه الانحرافات على الانحراف المعياري لكل منهما، فنحصل على ما يسمى بالوحدات المعيارية لكل متغير. ويكون معامل ارتباط بيرسون هو " متوسط حاصل ضرب هذه الوحدات المعيارية ". ومعامل الارتباط يكون بدون تمييز.

إن الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين .

ويتم حساب معامل ارتباط بيرسون من خلال العلاقة التالية:

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{n}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{n}\right]}}$$

حيث :

$\sum XY$ تعني مجموع حاصل ضرب كل قيمة من X في Y.

$(\sum X)$ تعني مجموع قيم المتغير X .

$(\sum Y)$ تعني مجموع قيم المتغير Y .

$\sum X^2$ تعني مجموع مربع قيم المتغير X .

$(\sum X)^2$ تعني مربع مجموع قيم المتغير X .

$\sum Y^2$ تعني مجموع مربع قيم المتغير Y .

$(\sum Y)^2$ تعني مربع مجموع قيم المتغير Y .

n عدد قيم الدراسة (عدد الأزواج المطلوب حساب الارتباط بينها) .

لحساب معامل ارتباط الرتب هناك طرق مختلفة أهمها معاملي سبيرمان وكيندال.

لحساب معامل سبيرمان لارتباط الرتب نقوم بترتيب كل من المتغيرين ترتيباً تصاعدياً أو تنازلياً (أما تصاعدياً لكلا المتغيرين أو تنازلياً لكليهما).

الإرتباط الجزئي :

هو عبارة عن مقياس لقوة واتجاه الارتباط بين متغيرين كميين بعد استبعاد اثر متغير كمي ثالث .

اختبار معنوية معامل الارتباط:

إذا كانت قيمة معامل ارتباط العينة r قريبة من + 1 أو – 1 فإن هناك علاقة خطية قوية بين المتغيرين، وإذا كانت r = 0 فإنه لا توجد علاقة خطية بينهما، أما إذا كانت قيم r متوسطة فإنه يجب اختبار معنوية (أو دلالة) معامل ارتباط العينة، وهل هناك ارتباط حقيقي بين المتغيرين في المجتمع، أم أن الارتباط بينهما زائف وغير حقيقي.

الطرق الإحصائية اللامعلمية :

تتطلب معظم التحليلات تحديد بعض الافتراضات أو الشروط حول المجتمع أو المجتمعات التي اختيرت منها العينة أو العينات. ففي كثير من الحالات يتم افتراض أن المجتمعات موضع الدراسة تتصف بالتالي:

- افتراض أن المجتمعات موضع الدراسة تتبع توزيعاً طبيعياً
- افتراض أن تباينات هذه المجتمعات معلومة
- افتراض أن تباينات هذه المجتمعات غير معلومة ولكنها متساوية
- افتراض أن العينات المختارة مستقلة

وحيث أنه توجد مواقف أو حالات كثيرة يكون من الصعب التأكد من تحقق هذه الافتراضات، أو يكون هناك شك في تحققها، وحيث أننا نواجه في كثير من الأحيان بيانات واقعية يصعب فيها التعرف على صيغة التوزيع الاحتمالي الذي تتبعه. لذلك فقد طور الإحصائيين أساليباً وطرقاً إحصائية بديلة وهذه الطرق تتصف بالتالي:

- لا تتطلب افتراضات كثيرة .
- لا تتطلب معرفة صيغة التوزيع الاحتمالي للمجتمعات التي تختار منها العينات.

وهذه الطرق بالإضافة إلى أنه يمكن استخدامها تحت شروط وافتراضات عامة فإنها غالباً لا تحتاج إلى مجهود في العمليات الحسابية .كما انه يمكن التعامل معها لمتغيرات منفصلة ومتغيرات متصلة على السواء. ولهذه الأسباب أصبحت الطرق اللامعلمية مرغوبة بكثرة .

الفرق بين الطرق الإحصائية المعلمية والطرق الإحصائية غير المعلمية:

الطرق الإحصائية المعلمية	الطرق الإحصائية اللامعلمية
يتطلب استخدام الطرق المعلمية الوفاء بافتراضات معينة حول التوزيع الأساسي للمجتمع، من حيث أن يكون توزيع المجتمع طبيعياً	لا يتطلب استخدام الطرق اللامعلمية اي افتراضات أو معلومات حول خصائص التوزيع الأساسي للمجتمع (لذلك يطلق بعض الإحصائيين عليها إحصاءات التوزيعات الحرة)
لا يمكن استخدام الطرق الإحصائية المعلمية لمعالجة وتحليل البيانات في المواقف التجريبية التي يكون فيها حجم العينة صغير جداً، لأن الحجم الصغير للعينة يؤثر على خصائص التوزيع التكراري للعينة الصغيرة، فتبتعد بذلك عن اعتدالية التوزيع التكراري للمجتمع الأب	يمكن استخدام بعض الطرق اللامعلمية لمعالجة وتحليل البيانات في المواقف التجريبية التي يكون فيها حجم العينة صغير جداً
تكون الطرق المعلمية أكثر ملائمة لتحليل البيانات الفترية والبيانات النسبيـة فضلاً عن أنها يمكن استخدامها (مع بعض التحفظ) لتحليل يانات الرتبية فقط وذلك بعد أن يتم تحويلها إلى بيانات فترية من خلال إعطاء كل رتبة درجات تتناسب مع قيمتها، مع مراعاة أن تكون البيانات في صورة رتب ذات درجات متتابعة مثل: موافق بشدة، موافق، غير موافق، غير موافق بشدة	تكون الطرق اللامعلمية عادة أكثر ملاءمة للاستخدام عندما تكون البيانات الخاصة بالبحث من النوعين الاسمي والرتبي فضلاً عن أنها يمكن استخدامها أحياناً عندما تكون البيانات فترية أو نسبية، وذلك بعد أن يتم تحويلها الى بيانات اسمية او رتبية (وهذا فيه إهدار لجزء كبير من البيانات التي نحللها، لان الطرق الإحصائية اللامعلمية لا تستخدم الأجزاء يسيراً من تلك البيانات) لذلك أوضح بعض الإحصائيين انه عندما نستخدم الطرق الإحصائية اللامعلمية للبيانات الفترية والنسبية فإنها لا تستخدم إلا للتقدير المبدئي، على أن يتلوه بعد ذلك استخدام الطرق المعلمية
تعتبر الطرق الإحصائية المعلمية بشكل عام أقوى من الطرق الإحصائية اللامعلمية، حيث أن الطرق المعلمية تميل إلى رفض الفرضية الصفرية اكثر	بالرغم من تحرر الطرق الإحصائية اللامعلمية من الشروط والخصائص التي قد تعوق أحيانا استخدام الطرق الإحصائية المعلمية، إلا أن الطرق

من ميل الطرق اللامعلمية لرفض نفس الفرضية	الإحصائية اللامعلمية بشكل عام أقل قوة من الطرق الإحصائية المعلمية
تعتمد الطرق الإحصائية المعلمية بشكل عام على الدرجات الأصلية والتي يتم تحليلها كما هي	تعتمد الطرق اللامعلمية في اغلب الأحيان على البيانات التي هي بشكل تكرارات أو رتب مما يؤدي إلى ضياع بعض المعلومات المفيدة
تعتبر الطرق الإحصائية المعلمية بصورة عامة أصعب في الاستخدام من الطرق الإحصائية اللامعلمية، وبالتالي فإن الوقت الذي يحتاجه الباحث لتحليل بياناته يكون أطول، مما يؤدي إلى تأخر الحصول على النتائج (نسبياً) والإفادة منها تطبيقياً	تعتبر الطرق الإحصائية اللامعلمية أفضل وسيلة للتقدير المبدئي السريع، فهي بصورة عامة اسهل استخداماً من الطرق المعلمية وبالتالي فإن الوقت الذي يحتاجه الباحث لتحليل بياناته يكون اقل، مما يؤدي إلى الإسراع في الحصول على النتائج والإفادة منها تطبيقياً
تعتبر الطرق الإحصائية المعلمية أكثر شيوعاً واستخداماً لدى المختصين في الإحصاء	تعتبر الطرق الإحصائية اللامعلمية أكثر شيوعاً واستخداماً لدي غير المتخصصين في الإحصاء
ينبغي أن يكون اختيار العينة من المجتمع بصورة عشوائية، وأن تكون إحصاءات العينة (مقاييس النزعة المركزية والتشتت) صورة مقربة للمعلمات الإحصائية للمجتمع	ليس هناك ضرورة أن يكون اختيار العينة من المجتمع بصورة عشوائية
تستخدم الطرق الإحصائية المعلمية لمعالجة وتحليل البيانات الكمية	تستخدم الطرق الإحصائية اللامعلمية لمعالجة وتحليل البيانات النوعية، و التي لا يمكن عادة استخدام أي طريقة إحصائية معلمية لتحليلها
من الأمثلة على الطرق الإحصائية المعلمية اختبار (Z) واختبار (t) واختبار (F) ... الخ.	من الأمثلة على الطرق الإحصائية اللامعلمية اختبار (كا^2) واختبار مان ويتني واختبار كروسكال

إن الميزة السيئة للاختبارات اللابارامترية هي أنها غير جيدة عادة لإيجاد الفروقات عندما يكون هناك فروقات في المجتمع .

ماذا علينا أن نفعل عندما لا نكون متأكدين من اختيار الاختبار المناسب: بارامتري–لابارامتري؟

عندما تكون في حيرة وشك استخدم كليهما. إذا حصلت على نتيجة واحدة فلا شيء يدعو للقلق، أما إذا كانت النتائج التي حصلت عليها من الاختبار اللابارامتري تفتقد إلى مدلول بينما تلك التي حصلت عليها باستخدام الاختبار البارامتري أكثر دلالة حاول أن تظهر السبب، هل يوجد بين البيانات قيم متطرفة (أصغر أو أكبر من بقية القيم بشكل ملحوظ)، إذا كانت تلك

الحالة موجودة فهذا يعني أن تلك القيم يمكن أن تؤثر على المتوسط وأن لها تأثير أكبر على نتائجك.

اختبار مان وتني Mann – Whitney U :

استخدامه:

يعتبر هذا الاختبار بديل لا معلمي للاختبار الخاص بالفرق بين متوسطي مجتمعين والمبني على أساس عينتين مستقلتين أي أن هذا الاختبار بديل لاختبار t لعينتين مستقلتين، بل أنه أفضل منه خاصة إذا كانت العينتان مختارتين من مجتمعين لا يتبعان توزيعاً طبيعياً.

ويعد هذا الاختبار أكثر الاختبارات اللابارامترية استخداماً فى البحوث عندما يكون المتغير التابع من المستوى الرتبى بدلاً من الدرجات الأصلية، كما يمكن استخدام هذا الاختبار إذا كانت المتغيرات من المستوى الفترى أو النسبى ولكنها لا تفى بشروط اختبار النسبة التائية مثل عدم إعتدالية التوزيع أو اختلاف التباين بين المجموعتين اختلافاً كبيراً.

إختبار ويلكوكسون Wilcoxon Test :

استخدامه:

ويسمى باختبار اشارات الرتب Sign –rank، ويستخدم هذا الاختبار فى تحديد ما إذا كان هناك اختلاف أو فروق بين عينتين مرتبطتين فيما يتعلق بمتغير تابع معين، ويعد بديلاً لابارامترياً لاختبار T لعينيتين مرتبطين، وتشتمل العينتان على نفس المجموعة من الأفراد يجرى عليهم قياس قبلى Pre test، وقياس بعدى Post test وفى مثل هذه الحالة يكون لكل فرد من أفراد العينة درجتان أحداهما تمثل درجته فى الاختبار القبلى والثانية تمثل درجته فى الاختبار البعدى. ويستخدم مع البيانات العددية فقط دون الاسمية .

حتى نحسب اختبار ويلكوكسن يجب اولا أن نجد الفرق بين القيمتين من أجل كل زوج ومن ثم من أجل كافة الحالات التي يكون عندها الفرق غير معدوم، نرتب الفروقات بشكل تصاعدي متجاهلين إشارة الفروقات، ذلك يعني بأن نسند إلى الفرق الصغير في القيمة المطلقة الرتبة 1 ونسند إلى الفرق الصغير التالي الرتبة 2 وهكذا، أما في حالة الفروقات المتساوية (الحالات المتعادلة) نسند رتبة المتوسط إلى تلك الحالات.

اختبار كروسكال واليس Kruskal-Wallis Test :

استخدامه:

يعتبر هذا الاختبار بديلاً لامعلميا لاختبار تحليل التباين في اتجاه واحد، وهو مبني على مجموع الرتب ويستعمل لاختبار الفروق بين ثلاث مجموعات أو أكثر.

اختبارات الفروض باستخدام توزيع كاي تربيع (χ^2)

يعتبر توزيع كاي تربيع χ^2 من التوزيعات الإحتمالية الشائعة الاستخدام حيث توجد له تطبيقات عديدة بدرجة يمكن معها القول أنه يأتي في المرتبة الثانية للتوزيع المعتدل من حيث كثرة تطبيقاته.

توزيع كاي تربيع χ^2 :

يعتمد توزيع χ^2 مثل توزيع t اعتمادا كاملا على درجات الحرية، وعلى الرغم من ذلك يوجد اختلاف رئيس بين التوزيعين حيث نجد أن توزيع t متماثل حول وسطه الحسابي ($\mu=0$)، بينما يعتبر توزيع χ^2 توزيعا ملتويا جهة اليمين (التواء موجب) وخاصة عندما تكون درجات الحرية صغيرة، **وكلما زادت درجات الحرية كلما قل التواء التوزيع واقترب من التماثل.**

استخدامات توزيع χ^2

اختبار تباين المجتمع

يستخدم توزيع χ^2 في إجراء العديد من الاختبارات الإحصائية مثل:

الاختبارات المتعلقة بتباين مجتمع ما (وذلك لاختبار المشاكل التي تتطلب اختبار تشتت مجتمع ما) ، ويتم ذلك من خلال استخدام المعادلة التالية:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

اختبار مربع كاي لجودة التوفيق

ويهتم هذا النوع من الاختبارات الإحصائية باختبار ما إذا كانت مشاهدات عينة تم اختيارها من مجتمع له توزيع احتمالي معين أو نظرية معينة.

ويستخدم هذا الاختبار عندما تكون البيانات اسمية أو على شكل تكرارات ويقصد بجودة التوفيق هنا دراسة مدى تشابه تكرارات العينة والتي تسمى عادة بالتكرارات الملاحظة Observed مع التكرارات المتوقعة Expected للمتغير موضوع الدراسة فى المجتمع الأصلى.

ويستخدم اختبار كا٢ كطريقة إحصائية للمقارنة بين التكرارين الملاحظ والمتوقع. فإذا كانت العينة ممثلة للمجتمع فى تكراراتها ومتطابقة معه فإن قيمة كا٢ تكون عادة صفراً وتزداد هذه القيمة لتصبح أكثر من صفر كلما كان هناك فرق بين تكرارات العينة (الملاحظة) وبين تكرارات التوزيع النظرى للمجتمع (المتوقعة).

<u>اختبار مربع كاي للاستقلالية (الإعتمادية)</u>

كاي تربيع للاستقلالية (Chi-Square test of independency) هو اختبار بسيط يقوم به الباحث لمعرفة ما إذا كان هناك علاقة بين شيئين أو متغيرين. يجرى هذا الاختبار عن طريقة مقارنة قيمة يحددها الباحث مسبقا تعرف بمستوى المعنوية (الفا) بالقيمة المسماة p-Value تحسب من البيانات المتوفرة، حيث سيتضح عن طريق المقارنة بين القيمتين ما إذا كانت هنالك علاقة بين الاثنين أم لا .

<u>اختبار كولومجروف سيمرنوف لجودة التوافيق</u>

استخدامه:

يستخدم هذا الاختبار لمعرفة إذا ما كانت العينة موضع الاهتمام تتبع توزيعاً احتمالياً معينا **<u>ويستخدم عوضاً عن اختبار مربع كاي</u>** عندما يكون مجموع التكرارات أقل من 30 أو يكون التكرار المتوقع لأي خلية أقل من خمسة وعملية ضم الخلايا تؤدي إلى فقد كثير من درجات الحرية مما يتعذر معه إجراء الاختبار أو أن تكون عملية الضم غير مناسبة. ويفضل استخدامه أيضاً في حالة كون التوزيع الاحتمالي لمتغير متصل.

تمت بحمد الله .

مع تمنياتي لكم بالتوفيق .. اخوكم ابو جنى الجوني .