الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي
جامعة العربي التبسي - تبسة

كلية العلوم الدقيقة وعلوم الطبيعة والحياة

قسم الرياضيات والإعلام الآلي

الإحصاء الحيوي

مقياس: الإحصاء الحيوي

اختصاص: بيولوجيا

المستوى: الثانية ليسانس

إعداد: الدكتور لعمايرية عبد الحكيم

2020-2021

الفهرس

الفصل لأول

الإحصاء الوصفي

الفصل الثاني

الإحصاء الإستدلالي

الفصل الأول

الإحصاء الوصفي

1.1 مقدمة:

ان علم الإحصاء يعتبر من اهم الركائز التي ترتكز عليها عملية البحث العلمي في ميادينه المختلفة ويمكن القول انه لا يوجد مجال من مجالات الفكر والعمل إلا واستعمل الإحصاء فيه بأساليبه المختلفة ومن اهم المجالات العلوم الحياتية.

تعريف علم الإحصاء (Statistics):

هو العلم الذي يهتم بجمع البيانات وتصنيف وتبويب وتحليل البيانات واستخلاص النتائج والإستنتاجات منها.

ويقسم علم الاحصاء الى قسمين هما:

1- الإحصاء الوصفي (Descriptive statistic)

يتضمن هذا القسم الطرق والأساليب المستخدمة لجمع البيانات وتصنيفها وتبويبها مع امكانية عرضها في جداول ورسوم بياتية وحساب بعض المؤشرات الإحصائية.

2- الإحصاء الإستدلالي (Inferential statistic)

يهتم هذا القسم بموضوع التقدير او التخمين واختيار الفرضيات.

تعريف الإحصاء الحيوي:

يعني الإحصاء الحيوي اشياء مختلفة للاشخاص المختلفين فهو للعامة جداول واعداد عن البيانات الحيوية اما المعنى الإصطلاحي للإحصاء فهو رياضيات جمع البيانات للظواهر البيولوجية وتنظيمها وتحليلها وتفسيرها والتعميم من الخاص الى العام عن طريق استدلال خواص المجتمع من خواص العينة.

اهمية علم الإحصاء:

يحتل الإحصاء مكاناً بين العلوم لما له من استعمالات واسعة للوصول الى قرارات صائبة لوصف او تفسير الظواهر المختلفة في جميع العلوم وهو المستعمل من قبل للافراد والجماعات المختلفة والدول على حد السواء وفي الحقيقة ان الإنتصار العظيم في نزول الإنسان على القمر ماكان يحدث لولا مساعدة علم الإحصاء، واستخدم الإحصاء في مجالات كثيرة ونركز على اهمية علم الإحصاء في العلوم البيولوجية والطبية والصحة العامة والكيمياء.

1- في علم الأحياء (البيولوجي): تستخدم الطرق الإحصائية في دراسة الأجناس والفصائل المختلفة للحيوان والنبات ومعرفة خواص كل جنس بما يتميز عن غيره واختلاف مفردات الجنس الواحد في اية خاصية معينة من الناحية الإحصائية، فمثلاً نرى الذكور في الجنس البشري اطول قامة من الإناث مع ان الذكور فيما بينهم يختلفون في الطول الى درجة ما وكذلك الإناث، كل ذلك يتم عن طريق جمع البيانات وتبويبها ودراستها دراسة احصائية والخروج بنتائج من هذه الصفات.

2- في الطب: يستخدم الإحصاء لدراسة العلاقة بين متغيرات كثيرة منها على سبيل المثال العلاقة بين العمر وضغط الدم وكذلك العلاقة بين الوراثة والبيئة وتأثيراتها على تكوين الفرد.

3- في الصحة العامة: يستخدم الإحصاء لدراسة الأمراض السارية ونسبة زيادتها ونقصها في المجتمع وكذلك دراسة حالة المعوقين والوفيات ونسبة الزيادة في السكان.

4- في الكيمياء: يستخدم الإحصاء لتحليل البيانات المتعلقة بتكرير النفط ومعرفة نسبة مكوناته وكذلك دراسة العلاقة بين الغازات او الفلزات او العمليات الكيمياوية من ناحية تحليل البيانات المتعلقة بها وكذلك التجارب الكيمياوية في اعداد بحوث الماجستير والدكتوراه والبحوث العلمية الأخرى وغيرها من التجارب في مجال النفط والمعادن وجمع البيانات المتعلقة بها ودراستها دراسة احصائية لغرض الإستفادة منها في اعداد خطط التنمية الصناعية والبتروكيمياوية.

المعالم والرموز الإحصائية

الرمز	مدلوله بالعربية	مدلوله بالإنجليزية
$\sum$	دلالة للجمع	Sum
$\bar{y}$	الوسط الحسابي للعينة	The arithmetic mean of the sample
M	الوسط الحسابي للمجتمع	The arithmetic mean of the population
S^2	تباين العينة	Sample variance
S	الإنحراف المعياري للعينة	Sample standard deviation
σ^2	تبياين المجتمع	population variance
σ	الإنحراف المعياري للمجتمع	The population standard deviation
$S_{\bar{y}}$	الإنحراف المعياري لمتوسط العينة (خطأ القياس)	Standard deviation of the sample mean (measurement error)

common variance	التباين المشترك	S_p^2
Common standard deviation	الإنحراف المعياري المشترك	S_p
deviation coefficient	معامل الإنحراف	$C.V$
degree of freedom	درجة الحرية	$d.f$
sum of squares	مجموع المربعات	$S.S$
null hypothesis	فرضية العدم	H_0
Alternative (counter) hypothesis	الفرضية البديلة (المضادة)	H_1
Substitution of r elements out of n elements	تبديلة لـ r عنصر من بين n عنصر	P_n^r
Match r elements out of n elements	توفيقة لـ r عنصر من بين n عنصر	C_n^r
Correlation coefficient	معامل الإرتباط	r
regression coefficient	معامل الإنحدار	B
Calculated F	F المحسوبة	F_{cal}
F tabular	F الجدولية	F_{tabl}
Calculated t	t المحسوبة	t_{cal}
t tabular	t الجدولية	t_{tabl}
Qui-squares	كاي مربع كاي	χ^2
normal distribution	التوزيع الطبيعي	$\mathcal{N}$
Gemma distribution	توزيع غاما	Γ

2.1 بعض المفاهيم الإحصائية:

المتغير: يقصد به اي صفة او عنصر قابل للتغير في النوع والكم من فرد الى أخر في نفس المجتمع ويكون المتغير اما:

أ-متغيرات وصفية او نوعية:

وهي الصفة التي لا يمكن قياسها مباشرة بأرقام عددية لان الفرق بين المفردات تكون في النوع وليس في الكم ومن الأمثلة على ذلك (الصحة، اللون، الذكاء، والجنس، والحالة الاجتماعية،...)

ب- صفة كمية: وهي الصفة التي يمكن قياسها مباشرة بأرقام عددية كالاختلاف بين الأفراد في الطول والوزن ومستوى الهيموغلوبين والهرمونات وعدد خلايا الدم الحمراء ومستوى الدهون في مصل الدم Lipid profile (TC ، TG ،HDL-C ، LDL-C ، VLDL) ويمكن قياسها بوحدات القياس المختلفة كالسنتيمتر والكيلوغرام (g, pg , mg) وتنقسم المتغيرات الكمية الى :

1- متغيرات متصلة او مستمرة

المتغير المتصل هو المتغير الذي تأخذ كل مفردة قيمة رقمية او كسر بين حدي المتغير الكلي مثلا لو فرضنا اطوال الطلبة يتراوح بين (130.5 و 170 سم) ,كمية الهيموغلوبين (12.5 – 14 ملغم لكل لتر من الدم)

2- متغيرات غير متصلة او مستمرة

هي المتغيرات التي تأخذ المشاهدة او المفردة فيها قيم متباعدة او متقطعة غير مستمرة اي هو الذي لا تأخذ كل مفردة فيه قيمة كسرية بل لا تزيد قيمة المتغير او تنقص بأقل من واحد فعدد الطلاب عدد الكتب كلها متغيرات غير متصلة او مستمرة.

المشاهدة:

تعتبر المشاهدة بمثابة المواد الأولية التي يتعامل معها الباحث فإذا اراد باحث ان يقيس مستوى الغلوكوز في مصل دم احد الجرذان ولنفرض ان مستوى الغلوكوز في مصل دم هذا الجرذ هو (120 ملغم/ 1مل) فإن هذا العدد يمثل المشاهدة، لذا فإن المشاهدة هي سجل رقمي لحادثة وان مجموع المشاهدات تكون البيانات Data.

المجتمع:

المجتمع من الناحية الإحصائية يمثل جميع الأفراد او العناصر التي تشترك في صفة متغير واحد او اكثر تميزه تماماً عن بقية المجتمعات ويتعلق مفهوم المجتمع بالهدف المحدد للبحث الإحصائي فقد يشكل طلبة جامعة تبسة مجتمعاً، والمجتمع هو عبارة عن جميع القيم التي يمكن ان يأخذها المتغير، فمثلاً عند دراسة مستوى الهيموغلوبين في دم طلبة جامعة تبسة وصفة مستوى الهيموغلوبين في دم طلبة جامعة تبسة هي متغير تأخذ مدى معين لمجتمع طلبة جامعة تبسة، والمجتمع اما ان يكون:

أ-مجتمع محدود:

وهو المجتمع الذي يمكن حصر مفرداته كما هو الحال في مستوى الهيموغلوبين في دم طلبة جامعة تبسة او عدد المرضى في مستشفى ما.

ب- مجتمع غير محدود:

هو المجتمع الذي من الصعب او المستحيل حصر مفرداته مثل عدد البكتريا في مستعمرة بكتيرية او حقل معين.

العينة:

العينة هي جزء المجتمع وهي عبارة عن مجموعة من المشاهدات اختير بطريقة ما من المجتمع حيث ان دراسة المجتمع ككل قد يكون صعباً ويحتاج الى وقت وجهد ومال لذا فقد استعيض عن دراسة المجتمع بدراسة العينة ومنها نستطيع ان نستنتج خواص المجتمع الذي اخذت منه العينة، فقد تكون العينة انسان او حيوان او نبات او جزء معلوم من نبات معين تجري عليه التجارب في المختبرات والعينة هي احدى ادوات البحث العلمي.

ومن اهم انواع العينات:

1- **العينة العشوائية البسيطة:**
وهي تلك العينة التي تسحب من مجتمع الدراسة بحيث يكون احتمال فرض ظهور اية مفردة من مفردات المجتمع الإحصائي في العينة متساوياً وبمعنى اخر تعني اعطاء كل فرد من المجتمع نفس الفرصة للظهور في العينة ويتم اختيارها كما يلي:
مثل استخدام طريقة البطاقات او القرعة
إذا كان لدينا (5) مرضى واردنا اختيار مريضين عشوائياً فما عدد الطرق الممكنة لأختيار مريضين لاجراء بعض الفحوصات.

ان عدد الطرق الممكنة $C_n^r = \frac{n!}{r!(n-r)!}$

$C_5^2 = \frac{5!}{2!(5-2)!}$ توفيقات

$C_5^2 = \frac{5\times4\times3\times2\times1}{2\times1\times(3\times2\times1)} = 10$ طرق مختلفة

بمعنى انه يوجد عشر بطاقات يكتب عليها اسم مريضين ويتم اختيار بطاقة من العشرة بطاقات عشوائياً.

فإذا كانت اسماء المرضى c,e,d,b,a فإن العشر بطاقات يكون مكتوب عليها ab , ac , ad , ae , bc , bd, be , cd, ce, de ويتم سحب اي بطاقة من العشرة .

2- **العينة المنتظمة:** وهي اختيار العينات بشكل منتظم من قائمة المجتمع حيث يتم اختيارها من خلال ترقيم عناصر المجتمع الإحصائي بحيث يتم تحديد قاعدة للاختيار تستند على تحديد اختيار العنصر الأول ولتبسيط الشرح لو كان مجتمع الأصل (100 مريض) وتريد اختيار (10 مرضى) لإجراء بعض الفحوصات عليهم فمثلاً تأخذ الأرقام العشرة الأولى وتوضع في صندوق ويتم السحب، فمثلاً حصلنا على الرقم (3) فيكون العينات العشرة المرضى هي كالآتي: 3, 13 , 23, 33, 43 , 53, 63, 83,73, 93 حيث

$$\text{المسافة} = \frac{\text{مجموع المرضى}}{\text{مجموع العينة}} = \frac{100}{10} = 10$$

ونقول ان الفاصلة = 10 بين مريض و آخر وتسمى هذه العينة المختارة (عينة منتظمة Systematic Sample).

3- العينة الطبقية:

يتم في هذا النوع من العينة تقسيم المجتمع الإحصائي اولاً الى مجموعات فرعية تسمى كل منها (طبقة Strate) ومن ثم تتم عملية المعاينة من كل طبقة ، وعادة تكون جميع عناصر الطبقة الواحدة متجانسة فيما يتعلق بالخصائص موضوع الدراسة فعلى سبيل المثال لو اريد اجراء دراسة معينة على مجتمع كلية طب الأسنان ونحتاج اخذ عينة من مجتمع كلية طب الأسنان عددها (20 عنصراً) علما ان مجتمع كلية طب الأسنان عدده (1000 فرد) حيث كان مجتمع كلية الطب مقسم الى الطبقات التالية:

رقم الطبقة	اسم الطبقة	عدد افراد الطبقة
1	اساتذة	150
2	موظفين	250
3	طلبة	600

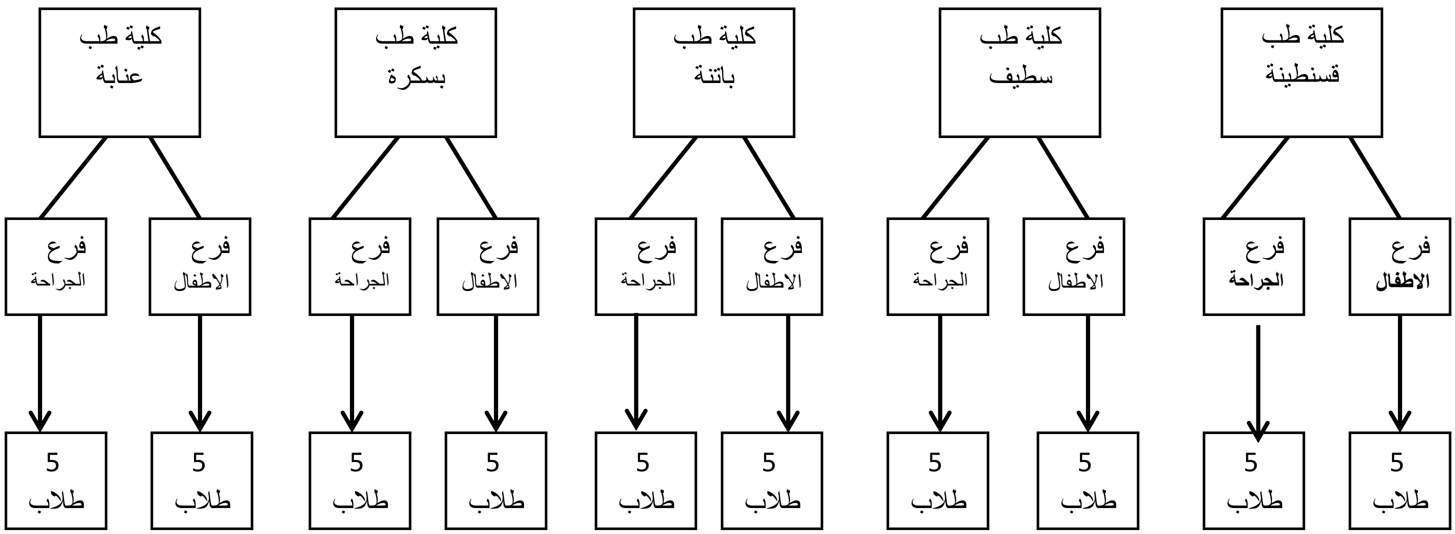
الحل: يتم اختيار عدد مفردات كل طبقة حسب العلاقة

$$\text{عدد افراد كل طبقة} = \frac{\text{حجم الطبقة}}{\text{حجم المجتمع}} \times \text{حجم العينة}$$

	عدد افراد الطبقة في العينة	عدد افراد الطبقة في المجتمع	رقم الطبقة
عدد افراد طبقة الأساتذة= $\frac{150}{1000} \times 20 =$ 3	3	150	1 اساتذة
عدد افراد طبقة الموظفين= $\frac{250}{1000} \times 20 =$ 5	5	250	2 موظفين
عدد افراد طبقة الطلاب= $\frac{600}{1000} \times 20 =$ 12	12	600	3 طلبة

4- العينة العنقودية متعددة المراحل:

تعتبر المعاينة العنقودية أحد الآليات التي يمكن استخدامها لاختيار العينات من خلال تقسيم المجتمع الى مجموعات او عناقيد على سبيل المثال نريد التعرف على مستوى التعليم الطبي في الجزائر فنختار 5 كليات طب ومن كل كلية فرعين ومن كل فرع 5 طلاب



الخطوات العامة لتكوين جدول توزيع تكراري:

1- **استخراج المدى الكلي**: يرمز له بالرمز R ويكتب

$$R = x_{max} - x_{min} + 1$$

2- **تحديد عدد الفئات**: ويرمز لعدد الفئات M

يفضل ان لا يقل عدد الفئات في التوزيع عن 5 ولا يزيد عن 15 فإذا قل عدد الفئات في التوزيع عن (5) فئات فإن عملية التبويب قد تؤدي الى عدم كشف الصفات الأساسية للمجتمع اي عدم اعطاء صورة واضحة لصفات المجتمع اما اذا زاد عدد الفئات عن (15) فئة فإن ذلك فيه صعوبات في اجراء العمليات الحسابية لبعض المؤشرات ويمكن حساب عدد الفئات حسب الصيغ التالية:

أ- صيغة يول $M = (2.5)\sqrt[4]{n}$

حيث n هو عدد المشاهدات

ب- صيغة ستيرجس Steruges : $M = 1 + (3.3)log(n)$

3- **ايجاد طول الفئة:**

طول الفئة = $\frac{\text{المدى}}{\text{عدد الفئات}}$ هنا تقرب النتيجة الى اقرب عدد صحيح

4- كتابة حدود الفئات:

بحيث ان جميع قيم المتغير عند كتابة حدود الفئات تضع بين الحد الأدنى للفئة الأولى والحد الأعلى للفئة الأخيرة.

5- استخراج عدد التكرارات لكل فئة:

مثال: البيانات التالية تمثل درجات 13 طالب من طلبة كلية الصيدلة في مادة الأنسجة

50, 51, 53, 59, 61, 61, 62, 63, 68, 69, 71, 74, 79

الحل:

1- استخراج المدى

$$R = x_{max} - x_{min} + 1$$

R= 79 – 50 +1 = 30

2- تحديد عدد الفئات

طريقة يول

$M = 2.5 \sqrt[4]{13}$

M= 2.5 × 1.898 = 4.75 ≃ 5

3- طريقة سترجس

M= 1 + 3.3 log (13)

M = 1 + 3.3 × 1.106 = 4.65 ≃ 5

4- طول الفئة = $\frac{R}{M}$ = $\frac{30}{5}$ = 6

الحد الادنى للفئة الأولى هو (50)

طول الفئة الأولى = الحد الأعلى – الحد الأدنى +1

6 = الحد الأعلى – 50 +1

6+50 -1= الحد الأعلى

الحد الأعلى = 55

إذن الحد الأعلى للفئة الأولى هو 55

بإضافة طول الفئة للحد الأدنى للفئة الأولى نحصل على الحد الأدنى للفئة الثانية والحد الأعلى للفئة الثانية هو الحد الأدنى للفئة الثانية زائد طول الفئة ناقص 1 وبتكرار العملية نحصل على الفئات الأخرى. (كما في الجدول)

الفئات	**التكرار** n_i	**مركز الفئات** x_i	**الحدود الحقيقية**	**التكرار النسبي**	**التكرار المئوي**
50 – 55	3	52.5	49.5 – 55.5	0.23	23
56 – 61	3	58.5	55.5 – 61.5	0.23	23
62 – 67	2	64.5	61.5 – 67.5	0.15	15
68 – 73	3	70.5	67.5 – 73.5	0.23	23
74 - 79	2	76.5	73.5 – 79.5	0.15	15
	$\sum_{i=1}^{5} n_i = 13$				

مثال: البيانات التالية تمثل تركيز المونولديهايد في اناث الأرنب المزالة من المبايض والتي عددها 40 انثى علماً ان تركيز المونولديهايد في مصل الدم مقاس $m.mol/L$. اعرض هذه البيانات في جدول توزيع تكراري:

2.6	2.2	4.1	3.5	2.0	3.2	3.7	3.0
3.7	3.4	1.6	3.1	3.3	3.8	3.1	2.4
3.1	2.5	4.3	3.4	3.6	2.9	3.3	3.9
3.4	3.3	3.1	3.7	4.4	3.2	4.1	1.9
3.5	2.3	3.8	3.2	2.6	3.9	3.0	4.2

الحل:

1- استخراج المدى

$$R = x_{max} - x_{min} + 1$$

$$R = 4.4 - 1.6 + 0.1$$

$$R = 2.9$$

2- تحديد عدد الفئات

$$M = 2.5\sqrt[4]{40}$$

M= 2.5 × 2.51 = 6.28 ≃ 6

3- ايجاد طول الفئة = $\frac{R}{M}$ = $\frac{2.9}{6}$ = 0.48 ≃ 0.5

4- كتابة حدود الفئات: بما ان اقل قيمة (1.6) للمتغير نأخذ الحد الأدنى للفئة الأولى 1.6

:. الحد الأدنى للفئة الأولى = 1.6

الحد الأعلى للفئة الأولى

طول الفئة = الحد الأعلى – الحد الأدنى + 0.1

0.5 = الحد الأعلى – 1.6 +0.1

الحد الأعلى = 0.5 + 1.6 – 0.1 = 2.0

الحد الأعلى للفئة الأولى هو 2.0

بإضافة طول الفئة للحد الأدنى للفئة الأولى نحصل على الحد الأدنى للفئة الثانية والحد الأعلى للفئة الثانية هو الحد الأدنى للفئة الثانية زائد طول الفئة ناقص 0.1 وبتكرار العملية نحصل على الفئات الأخرى.

حدود الفئات	**التكرار**	**مركز الفئات**	**الحدود الحقيقية**
1.6 – 2.0	3	1.8	1.55 – 2.05
2.1 – 2.5	4	2.3	2.05– 2.55
2.6 – 3.0	5	2.8	2.55 – 3.05
3.1 – 3.5	15	3.3	3.05 – 3.55
3.6 – 4.0	8	3.8	3.55 – 4.05
4.1 – 4.5	5	4.3	4.05 – 4.55
	40		

3.1 المقاييس الإحصائية:

المقاييس الاحصائية تتمثل في:

1.3.1- مقاييس النزعة المركزية: او ما تسمى بمقاييس التمركز او التوسط:

يشير مفهوم مقاييس النزعة المركزية الى ميل البيانات للتمركز حول قيمة ممثلة او نموذجية في التوزيع وتستخدم مقاييس النزعة المركزية لغايات المقارنة بين مجموعتين من البيانات ولوصف توزيع المشاهدات وتساعد هذه المقاييس في فهم وتفسير سلوك الظواهر وهذه المقاييس:

1- الوسط الحسابي
2- الوسيط
3- المنوال
4- الوسط الهندسي
5- الوسط التربيعي
6- الوسط التوافقي

ومن اهم مقاييس النزعة المركزية التي يمكن ان نستفاد منها في دراستنا هي:

1- الوسط الحسابي

هو عبارة عن القيمة التي يحصل عليها من خلال قسمة المجموع الكلي للقيم على عددها

أ- الوسط الحسابي للبيانات الغير مبوبة

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

حيث $\bar{x}$ = الوسط الحسابي

n = عدد المشاهدات

مثال1: اوجد الوسط الحسابي للبيانات التالية التي تمثل مستوى الهيموغلوبين في دم 6 رجال ملغم / ديسلتر.

x_i = 11 ,12, 13, 12, 13, 11

1- الطريقة الإعتيادية

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i}{n}$$

$$\bar{x} = \frac{11+12+13+12+13+11}{6} = \frac{72}{6} = 12$$

مثال$_2$: إذا كان متوسط مستوى الهرمون المحفز لنمو الحويصلات يساوي 18 Mg/dL حيث كان مستوى الهرمون المحفز لنمو الحويصلات في انثى الأرنب الأولى هو 18 وفي الأنثى الثانية هو 19 وفي الأنثى الثالثة هو 17 والأنثى الرابعة هو 19 اوجد مستوى الهرمون في انثى الأرنب الخامسة:

$$\bar{x} = \frac{\sum_{i=1}^{5} x_i}{n}$$

$$\bar{x} = \frac{x_1+x_2+x_3+x_4+x_5}{5}$$

$$18 = \frac{18+19+17+19+x_5}{5}$$

$$90 = 73 + x_5$$

$$x_5 = 90 - 76 = 17$$

طريقة الوسط الفرضي: تستخدم هذه الطريقة عندما تكون قيم مفردات العينة اعداد كبيرة ويصعب التعامل معها وخصوصاً عند عدم توفر الحاسبة تفي هذه الطريقة بالغرض

$$\boxed{\bar{x} = a + \frac{\sum_{i=1}^{n} d_i}{n}}$$

حيث a الوسط الفرضي

$\sum_{i=1}^{n} d_i$ مجموع الإنحرافات عن الوسط الفرضي

n = عدد المشاهدات

مثال: إذا كانت اوزان ستة طلاب من طلبة كلية الصيدلة كالآتي:

85, 67, 80, 75, 60, 55

اوجد الوسط الحسابي ؟

الحل:

نختار وسط فرضي وليكن = 75

x_i	$d_i = x_i - 75$
85	10
67	-8
80	5
75	0
60	-15
55	-20
$\sum_{i=1}^{6} x_i = 422$	$\sum_{i=1}^{6} d_i = -28$

ومنه الوسط الحسابي بهذه الطريقة هو

$$\overline{x} = 75 - \frac{28}{6} = 70.33$$

وحسب الطريقة الإعتيادية هو

$$\overline{x} = \frac{422}{6} = 70.33$$

ملاحظة: لا يتغير الوسط الحسابي بتغير الوسط الفرضي

ب – الوسط الحسابي في حالة البيانات المبوبة

$$\overline{x} = \frac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i}$$

مثال: اوجد الوسط الحسابي للبيانات التالية التي تبين توزيع (100) طالب من طلبة كلية الصيدلة حسب صفة الوزن، اوجد الوسط الحسابي لوزن طلبة الكلية ؟

$n_i x_i$	(مركز الفئات) x_i	التكرار n_i (عدد طلبة الفئة i)	الفئات (الوزن كغم)
305	61	5	$60 - 62$
960	64	15	$63 - 65$
3015	67	45	$66 - 68$
1890	70	27	$69 - 71$
584	73	8	$72 - 74$
$\sum_{i=1}^{5} n_i x_i = 6754$		$\sum_{i=1}^{5} n_i = 100$	

$$\overline{x} = \frac{6754}{100} = 67.54$$

جـ - الوسط الحسابي المرجح أو الموزون

من الناحية العملية هناك الكثير من الحالات تكون بعض المفردات اكثر اهمية من الأخرى مما يتوجب الأمر اخذ ذلك بعين الإعتبار لدى حساب الوسط الحسابي ، فمثلا عند حساب معدل نقاط الطالب فإن الأمر يستوجب الأخذ بعين الإعتبار عدد الساعات الأسبوعية المخصصة لكل درس وهذا يعني ترجيح المفردات بأوزان معينة تمثل اهمية كل منها وعند ادخال اهمية المفردات في حساب الوسط الحسابي عندئذ يسمى الوسط الحسابي المرجح وبتعبير آخر لكل قيمة من المشاهدات (x_i) وزن خاص يتناسب مع اهميتها (w_i) فالوسط الحسابي لهذه القيم يحسب كما يلي:-

$$\boxed{\overline{x}_w = \frac{\sum w_i x_i}{\sum w_i}}$$

حيث ان $\bar{x}_w$ الوسط الحسابي الموزون.

مثال: إذا كانت نقاط أحد الطلبة في الصف الأول في كلية الصيدلة في الدروس المقررة في هذه المرحلة حسب الساعات الأسبوعية المحددة لكل درس، المطلوب حساب معدل الطالب؟

النقطة	62	80	75	88	84	84	86
عدد الساعات	2	2	2	3	3	3	3

الحل:

النقطة x_i	الأهمية w_i	$w_i x_i$
62	2	124
80	2	160
75	2	150
88	3	264
84	3	252
84	3	252
86	3	258
	18	1460

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i} = \frac{1460}{18} = 81.11$$

الوسط الحسابي الموزون في حالة البيانات المبوبة:

$$\bar{x}_w = \frac{\sum w_i n_i x_i}{\sum w_i n_i}$$

x_i= مركز الفئة

n_i= التكرار

w_i= الأهمية

مثال: اوجد الوسط الحسابي الموزون للبيانات التالية التي تمثل إنتاج معمل الأدوية في الجزائر من الأدوية بالطن وعدد المكائن العاملة وعدد ساعات العمل؟

$w_i n_i x_i$	$w_i n_i$	x_i	عدد ساعات العمل w_i	عدد المكائن العاملة n_i	فئات الإنتاج بالطن
72	24	3	6	4	2 – 4
125	25	5	5	5	4 – 6
252	36	7	6	6	6 – 8
108	12	9	4	3	8 – 10
88	8	11	4	2	10 - 12
645	105			20	

$\bar{x}_w = \frac{645}{105} = 6.14$ طن

خصائص الوسط الحسابي:

1- مجموع انحرافات القيم عن وسطها الحسابي = صفر

بالفعل

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i - n\bar{x}$$
$$= \sum_{i=1}^{n} x_i - n\frac{\sum_{i=1}^{n} x_i}{n} = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0$$

من اجل $\bar{x} = \frac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i}$

$$\sum_{i=1}^{k} n_i(x_i - \overline{x}) = \sum_{i=1}^{k} n_i x_i - \sum_{i=1}^{k} n_i \overline{x} = \sum_{i=1}^{k} n_i x_i - \sum_{i=1}^{k} n_i \frac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i} = 0$$

2- مجموع مربعات انحرافات القيم عن وسطها الحسابي

$$\sum_{i=1}^{n} (x_i - \overline{x})^2$$

تلعب دور كبير في تقدير الوسائط.

3- يأخذ الوسط الحسابي بعين الإعتبار جميع القيم في حسابه.

4- يتأثر الوسط الحسابي بالقيم الشاذه او المتطرفة لان الوسط الحسابي يأخذ بعين الإعتبار جميع القيم.

5- هناك صعوبة في حساب الوسط الحسابي في حالة الفئات المفتوحة لأنه من الصعب تحديد مراكز الفئات وهذه المشكلة تحل بتحديد مراكز الفئات بصورة تقريبية.

الوسيط:

يعرف الوسيط بأنه القيمة التي تمثل المرتبة الوسطى عندما ترتب القيم قيد الدراسة تصاعدياً او تنازلياً وهذا يعني ان نصف القيم تقل عن قيمة الوسيط والنصف الآخر يزيد عنها.

أ-ايجاد الوسيط لبيانات غير مبوبة

1- يتم ترتيب القيم تصاعدياً او تنازلياً

2- إذا كان عدد القيم فردي (n) فالوسيط يكون القيمة التي ترتيبها $\frac{n+1}{2}$ وإذا كان عدد القيم زوجي (n) فالوسيط هو الوسط الحسابي للقيمتان اللتان ترتيبهما $\frac{n}{2}$ و $\frac{n}{2}+1$

مثال$_1$: اوجد الوسيط للبيانات التي تمثل مستوى الهيموغلوبين في دم (7) رجال ملغم / ديسلتر

11, 12, 13, 12, 13, 11, 14

الحل:

1- ترتب البيانات ترتيب تصاعدي 11 , 11 , 12 , 12 ,13 , 13 ,

2- إيجاد رتبة الوسيط

بما ان عدد القيم (n) = فردي

رتبة الوسيط = $\frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4$

$Me = 12$

مثال$_2$: اوجد الوسيط للبيانات التي تمثل مستوى الهيموغلوبين في دم 8 رجال ملغم/ديلتر

11, 12, 13 ,12, 13, 11, 14, 10

الحل:

1- ترتيب البيانات تصاعدياً 10 , 11 , 11, 12 , 12 , 13 , 13 , 14

2- ايجاد رتبة الوسيط

بما ان عدد القيم زوجي = 8

فالوسيط هو الوسط الحسابي للقيمتان التي ترتيبهما $\frac{n}{2}$, $\frac{n}{2}+1$

$$4=\frac{8}{2}=\frac{n}{2}$$

$$5=\frac{8}{2}+1 = \frac{n}{2}+1$$

$$Me=\frac{12+12}{2}=\frac{24}{2}=12$$

ب – ايجاد الوسيط لبيانات مبوبة:

$$\boxed{Me=L_j+\left[\frac{\frac{\sum_{i=1}^{k}n_i}{2}-N}{n_j}\right]\times C}$$

L_j = هي الحد الادنى الحقيقي لفئة الوسيط

$\frac{\sum_{i=1}^{k}n_i}{2}$ = رتبة الوسيط في حالة مجموع التكرارات عدد زوجي

$\frac{\sum_{i=1}^{k}n_i+1}{2}$ = رتبة الوسيط في حالة مجموع التكرارات عدد فردي

N = التكرار المتجمع الصاعد عند بداية فئة الوسيط

C = طول الفئة (طول فئة الوسيط)

n_j = التكرار المتجمع الصاعد عند نهاية فئة الوسيط - التكرار المتجمع الصاعد عند بداية فئة الوسيط

مثال: اوجد الوسيط للبيانات التالية التي تبين توزيع 100 طالب من طلبة كلية الصيدلة حسب صفة الوزن.

الفئات	التكرار n_i	التكرار المتجمع الصاعد
[60,62]	5	5
[63,65]	15	20
[66,68]	45	65
[69,71]	27	92
[72,74]	8	100
	$\sum_{i=1}^{k=7} n_i = 100$	

لإيجاد الوسيط نتبع مايلي:

1- إيجاد التكرار المتجمع الصاعد

2- إيجاد الفئة الوسيطية وهي [66,68]

3- إيجاد رتبة الوسيط رتبة الوسيط هي $\frac{\sum_{j=1}^{5} n_j}{2} = \frac{100}{2} = 50$

4- إيجاد الحد الأدنى للفئة الوسيطية وفي هذه الحالة نستعمل الحد الأدنى الحقيقي وهو $L_j = 65.5$ (لأن هناك إرتياب في دقة القياس).

5- طول الفئة الوسيطية هو الطول الحقيقي والذي هو $C = 68.5 - 65.5 = 3$.

6- $n_j = 45$

$$Me = 65.5 + \left[\frac{50 - 20}{45}\right] \times 3$$

$$Me = 65.5 + \frac{30}{45} \times 3$$

$$Me = 65.5 + 0.67 \times 3$$

$$Me = 65.5 + 2.01$$

$$Me = 67.51$$

ملاحظات عن الوسيط:

1- يستخدم الوسيط كمقياس للنزعة المركزية بدلاً عن الوسط الحسابي عندما تكون هناك قيمة شاذة في التوزيع.

2- الوسيط قليل الحساسية للمتغيرات التي تحدث في قيم البيانات الأصلية لأنه يهتم بالقيم الواقعة في المنتصف ويهمل الأطراف على عكس الوسط الحسابي الذي يعتبر شديد الحساسية لأنه يأخذ بعين الإعتبار جميع القيم في حسابه.

3- يمكن استخدامه في حالة المتغيرات الوصفية التي لا تعبر عنها بالأرقام كما هو الحال في ترتيب الأشخاص وفقاً لخصائصهم.

3- المنوال:

هي القيمة الأكثر شيوعاً او تكراراً في التوزيع وهو ابسط مقاييس النزعة المركزية.

أ-المنوال في حالة البيانات غير المبوبة

مثال$_1$: اوجد المنوال للبيانات التالية 4 , 6 , 8 , 4 , 7

$$Mo = 4$$

مثال$_2$: اوجد المنوال للبيانات التالية 3 , 6, 5 , 8 , 3 , 6 , 7

$$Mo = 3$$

$$Mo = 6$$

التوزيع ثنائي المنوال

ب – المنوال في حالة البيانات المبوبة

$$\boxed{Mo = L_j + \left[\frac{d_1}{d_1 + d_2}\right] \times C}$$

L_j = هي الحد الأدنى الحقيقي للفئة المنوالية

d_1 = الفرق بين فئة المنوال والفئة السابقة لها في التكرار

d_2 = الفرق بين فئة المنوال والفئة اللاحقة لها في التكرار

C = طول الفئة

مثال: اوجد المنوال للبيانات التالية التي تمثل توزيع طلبة كلية الصيدلة حسب صفة الوزن

الفئات	التكرار n_i
60 – 62	5

15	63 – 65
45 (الفئة المنوالية)	66 – 68
27	69 – 71
8	72 - 74

الحل:

$$d_1 = 45 - 15 = 30$$

$$d_2 = 45 - 27 = 18$$

$$Mo = L_i + \left[\frac{d_1}{d_1+d_2}\right] \times C$$

$$Mo = 65.5 + \left[\frac{30}{30+18}\right] \times 3$$

$$Mo = 65.5 + \frac{30}{48} \times 3$$

$$Mo = 65.5 + 0.625 \times 3$$

$$Mo = 65.5 + 1.88 = 67.38$$

2.3.1 مقاييس التشتت:

تحدد مقاييس التشتت مدى تقارب او تباعد او تناثر او اختلاف القيم عن بعضها البعض او عن نقطة معينة كالوسط الحسابي مثلاً، فالتشتت يرجع إلى اختلاف قيم المشاهدات، لذلك فإن القيم المتشابهة لا يوجد لها تشتت، ولكن باختلاف قيم المشاهدات فلابد من وجود تشتت بين هذه القيم، حيث يختلف مقدار التشتت من مجموعة إلى اخرى.

تعريف مقاييس التشتت: هي المقاييس التي تقيس مدى تباعد القيم او تقاربها والتي تستعمل كمؤشر احصائي لتحديد درجة التقارب او التشتت.

ومن اهم مقاييس التشتت:

أ - مقاييس التشتت المطلق

1- المدى: وهو ابسط مقاييس التشتت ويمكن حساب المدى للبيانات عن طريق المعادلة:

المدى = اعلى قيمة – اقل قيمة +1

مثال: اوجد المدى للمجموعتين التاليتين تمثل نقاط طلبة كلية الصيدلة في مادة كيمياء الأدوية؟

المجموعة A= 70, 60, 50, 40, 30

المجموعة B = 52, 51, 50, 49, 48

$E_A = 70 - 30 + 1 = 41$

$E_B = 52 - 48 + 1 = 5$

المجموعة A تشتتها أكبر من المجموعة B.

2- الإنحراف المتوسط

يعرف الإنحراف المتوسط بأنه معدل مجموع انحرافات القيم المطلقة عن متوسطها.

أ- الإنحراف المتوسط لبيانات غير مبوبة

$$\boxed{MD = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}}$$

مثال: اوجد الإنحراف المتوسط للبيانات التالية التي تمثل مستوى الهيموغلوبين في دم 6 رجال ملغم/ديلتر

$x_i = 11, 12, 13, 12, 13, 11$

الحل:

$\lvert x_i - \bar{x} \rvert$	x_i
1	11
0	12
1	13
0	12
1	13
1	11
$\sum_{i=1}^{6}\lvert x_i - \bar{x}\rvert = 4$	$\bar{x} = 12$

$$MD = \frac{\sum |x_i - \bar{x}|}{n}$$

$$MD = \frac{4}{6} = 0.67$$

ب – الانحراف المتوسط لبيانات مبوبة:

$$\boxed{MD = \frac{\sum_{i=1}^{k} n_i |x_i - \bar{x}|}{\sum_{i=1}^{k} n_i}}$$

مثال: اوجد الإنحراف المتوسط للبيانات التالية التي تمثل توزيع طلبة كلية الصيدلة حسب الوزن.

$n_i \mid x_i - \bar{x} \mid$	$\mid x_i - \bar{x} \mid$	$n_i x_i$	مركز الفئات x_i	التكرار n_i	الفئات
32.7	6.54	305	61	5	60 – 62
53.1	3.54	960	64	15	63- 65
24.3	0.54	3015	67	45	66 – 68
66.42	2.46	1890	70	27	69 – 71
43.68	5.46	584	73	8	72 - 74
220.2		6754		100	

$$MD = \frac{\sum_{i=1}^{5} n_i |x_i - \bar{x}|}{\sum_{i=1}^{5} n_i}$$

$$\bar{x} = \frac{\sum_{i=1}^{5} n_i x_i}{\sum_{i=1}^{5} n_i} = \frac{6754}{100} = 67.54$$

$$MD = \frac{220.2}{100} = 2.202$$

التباين والإنحراف المعياري:

يعد كل من الإنحراف المعياري والتباين كمقياس للتشتت من انسب المقاييس نظراً لتجاوزها المقاييس السابقة من ناحية واستخدامها على نطاق واسع في التحليل من ناحية ثانية ويعرف التباين بأنه معدل مجموع مربعات انحرافات القيم عن متوسطها.

أ- التباين في حالة البيانات غير المبوبة:

تباين العينة $S^2 = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}$ (الطريقة الإعتيادية)

(الطريقة السريعة) $S^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$

حيث ان التباين $S^2 = \frac{ss}{d.f}$

$$ss = \sum_{i=1}^{n}(x_i - \overline{x})^2$$

او $ss = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$

$d.f =$ هي درجات الحرية او عدم السيطرة

$$d.f = n - 1$$

حيث n هي عدد القيم.

مثال: احسب التباين للقيم التالية التي تمثل نقاط7 طلاب من طلبة كلية الصيدلة في مادة كيمياء الادوية العملي ؟ 9, 4, 6, 8, 10, 5, 7

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
9	+2	4
4	- 3	9
6	- 1	1

8	+1	1
10	+ 3	9
7	0	0
5	- 2	4
49		28

$$\bar{x} = \frac{\sum_{i=1}^{7} x_i}{n} \quad \rightarrow \quad \bar{x} = \frac{49}{7}$$

$$S^2 = \frac{\sum_{i=1}^{7}(x_i - \bar{x})^2}{n-1}$$

$$S^2 = \frac{28}{7-1} = 4.67$$

الطريقة السريعة

x_i	x_i^2
9	81
4	16
6	36
8	64
10	100
7	49
5	25
49	371

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n-1}$$

$$S^2 = \frac{371 - \frac{(49)^2}{7}}{7-1}$$

$$S^2 = \frac{371-343}{7-1} = 4.67$$

ب – **في حالة البيانات المبوبة:**

$$S^2 = \frac{\sum_{i=1}^{k} n_i (x_i - \overline{x})^2}{\sum_{i=1}^{k} n_i - 1}$$

$$S^2 = \frac{\sum_{i=1}^{k} n_i x_i^2 - \frac{\left(\sum_{i=1}^{k} n_i x_i\right)^2}{\sum_{i=1}^{k} n_i}}{\sum_{i=1}^{k} n_i - 1}$$

مثال: احسب التباين للبيانات التالية التي تمثل توزيع طلبة كلية الصيدلة حسب الوزن.

الفئات	التكرار n_i	مركز الفئات x_i	$n_i x_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$n_i(x_i - \bar{x})^2$
60 – 62	5	61	305	- 6.54	42.7716	213.858
63- 65	15	64	960	- 3.54	12.5316	187.974
66 – 68	45	67	3015	0.54	0.2916	13.122
69 – 71	27	70	1890	2.46	6.0516	163.3932
72 - 74	8	73	584	5.46	29.8226	238.4928
	100		6754			816.84

$$\bar{x} = \frac{\sum_{i=1}^{5} n_i x_i}{\sum_{i=1}^{5} n_i} = \frac{6754}{100} = 67.54$$

$$S^2 = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x})^2}{\sum_{i=1}^{k} n_i - 1}$$

$$S^2 = \frac{816.84}{100-1} = \frac{816.84}{99} = 8.25$$

الفئات	n_i التكرار	x_i	$n_i x_i$	x_i^2	$n_i x_i^2$
60 – 62	5	61	305	*3721*	18605
63 - 65	15	64	960	4096	61440
66 – 68	45	67	3015	4489	202005
69 – 71	27	70	1890	4900	132300
72 - 74	8	73	584	5329	42632
	100		6754		456982

$$S^2 = \frac{\sum_{i=1}^{k} n_i x_i^2 - \frac{\left(\sum_{i=1}^{k} n_i x_i\right)^2}{\sum_{i=1}^{k} n_i}}{\sum_{i=1}^{k} n_i - 1}$$

$$S^2 = \frac{456982 - \frac{(6754)^2}{100}}{100 - 1}$$

$$= \frac{456982 - 456165.16}{99} = \frac{816.84}{99} = 8.25$$

الإنحراف المعياري:

هوالجذر التربيعي للتباين ويستخدم على نطاق واسع كونه يتعامل مع نفس وحدات القياس للمشاهدات الأصلية ويعتبر الإنحراف المعياري اهم مقاييس التشتت واكثرها استعمالاً في مجال التحليل الإحصائي.

$$S = \sqrt{s^2}$$

مثال: اوجد الإنحراف المعياري اذا كان التباين (4.67) وكذلك اذا كان التباين (8.25)

$$S = \sqrt{s^2}$$

$$S = \sqrt{4.67} = 2.16$$

$$S = \sqrt{8.25} = 2.87$$

الخطأ القياسي:

يسمى الإنحراف المعياري لمتوسط العينة ويستخدم للدلالة على التشتت فكلما كان الخطأ القياسي قليلاً كلما كان هناك تقارب او تجانس اكثر بين القيم وكلما زاد الخطأ القياسي كلما قلت دقة القياس ودل ذلك على تشتت القيم

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

مثال: اوجد الخطأ القياسي إذا كان التباين (4.67) وعدد القيم 6

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

$$= \frac{2.16}{\sqrt{6}} = \frac{2.16}{2.45} = +0.8$$

3.3.1 مقاييس التشتت النسبي:

من اهم مقاييس التشتت النسبي:ول

1- معامل الإختلاف:

يستخدم للمقارنة بين المجموعات المختلفة او بين العينات فإننا لا نستطيع اجراء مقارنة بناء على الإنحراف المعياري لكل مجموعة لأننا بحاجة الى توحيد القياس بالنسبة للمجموعتين لذلك يتم استخدام معامل الإختلاف.

$$C.V = \frac{S}{\bar{x}} \times 100$$

فيما يلي نقاط مجموعتين من الطلاب

المجموعة الأولى	المجموعة الثانية
2	2000
2	2000
4	4000
5	5000
12	12000

فعند حساب الوسط الحسابي والإنحراف المعياري لكلتا المجموعتين

المجموعة	الوسط الحسابي	الإنحراف المعياري
الأولى	5	4.123
الثانية	5000	4123.11

فهل نستطيع المقارنة بين المجموعتين بناءاً على الإنحراف المعياري، كما أشرنا سابقاً لا نستطيع حيث نحن بحاجة الى توحيد القياس لابد من استخدام معامل الإختلاف.

معامل الإختلاف للمجموعة الأولى $C.V = \frac{4.123}{5} \times 100 = \% \ 82.46$

معامل الاختلاف للمجموعة الثانية $C.V = \frac{4123.11}{5000} \times 100 = \% \ 82.46$

بناءاً على هذه النتيجة فإن معامل الإختلاف هو واحد بالنسبة للمجموعتين او ان التباين متساوي.

يعرف معامل الإختلاف على انه النسبة المئوية التي يشكلها الإنحراف المعياري.

مثال: نتائج الإمتحانات لدرس الإحصاء الحيوي والكيمياء السريرية لطلبة كلية الصيدلة كانت كما مبين ادناه. اي من النقاط بالنسبة للدرسين أكبر تشتتاً؟

	الإحصاء	الكيمياء
الوسط الحسابي	78	73
الإنحراف المعياري	8	76

معامل الإختلاف للإحصاء $C.V = \frac{8}{78} \times 100 = \%\ 10.25$

معامل الإختلاف للكيمياء $C.V = \frac{76}{73} \times 100 = \%\ 104.11$

ملاحظة: الحد الأعلى لمعامل الإختلاف للتجارب المختبرية يجب ان لا يتجاوز (10%) وللتجارب الحقيقية يجب ان لا يتجاوز (20%)

2- النقطة المعيارية:

إن المقارنة بين النقاط للفرد بناءاً على النقطة الخام ليس له معنى وبالتالي لابد من تحويل هذه النقطة الى نقطة جديدة، واحد هذه التحويلات تسمى بالنقطة المعيارية ومن خصائصها ان متوسطها (صفر) وانحرافها المعياري (1) وتستخدم النقطة المعيارية لمقارنة اداء طالب معين في مواد مختلفة مثل:

مثال: لمقارنة اداء طالب من طلبة كلية الصيدلة في مواد دراسية مختلفة؟

النقطة المعيارية $z = \frac{x_i - \overline{x}}{s}$

	اللغة الإنجليزية	الإحصاء الحيوي	كيمياء الأدوية
النقطة	75	70	60

الوسط الحسابي	70	55	50
الإنحراف المعياري	10	15	5

عند النظر الى النقاط نقول اداء الطالب في اللغة الإنجليزية أفضل ولكن عند التحويل الى النقطة المعيارية

$$Z = \frac{x_i - \bar{x}}{s}$$

اللغة الإنجليزية $$Z = \frac{75-70}{10} = 0.5$$

الإحصاء $$Z = \frac{70-55}{15} = +1$$

كيمياء الأدوية $$Z = \frac{60-50}{5} = +2$$

اذن اداء الطالب أفضل في مادة كيمياء الأدوية.

4.1 عرض البيانات:

تعرض البيانات بأشكال مختلفة كالدوائر المجزأة او الأعمدة والخطوط المتكسرة وغيرها وان هذه الأشكال الهندسية ماهي إلا تعبير يوضح البيانات بطريقة جذابة وسهلة وفعالة تساعد القارئ على فهم واستيعاب قيم الظاهرة وفهمها ووسائل التمثيل البياني كثيرة منها:

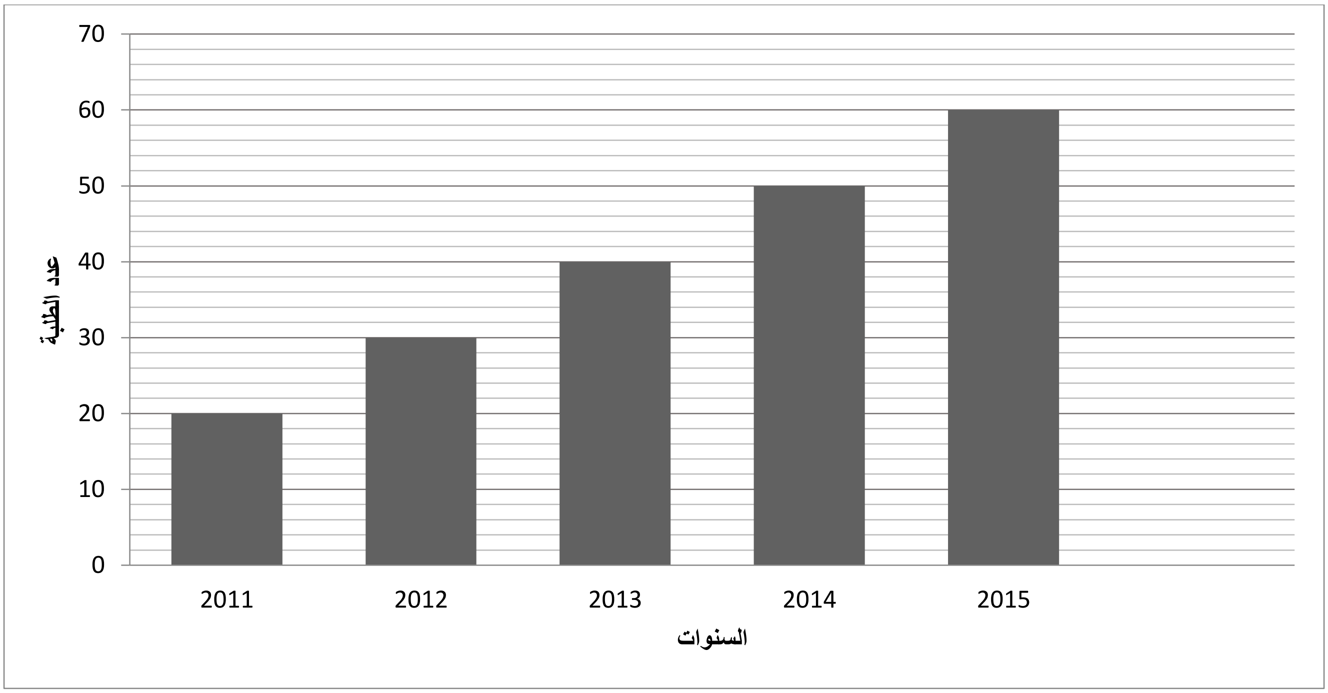
أ - الأشرطة البيانية:

وهي عبارة عن مستطيلات الرئيسية او الأفقية قواعدها متساوية وتمثل الصفة التي يتم على اساسها التبويب (سنة، شهر، محافظة وضغط الدم.. الخ) وارتفاعها تمثل البيانات المقابلة لتلك الصفة مثل عدد الطلبة، عدد المرضى، درجة الحرارة.

مثال: كانت خطة القبول لكلية الصيدلة للسنوات المبينة ادناه، ارسم شريط بياني لخطة القبول؟

السنة	عدد الطلبة
2011	20
2012	30

40	2013
50	2014
60	2015



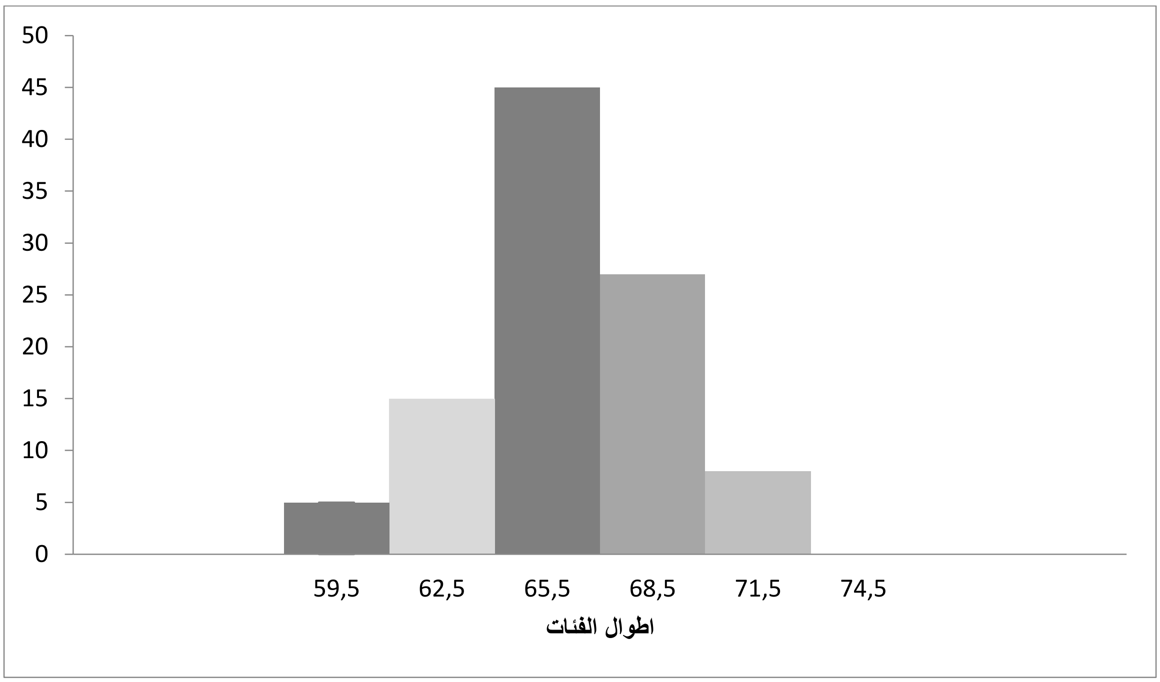
ب- المدرج التكراري:

هو عباره عن مجموعة من المستطيلات تمتد قواعدها على المحور الأفقي لتمثل اطوال الفئات بينما ارتفاعها تمثل تكرار لتلك الفئة.

مثال: ارسم المدرج التكراري للبيانات التالية التي تمثل توزيع طلبة كلية الصيدلة حسب صفة الوزن؟

التكرار n_i	الفئات
5	$60 - 62$

15	63- 65
45	66 – 68
27	69 – 71
8	72 -74

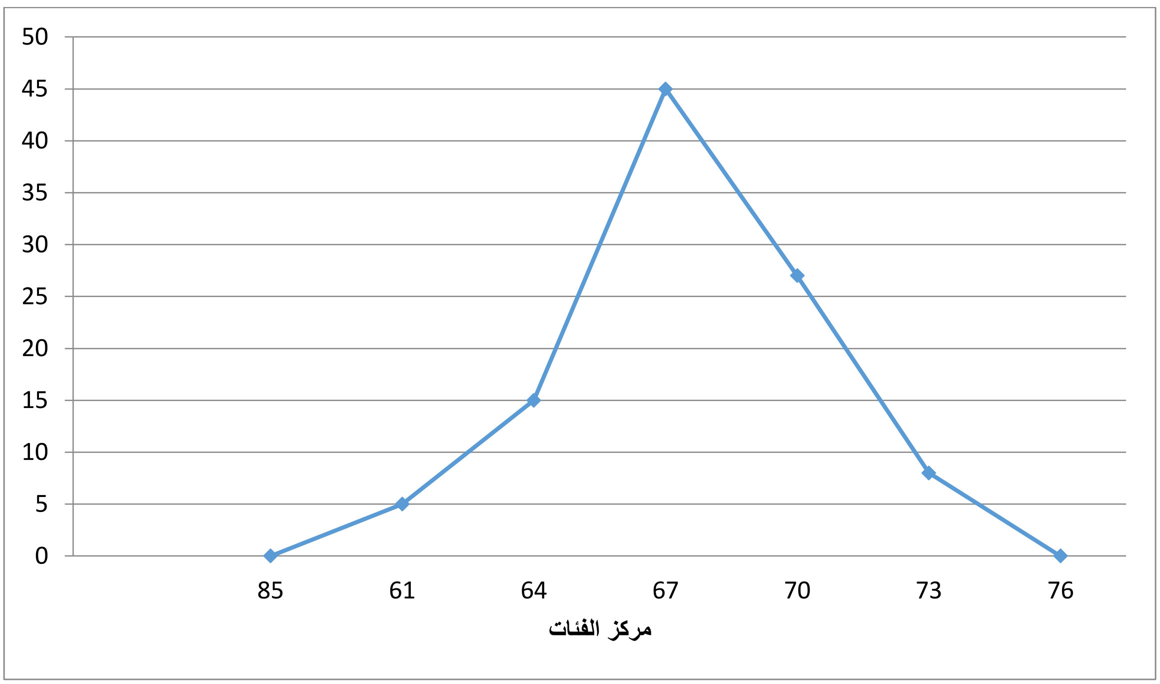


جـ - المضلع التكراري:

هو عبارة عن خطوط مستقيمة متكسرة تصل بين نقاط كل منها واقعة فوق مركز فئة على ارتفاع يمثل تكرار تلك الفئة وعادة يقفل المضلع عن طريق ايصال بداية المضلع بالمحور الأفقي بمركز فئة خيالية واقعة على يسار اول فئة بتكرار يساوي (صفر) ويتصل نهاية المضلع بالمحور الأفقي بمركز فئة خيالية واقعة الى يمين اخر مركز فئة بتكرار يساوي (صفر).

مثال: ارسم المضلع التكراري للبيانات التالية التي تمثل توزيع طلبة كلية الصيدلة حسب صفة الوزن؟

التكرار n_i	الفئات
5	60 – 62
15	63- 65
45	66 – 68
27	69 – 71
8	72 - 74



د- الدائرة البيانية:

وهي عبارة عن شكل هندسي دائري يقسم إلى مجموعة من القطاعات بحيث ان مجموع مساحات القطاعات تمثل المساحة الكلية للدائرة وبهدف تحديد كل قطاع فإن ذلك يتطلب تحديد زوايا كل قطاع

$$\text{زاوية القطاع} = \frac{\text{عدد بيانات الصنف}}{\text{مجموع البيانات الكلية}} \times 360$$

مثال: كانت المصاريف الشهرية لأحد العوائل على النحو التالي:

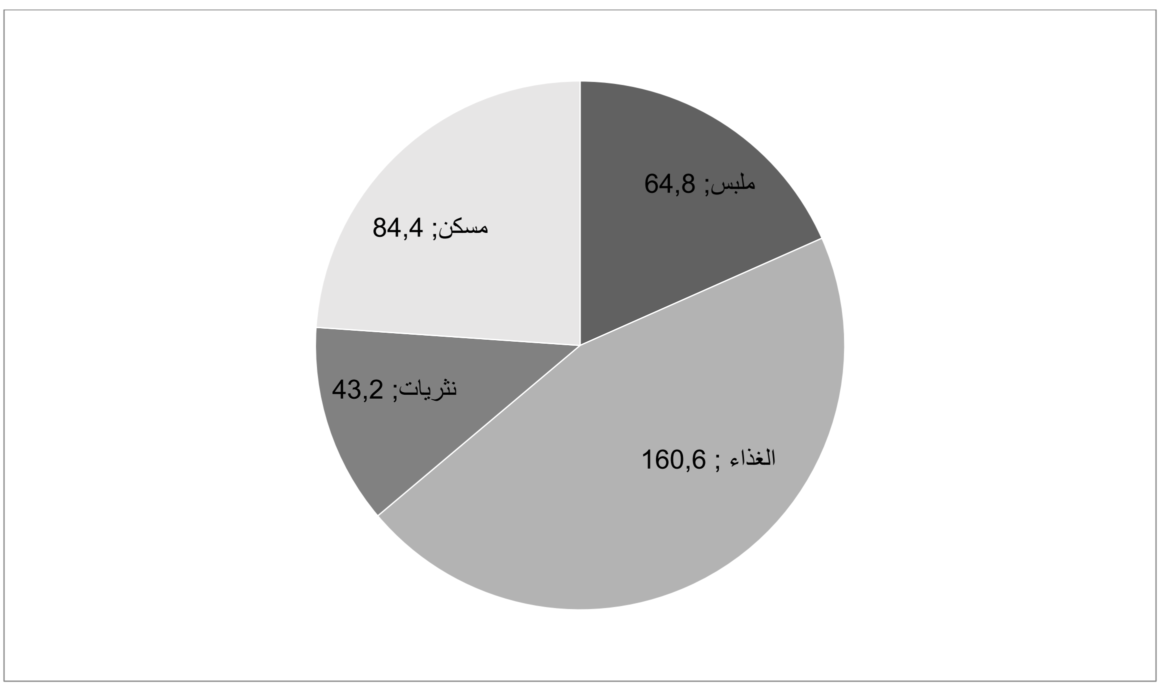
المصروفات	التفاصيل
460.000 دينار	غذاء
180.000 دينار	ملبس
240.000 دينار	مسكن
120.000 دينار	نثريات
1000,000	

زاوية القطاع الخاص بالغذاء = $\frac{460.000}{1000,000} \times 360 = 165.6^{\circ}$

زاوية القطاع الخاص بالملبس = $\frac{180.000}{1000,000} \times 360 = 64.8^{\circ}$

زاوية القطاع الخاص بالمسكن = $\frac{240.000}{1000,000} \times 360 = 84.4^{\circ}$

زاوية القطاع الخاص بالنثرية = $\frac{120.000}{1000,000} \times 360 = 43.2^{\circ}$



الفصل الثاني

الإحصاء الإستدلالي

1.2 مدخل حول مبادئ نظرية الإحتمال:

إن نظرية الإحتمال تلعب دوراً هاماً في نظريات تطبيقات علم الإحصاء وهي تهتم بدراسة التجارب العشوائية ومن الأمثلة:

1- تجارب رمي زهرة النرد حيث زهرة النرد له 6 وجوه.

2- تجارب قطعة النقود، قطعة النقود لها وجهان كتابة (Q) وصورة (F).

3- تجارب صندوق الكرات وصندوق الكرات يحتوي على كرات مختلفة الألوان

4- تجارب اوراق اللعب;

إن مجموعة اوراق اللعب تتألف من 52 ورقة مقسمة الى اربعة مجاميع

أ - مجموعة un pique

ب - مجموعة القلب le cœur

جـ - مجموعة ماجة le trèfle

د - مجموعة ديتر le carreau

وكل مجموعة تحتوي على اوراق اربعة صور وتسعة اوراق تحمل ارقام من 2 الى 10 اي كل مجموعة تتكون من 13 ورقة كما ان مجموعة اوراق اللعب تتكون من لونين اسود واحمر كل لون 26 ورقة.

2.2 بعض المصطلحات والتعاريف:

1- التجربة العشوائية:

هي التجربة التي لا يمكن معرفة نتيجتها لخضوعها لقوانين الإحتمال، إن رمي زهرة النرد تجربة عشوائية لأن النتائج الممكنة لهذه التجربة تخضع لقوانين الإحتمال وإذا اراد صيدلي تقدير المادة الفعالة في بذور أحد النباتات التي تستخدم في صناعة عقار معين فإن العينة التي سيبني عليها تقديره للمادة الفعالة والنتائج الممكنة التي سيحصل عليها ستخضع إلى قوانين الإحتمال ولهذا فالتجربة عشوائية وان نسبة عالية من التجارب العلمية هي تجارب عشوائية.

2- فضاء العينة:

فضاء العينة هو مجموعة من النقاط تمثل جميع النتائج الممكنة لتجربة ما حيث ان كل نتيجة تمثل بنقطة او عنصر، مثلا:

في حالة رمي قطعة نقود فضاء العينة هو: $\Omega = \{Q \ , \ F\}$

اما إذا رمينا قطعتي نقود فإن فضاء العينة سيكون اربعة نتائج:

$$\Omega = \{QQ , QF , FQ , FF\}$$

اما إذا رمينا زهرة النرد مرة واحدة فإن فضاء العينة يكون 6 نتائج ممكنة:

$$\Omega = \{1,2,3,4,5,6\}$$

3. الحادث:

هو نقطة او عدة نقاط في فضاء العينة ويرمز له بالرمز (E_i) فالحصول على الصورة (F) في رمي قطعة النقود مرة واحدة يسمى حادثاً وهويتكون من نقطة واحدة (F) من مجموع نقاط فضاء العينة $\{Q , F\}$, وكذلك فإن الحصول على عدد زوجي في رمي زهرة النرد يسمى ايضاً حادثاً يتكون من النقاط $\{2,4,6\}$ من مجموع نقاط فضاء العينة$\{1,2,3,4,5,6\}$. والحادث يكون بسيط إذا تكون من نقطة واحدة في فضاء العينة اي حالة واحدة من الحالات التي تظهر نتيجة التجربة او يكون حادثاً مركباً إذا شمل حالتين او أكثر من الحالات التي تظهر نتيجة التجربة.

والحوادث انواع:

1- الحوادث المتنافية (المستبعدة):

يقال عن حدثين انهما متنافيان (مستبعدان) إذا استحال حدوثهما معاً.

مثلاً عند رمي قطعة نقود من المستحيل الحصول على صورة وكتابة في نفس الوقت.

2- الحوادث غير المتنافية:

وهي إما احداث مستقلة او احداث غير مستقلة

أ- الحوادث المستقلة:

هي الحوادث التي إذا وقع أحدهما لا يمنع او يؤثر على وقوع الأحداث الأخرى.

فمثلا عند رمي قطعتي نقود فالحصول على صورة في القطعة الأولى مثلاً لا يؤثر في نتيجة القطعة الثانية.

صندوق الكرات: عند سحب الكرة الأولى وارجاعها لا يؤثر في نتيجة السحب الثانية.

ب- الحوادث غير المستقلة:

هي الحوادث التي إذا تحقق أحدهما يؤثر في تحقق الأحداث الأخرى، ففي حالة صندوق به كرات فعند سحب كرتان على التوالي بدون إرجاع فإن نتيجة السحب الأولى تؤثر في نتيجة السحب الثانية.

4. الحالات الممكنة:

هي جميع الحالات المختلفة التي يمكن ان تظهر في تجربة ما، فعند رمي قطعة نقود فعدد الحالات الممكنة هنا حالتين صورة وكتابة وعند رمي زهرة النرد عدد الحالات الممكنة 6 وعند رمي زهرتي النرد عدد الحالات الممكنة $6 \times 6 = 36$.

5. الحالات المواتية:

هي الحالات التي تحقق طور الأحداث المراد دراستها وتسمى بحالات النجاح.

لذا فإحتمال حدوث الحدث:

$$P(E_i) = \frac{\text{عدد الحالات المواتية للحادث}}{\text{عدد الحالات الممكنة}} = \frac{n}{N}$$

$$P(E_i) + P(\bar{E_i}) = 1$$

احتمال النجاح + احتمال الفشل=1

احتمال الفشل $P(\bar{E}_i) = 1 - P(E_i)$

3.2. طرق العد:

أ- **التبديلات:** يقصد بالتبديلات بأنها عدد طرق الإختيار المرتب التي يمكن تكوينها من عدة اشياء يأخذها كلها او يعوضها ويرمز لها بـ : P_n^r

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

مثال$_1$: إذا كان لدينا اربعة حروف A , B , C , D واختير منها حرفان فما هي عدد الطرق التي يمكن بها اختيار هاذان الحرفان ؟

$$P_4^2 = \frac{4!}{(4-2)!} = \frac{4\times3\times2\times1}{2\times1} = 12$$

AC, AD, AB, BC, BD, CD, CA, DA, BA, CB, DB, DC

مثال$_2$: كتبت الأرقام من 1 الى 9 على بطاقات ووضعت في صندوق ثم سحبت منها 5 بطاقات الواحدة بعد الأخرى فكم عددا خماسياً ارقامه مختلفة يمكن تكوينه ؟

$$P_9^5 = \frac{9!}{(9-5)!} = \frac{9\times8\times7\times6\times5\times4\,!}{4!} = 15120$$

ملاحظة: إذا كان $n = r$ نأخذ اصطلاحا $0! = 1$.

مثال$_3$: إذا اراد طالب ان يرتب 4 كتب مختلفة المواضع على رف مكتبته فبكم طريقة يمكن ترتيبها ؟

من الوجهة العلمية إذا كان $n = r$ فإن عدد التبديلات هو عدد الطرق التي يمكن ترتيب n من الأشياء على خط مستقيم .

الحل:

يمكن اختيار الكتاب الأول بأربعة طرق 4

يمكن اختيار الكتاب الثاني بثلاثة طرق 3

يمكن اختيار الكتاب الثالث بطريقتين 2

يمكن اختيار الكتاب الرابع بطريقة واحدة 1

$$P_4^4 = \frac{4!}{(4-4)!} = \frac{4!}{0!} = \frac{4\times3\times2\times1}{1} = 24$$

ملاحظة: التبديلات في حالة وجود مجاميع متشابهة

إذا كان لدينا n عنصر مشكلة كمايلي: $n = n_1 + \cdots + n_r$ حيث

n_i هو عدد العناصر من نفس النوع i أين $i = 1,2,\dots,r$ عندئذ :

$$P = \frac{n!}{n_1! \times n_2! \times \dots \times n_r!}$$

يمثل عدد التبديلات من n عنصر.

مثال$_4$: ماهي الطرق التي يمكن بها ترتيب أحرف كلمة باب ؟

الحل:

تكرار الحرف ب = 2

تكرار الحرف أ = 1

عدد الطرق هو

$$P = \frac{3!}{1! \times 2!} = \frac{3 \times 2!}{1! \times 2!} = 3$$

مثال$_5$: ما هو عدد الترتيبات التي يمكن تكوينها من أحرف كلمة Statistique

الحل:

عدد الحروف = 11

حرف S تكرر 2 مرات

حرف t تكرر 3 مرات

حرف a تكرر 1 مرات

حرف i تكرر 2 مرات

حرف q تكرر 1 مرات

حرف u تكرر 1 مرات

حرف e تكرر 1 مرات

عدد الترتيبات هو:

$$P = \frac{n!}{n_1! \times n_2! \times n_3! \times n_4! \times n_5! \times n_6! \times n_7!}$$

$$P = \frac{11!}{2! \times 3! \times 1! \times 2! \times 1! \times 1! \times 1!} = \frac{11!}{4 \times 3!} = \frac{4! \times 5 \times 6 \times 7 \times 8 \times 9 \times 10 \times 11}{4!}$$
$$= 1663200$$

ب- **التوفيقات:**

يقصد بالتوفيقات طرق الإختيار الغير مرتب التي يمكن تكوينها من عدة اشياء يأخذها كلها او بعضها ويرمز للتوفيقات بالرمز: $\binom{n}{r}$ حيث

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

مثال$_1$: ما عدد طرق الإختيار التي يمكن الحصول عليها لاختيار لجنة مؤلفة من 5 صيادلة من مجموع 9 صيادلة؟

$$\binom{9}{5} = \frac{9!}{5!\,(9-5)!} = \frac{9\times8\times7\times6\times5!}{5!\times4\times3\times2\times1} = \frac{9\times8\times7\times6}{4\times3\times2\times1} = 126$$

ملاحظة: هناك قاعدتان اساسيتان يعتمد عليها كل من التبديلات والتوفيقات.

1- إذا كان عدد الطرق الممكنة لوقوع الحادث (E_1) هو n وعدد الطرق لوقوع الحادث (E_2) هو m وكان E_1 و E_2 حادثان متنافيان فإن عدد الطرق لوقوع الحادث E_1 او E_2 هو $(n+m)$ من الطرق.

2- إذا كان عدد الطرق الممكنة لوقوع الحادث E_1 هو n وان عدد الطرق الممكنة لوقوع E_2 هو m وكان E_1 و E_2 حادثان مستقلان فإن عدد الطرق الممكنة لوقوع الحادث E_1 و E_2 هو $(n \times m)$ من الطرق.

مثال$_2$: كم لجنة سباعية يمكن اختيارها من 6 اطباء و5 صيادلة على ان تضم 4 اطباء؟

عدد الطرق هو $\binom{6}{4}\times\binom{5}{3}$

$$\binom{6}{4} = \frac{6!}{4!\,(6-4)!} = \frac{6\times5\times4!}{4!\times2!} = \frac{6\times5}{2\times1} = 15$$

$$\binom{5}{3} = \frac{5!}{3!\,(5-3)!} = \frac{5\times4\times3!}{3!\times2\times1} = 10$$

عدد الطرق = 15×10= 150.

4.2 قانون الإحتمال:

لقد وضعت القوانين التالية لتسهيل حساب درجة الإحتمال عند وقوع حدثين او أكثر بدلاً من ايجادهما عن طريق تعريف الإحتمال الذي يكون من الصعوبة في مثل هذه الحالات حساب عدد الحالات المواتية والممكنة.

وقبل شرح قوانين الإحتمال نفرض ان هناك حادثان E_1 و E_2 فالتعابير التالية يقصد بها ما يلي :

$P(E_1 \cup E_2)$ احتمال وقوع الحدث E_1 أو E_2 . (اي احتمال وقوع اياً منهما فقط)

$P(E_1 \cap E_2)$ احتمال وقوع الحادث E_1 والحادث E_2 معاً .

$P(E_1/E_2)$ احتمال حدوث E_1 علماً بأن الحادث E_2 قد وقع. ويسمى بالإحتمال الشرطي.

1- قانون الجمع:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

أ- إذا كانت الاحداث متنافية

إذا كانت E_1 و E_2 حادثان متنافيان

بما ان الحوادث المتنافية دائماً يكون تقاطعها مجموعة خالية

عندئذ احتمال حدوث اياً منها هو حاصل جمع احتمال كل منهما أي أن:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

مثال$_1$: صندوق يحتوي 4 كرات سوداء و5 بيضاء و3 حمراء فإذا سحبت كرة واحدة فما هو احتمال ان تكون اما سوداء او بيضاء؟

عدد الكرات $4N + 5B + 3R = 12$

$P(N) = \frac{4}{12}$ احتمال كون الكرة سوداء

$P(B) = \frac{5}{12}$ احتمال كون الكرة بيضاء

$$P(N \cup B) = P(N) + P(B) = \frac{4}{12} + \frac{5}{12} = \frac{9}{12} = \frac{3}{4}$$

مثال$_2$: في حالة رمي زهرة نرد، ما هو احتمال الحصول على عدد زوجي ؟

$P(\{2\}) = \frac{1}{6}$

$P(\{4\}) = \frac{1}{6}$

$P(\{6\}) = \frac{1}{6}$

$$P(\{2\} \cup \{4\} \cup \{6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6}$$

ب - إذا كانت الاحداث غير متنافية:

إذا كان الحدثان E_1 و E_2 حدثان غير متنافيان فإن احتمال حدوث اي منهما (E_1 او E_2) هو حاصل جمع احتمال كل منهما ناقصا احتمال حدوثهما معاً أي :

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

إذا كان أكثر من حدثان غير متنافيان.

$$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3)$$

مثال$_1$: في كلية طب الأسنان 25% من الطلبة رسبو بالرياضيات و15% من الطلبة رسبو في الكيمياء و10% رسبو كلاً من الرياضيات والكيمياء، فإذا انتخب طالب منهم عشوائياً فما هو احتمال ان يكون راسباً في الرياضيات او الكيمياء

نرمز للرياضيات بالرمز M والكيمياء بالرمز C

$$P(M\cup C)=P(M)+P(C)-P(M\cap C)$$

$$=0.25+0.15-0.10=0.30$$

مثال$_2$: إذا كان الرجل A يصيب هدفاً باحتمال $\frac{1}{4}$ والرجل B يصيب نفس الهدف باحتمال $\frac{2}{5}$ ماهو احتمال اصابة الهدف اذا صوب A أو B نحو الهدف ؟

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

$$P(A\cup B)=\frac{1}{4}+\frac{2}{5}-\frac{1}{4}\times\frac{2}{5}$$

$$P(A\cup B)=\frac{5+8}{20}-\frac{2}{20}=\frac{13}{20}-\frac{2}{20}=\frac{11}{20}$$

طريقة اخرى للحل

$P(\bar{A})=1-p(A)=1-\frac{1}{4}=\frac{3}{4}$ احتمال ان الرجل A لايصيب الهدف

$P(\bar{B})=1-p(B)=1-\frac{2}{5}=\frac{3}{5}$ احتمال ان الرجل B لايصيب الهدف

$P(\bar{A}\cap\bar{B})=\frac{3}{4}\times\frac{3}{5}=\frac{9}{20}$ احتمال ان كليهما لايصيب الهدف

احتمال ان واحد منهم يصيب الهدف

$$P(A\cup B)=1-P(\overline{A\cup B})=1-P(\bar{A}\cap\bar{B})=1-\frac{9}{20}=\frac{20-9}{20}=\frac{11}{20}$$

مثال$_3$: إذا القي زهرة نرد مرة واحدة فما احتمال ظهور عدد يكون فرديا او يقبل القسمة على 3

A= الحادث عدد فردي وعدد الحالات المواتية {5} , {3} , {1}

$$P(A)=\frac{3}{6}$$

B= الحادث عدد يقبل القسمة على 3 وعدد الحالات المواتية {6} , {3}

$$P(B)=\frac{2}{6}$$

$A\cap B$= الحادث عدد فردي ويقبل القسمة على 3 وعدد الحالات الممكنة = 1

$$P(A\cap B)=\frac{1}{6}$$

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

$$P(A\cup B)=\frac{3}{6}+\frac{2}{6}-\frac{1}{6}$$

$$P(A \cup B) = \frac{5}{6} - \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$$

طريقة اخرى:

احتمال الحادث عدد غير فردي $P(\bar{A}) = 1 - \frac{3}{6} = \frac{1}{2}$

احتمال الحادث عدد لايقبل القسمة على 3 $P(\bar{B}) = 1 - \frac{2}{6} = \frac{2}{3}$

احتمال ان لايكون العدد فردي ولا يقبل القسمة على 3

$$P(\bar{A} \cap \bar{B}) = \frac{1}{2} \times \frac{2}{3} = \frac{2}{6}$$

احتمال ان يكون العدد فردي أو يقبل القسمة على 3

$$P(A \cup B) = 1 - \frac{2}{6} = \frac{4}{6} = \frac{2}{3}$$

مثال$_4$: إذا كان احتمال ان الطالب (A) يستطيع حل مسألة هو $\frac{4}{5}$ وان احتمال الطالب (B) يستطيع حل نفس المسألة هو $\frac{2}{3}$ واحتمال ان الطالب (C) يستطيع حلها هو $\frac{3}{7}$, فإذا ثلاثتهم حاولو حل نفس المسألة فما هو احتمال ان المسألة تحل ؟

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

$$P(A \cup B \cup C) = \frac{4}{5} + \frac{2}{3} + \frac{3}{7} - \left(\frac{4}{5} \times \frac{2}{3}\right) - \left(\frac{4}{5} \times \frac{3}{7}\right) - \left(\frac{2}{3} \times \frac{3}{7}\right) + \left(\frac{4}{5} \times \frac{2}{3} \times \frac{3}{7}\right) = \frac{101}{105}$$

الطريقة الثانية

$P(\bar{A}) = 1 - \frac{4}{5} = \frac{1}{5}$ احتمال ان A لا يستطيع حلها

$P(\bar{B}) = 1 - \frac{2}{3} = \frac{1}{3}$ احتمال ان B لا يستطيع حلها

$P(\bar{C}) = 1 - \frac{3}{7} = \frac{4}{7}$ احتمال ان C لا يستطيع حلها

احتمال ان جميعهم لا يحلونها

$$P(\bar{A} \cap \bar{B} \cap \bar{C}) = \frac{1}{5} \times \frac{1}{3} \times \frac{4}{7} = \frac{4}{105}$$

احتمال واحد منهم على الأقل يستطيع حلها

$$P(A \cup B \cup C) = 1 - \frac{4}{105} = \frac{101}{105}$$

2- قانون الضرب:

1- إذا كانت الأحداث مستقلة:

إذا كان E_1 و E_2 حادثين مستقلين فإن احتمال حدوثهما معاً هو حاصل ضرب احتمال كل منهما

$$P(E_1 \cap E_2)=P(E_1)\times P(E_2)$$

$$P(E_1 \cap E_2 \cap E_3)=P(E_1)\times P(E_2)\times P(E_3)$$

مثال$_1$: صندوقان الأول يحوي على 4 كرات بيضاء و2 سوداء والثاني يحتوي على 3 كرات بيضاء و5 سوداء فإذا سحبت كرة من كل منهما فما احتمال ان يكونا سوداوين ؟

نرمز الى احتمال ان تكون الكرة الأولى سوداء من الصندوق الأول $P(B1)$

نرمز الى احتمال ان تكون الكرة الثانية سوداء من الصندوق الثاني $P(B2)$

$$P(B1)=\frac{2}{6}$$

$$P(B2)=\frac{5}{8}$$

$$P(B1\cap B2)=P(B1)\times P(B2)$$

$$P(B1\cap B2)=\frac{2}{6}\times\frac{5}{8}=\frac{10}{48}=\frac{5}{24}$$

مثال$_2$: عند رمي قطعتي نقود ماهو احتمال الحصول على صور في كليهما ؟

نرمز لاحتمال الحصول على صورة في القطعة الاولى $P(F1)$

نرمز لاحتمال الحصول على صورة في القطعة الثانية $P(F2)$

$$P(F1\cap F2)=P(F1)\times P(F2)$$

$$P(F1\cap F2)=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$$

2- اذا كانت الأحداث غير مستقلة:

إذا كان E_1 و E_2 حدثين غير مستقلين فإن احتمال حدوثهما معاً يساوي حاصل ضرب احتمال وقوع الحادث الأول في احتمال وقوع الحدث الثاني مشروطاً بحدوث الأول

$$P(E_1 \cap E_2)=P(E_1)\times P(E_2/E_1)$$

$$P(E_1 \cap E_2 \cap E_3)=P(E_1)\times P(E_2/E_1)\times P(E_3/(E_1 \cap E_2))$$

مثال$_1$: صندوق به 5 كرات حمراء و3 سوداء فإذا سحبت كرتان سويا (او سحبت كرتان على التوالي بدون ارجاع الكرة الأولى الى الصندوق) فما هو احتمال ان تكون كلتاهما سوداء؟

الحل:

احتمال الحصول على كرة سوداء في السحبة الأولى = $\frac{3}{8}$

اما السحبة الثانية (بدون ارجاع الكرة المسحوبة الى الصندوق) فإن احتمال ان تكون الكرة سوداء هو

$$P(N2/N1) = \frac{2}{7}$$

$$P(N1 \cap N2)= P(N1).P(N2/N1) = \left(\frac{3}{8}\right)\left(\frac{2}{7}\right) = \frac{6}{56}$$

مثال$_2$: شعبة أ من الصف الاول في كلية الصيدلة تتألف من 25 طالباً و10 طالبات، فإذا اختير 3 اسماء عشوائياً فما هو احتمال ان يكونو من الذكور (من الطلاب)؟

$$P(M1\cap M2\cap M3)=P(M1)\times P(M2/M1)\times(M3/(M2\cap M1))$$

$$= \frac{25}{35}\times\frac{24}{34}\times\frac{23}{33}$$

$$= 0.71 \times 0.68 \times 0.7 = 0.34$$

5.2 الإحتمال الشرطي:

إذا كان A و B حادثين في فضاء العينة فإن احتمال وقوع A علماً ان B قد وقع ونرمز له (A/B) هو

$$P(A/B) = \frac{P(A\cap B)}{P(B)}$$

مثال$_1$: صندوق يحتوي 6 كرات حمراء و4 سوداء فإذا سحبت كرتان على التوالي (بدون ارجاع) ماهو احتمال ان تكون الكرة الأولى حمراء ايضاً؟

الحل:

$$P\left({}^{R_2}/_{R_1}\right) = \frac{P(R_1 \cap R_2)}{P(R_1)}$$

حيث ان $P(R_1 \cap R_2)$ هو احتمال الكرة الأولى والثانية حمراء

إن عدد اختيار كرتان حمراء $\binom{6}{2}$

إن عدد اختيار كرتان من الصندوق $\binom{10}{2}$

$$P\,(R_1 \cap R_2) = \frac{\binom{6}{2}}{\binom{10}{2}} = \frac{\text{عدد الحالات المؤاتية}}{\text{عدد الحالات الممكنة}}$$

$$P\,(R_1 \cap R_2) = \frac{\frac{6!}{2!(6-2)!}}{\frac{10!}{2!(10-2)!}}$$

$$P\,(R_1 \cap R_2) = \frac{\frac{6\times5\times4!}{2!\times4!}}{\frac{10\times9\times8!}{2\times1\times8!}} = \frac{\frac{15}{1}}{\frac{45}{1}} = \frac{15}{1} \times \frac{1}{45} = \frac{1}{3}$$

احتمال ان تكون الكرة الأولى حمراء = $\frac{6}{10}$

$$P\,(R_2/R_1) = \frac{\frac{1}{3}}{\frac{6}{10}} = \frac{1}{3} \times \frac{10}{6} = \frac{5}{9}$$

طريقة اخرى:

بما ان الكرة الأولى حمراء فإن عدد الكرات الحمراء الباقية = 5 ومجموع الكرات الكلية الباقية 9 فاحتمال الكرة الثانية حمراء = $\frac{5}{9}$

$$P\,(R_2/R_1) = \frac{5}{9}$$

مثال$_2$: صنف الشباب في مدينة تبسة كالآتي:

	له وظيفة	ليس له وظيفة	المجموع
ذكور (M)	460	40	500
اناث (F)	140	260	400
المجموع	600	300	900

فإذا اخذنا شاباً بصورة عشوائية فما هو احتمال ان يكون موظف ذكر ؟

نرمز للذكر = M

وللموظف = E

$$P\,(M/E) = \frac{P(M\cap E)}{P(E)} = \frac{\frac{460}{900}}{\frac{600}{900}}$$

$$P\,(M/E) = \frac{460}{900} \times \frac{900}{600} = \frac{460}{900}$$

$$=\frac{23}{30}$$

الإحتمال والتحليل التوفيقي:

إن التحليل التوفيقي يسهل كثيراً حساب درجة الإحتمال في كثير من الأحيان كما في الأمثلة التالية:

مثال: صندوق يحتوي على 8 كرات حمراء و3 بيضاء و9 سوداء فإذا سحبت ثلاث كرات عشوائياً احسب احتمال

أ- ثلاثتهم حمراء

$$P(R_1 \cap R_2 \cap R_3) = P(R_1) \times P\left({}^{R_2}/_{R_1}\right) \times P\left({}^{R_3}/_{R_2 \cap R_1}\right)$$
$$= \frac{8}{20} \times \frac{7}{19} \times \frac{6}{18} = \frac{14}{285}$$

الطريقة باستخدام التحليل التوفيقي

$$P(R_3) = \frac{\binom{8}{3}}{\binom{20}{3}} = \frac{14}{285}$$

ب- ثلاثتهم بيضاء

$$P(B_3) = \frac{\binom{3}{3}}{\binom{20}{3}} = \frac{1}{285}$$

جـ- 2 حمراء و 1 بيضاء

$$P(R_2 \cap B_1) = \frac{\binom{8}{2}\binom{3}{1}}{\binom{20}{3}} = \frac{7}{95}$$

د- على الأقل 1 بيضاء. نضع

$$P(\overline{W}) = \frac{\binom{17}{3}}{\binom{20}{3}} = \frac{34}{57}$$

P(على الأقل واحدة بيضاء)$=1- P(\overline{W}) = 1-\frac{34}{57} = \frac{23}{57}$

هـ- واحدة من كل لون

$$P(R_1 \cap B_1 \cap N_1) = \frac{\binom{8}{1}\binom{3}{1}\binom{9}{1}}{\binom{20}{3}} = \frac{18}{95}$$

و- الكرات سحبت حسب الترتيب التالي حمراء ثم بيضاء ثم سوداء

$$P(R_1 \cap B_2 \cap N_3) = P(R_1) \times P(B_2/R_1) \times (N_3/R_1 \cap B_2)$$
$$= \frac{8}{20} \times \frac{3}{19} \times \frac{9}{18} = \frac{3}{95}$$

قاعدة بايز:

تظهر في الظواهر العشوائية الآنية حيث الحظ يتدخل مرتين.

تعريف: لتكن Ω مجموعة، نقول عن A_1،A_2،.....، A_n أنها تشكل تجزئة لـ Ω إذا:

$$-\forall i \in \{1, \ldots, n\} \ , \ \ A_i \neq \emptyset$$

$$-\ \forall\ i \neq j \ \ , \ A_i \cap A_j = \emptyset$$

$$-\ A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$$

باعتبار ان $(A_i)_{i=1,\ldots,n}$ مجموعة من الحوادث و Ω هو الحدث الأكيد و ϕ الحدث المستحيل ، صيغة بابيز تعطى بالشكل:

$$P\left({}^{A_i}/_{B}\right) = \frac{P\left({}^{B}/_{A_j}\right).P(A_j)}{\sum_{i=1}^{n} P\left({}^{B}/_{A_i}\right).P(A_i)}$$

مثال: لدينا 100 قطعة زهرة النرد، من بينها 25 مزورة. إذا علمت ان:

إذا كان النرد مزور فاحتمال ان نحصل على رقم 6 (في عملية رمي النرد) يساوي $\frac{1}{2}$.

نأخذ بصفة عشوائية قطعة من النرد ثم نرمي هذه القطعة ونحصل على رقم 6، ما هو احتمال ان يكون هذا النرد مزور؟

نسمي T الحدث :" النرد مزور"

و الحدث A_1 : " نحصل على رقم 6 بعد الرمية الأولى "

عندئذ الإحتمال المطلوب حسابه هو : $P\left({}^{T}/_{A_1}\right)$

نعلم أن كل من T و $\overline{T}$ يشكلان تجزئة ل Ω ، بإستعمال صيغة بابيز نجد:

$$P\left({}^{T}/_{A_1}\right) = \frac{P\left({}^{A_1}/_{T}\right).P(T)}{P\left({}^{A_1}/_{T}\right).P(T) + P\left({}^{A_1}/_{\overline{T}}\right).P(\overline{T})} = \frac{\frac{1}{2}\cdot\frac{1}{4}}{\frac{1}{2}\cdot\frac{1}{4} + \frac{3}{4}\cdot\frac{1}{6}} = \frac{1}{2}$$

6.2 المتغيرات العشوائية:

بعد دراسة كيفية لظاهرة عشوائية (غير كافية) يجب دراسة كمية نحتاج فيها إلى معطيات عددية كي نصف الظاهرة العشوائية.

مثلا: لتكن ξ ظاهرة عشوائية ، وليكن $\Omega \supset \mathrm{A}$ حادث و $\omega \in \Omega$ حادث أولي.

نعرف تابع ξ_{A} كمايلي :

$$\xi_{\mathrm{A}} : \Omega \to \mathbb{R}$$

$$\forall \omega \in \Omega \,;\; \xi_{\mathrm{A}}(\omega) = \begin{cases} 1, & \omega \in \mathrm{A} \\ 0, & \omega \notin \mathrm{A} \end{cases}$$

أي ان $\xi_{\mathrm{A}} = 1$ عندما يتحقق A ، $\xi_{\mathrm{A}} = 0$ عندما يتحقق $\overline{\mathrm{A}}$.

المتغير العشوائي الحقيقي X:

نسمي متغير عشوائي حقيقي X كل تطبيق يرفق بكل عنصر ω من Ω عدد حقيقي x .

$$X : \Omega \longrightarrow \mathbb{R}$$

$$\omega \longmapsto x = X(\omega)$$

- نسمي ميدان تغير (أو حامل) لـ X المجموعة $D_X \subseteq \mathbb{R}$ مجموعة القيم التي يمكن ان يأخذها المتغير العشوائي X . في الحالة العامة نسمي X المتغير العشوائي وx مخرجه.
- المتغير العشوائي إما متقطع أي ان ميدان تغيره يحوي عدد منته او غير منته معدود من القيم.
- او متغير عشوائي مستمر وهو الذي ميدان تغيره يحوي قيم غير منتهية غير معدودة.

تابع التوزيع:

تعريف: نسمي تابع التوزيع التابع F معرف كمايلي:

$$\begin{aligned} F \,:\, D_X &\longrightarrow [0\,,1] \\ x &\longmapsto P(X \leq x) \end{aligned}$$

خواص:

1- $F(x) \in [0\,,1]$

2- F تابع متزايد

3- $\lim_{x \to -\infty} F(x) = 0$

4- $\lim_{x \to +\infty} F(x) = 1$

5- من اجل متغير عشوائي متقطع F هو تابع درجي.

6- من اجل متغير عشوائي مستمر F هو تابع مستمر.

كثافة الإحتمال:

تعريف: كثافة احتمال (او التوزيع) هي تابع معرف كمايلي:

$$f(x) = \frac{d}{dx}F(x) = \frac{d}{dx}P(X \leq x) = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x)}{\Delta x}$$

خواص:

1- $f(x) \geq 0$

2- $\int_{-\infty}^{+\infty} f(x)dx = 1$

3- من اجل $a < b$

$$P(a < X \leq b) = \int_{a}^{b} f(x)dx = F(b) - F(a)$$

التوقع الرياضي:

تعريف: التوقع الرياضي $E(X)$ للمتغير العشوائي X يعرف بـ :

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

- التوقع ليس بالضرورة جزء من D_X.
- التوقع ليس معرف دائما.

خواص:

1- $E(a + bX) = a + bE(X)$

2- $E(X + Y) = E(X) + E(Y)$

التباين و الإنحراف:

تعريف: التباين $Var(X)$ للمتغير العشوائي X يعرف بـ :

$$Var(X) = E\big(X - E(X)\big)^2$$

- الانحراف المعياري σ يعرف بـ : $\sigma = \sqrt{Var(X)}$

ملاحظة:

$$\begin{aligned} Var(X) &= E(X^2 - 2XE(X) + E(X)^2) \\ &= EX^2 - 2E(X)E(X) + E(X)^2 \\ &= EX^2 - E(X)^2 \end{aligned}$$

خواص:

- مثل التوقع التباين ليس موجود دائما.
- $Var(X) \geq 0$
- $Var(a + bX) = b^2 Var(X)$

7.2 بعض التوزيعات المألوفة:

1.7.2 القانون الطبيعي: (قانون غوص)

تعريف: نقول عن متغير عشوائي X انه يتبع القانون الطبيعي بوسيطين μ و σ^2 إذا:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

ونرمز : $X \sim \mathcal{N}(\mu , \sigma^2)$

تابع التوزيع:

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}} dt$$

التوقع الرياضي والتباين:

-إذا كان X متغير عشوائي بحيث $X \sim \mathcal{N}(\mu , \sigma^2)$ عندئذ :

$$a + bX \sim \mathcal{N}(a + b\mu , b^2\sigma^2)$$

-إذا كان X و Y متغيران عشوائيان مستقلان بحيث :

$$X \sim \mathcal{N}(\mu_1 , \sigma_1^2)$$

$$Y \sim \mathcal{N}(\mu_2 , \sigma_2^2)$$

عندئذ:

$$X + Y \sim \mathcal{N}\left(\mu_{1+\mu_2} , \sigma_1^2 + \sigma_2^2\right)$$

القانون الطبيعي المركز المختزل:

تعريف: نسمي قانون طبيعي مركز مختزل القانون $\mathcal{N}(0 , 1)$ مع تابع الكثافة :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

خاصية: إذا $X \sim \mathcal{N}(\mu , \sigma^2)$ عندئذ : $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0 , 1)$

2.7.2 قانون χ^2:

تعريف: لتكن $X_1, \dots , X_n$ متغيرات عشوائية مستقلة تتبع نفس القانون الطبيعي المركز المختزل عندئذ المتغير العشوائي :

$Y = X_1^2 + \cdots + X_n^2$ يتبع قانون مستمر يسمى قانون χ^2 بـ n درجة من الحرية ونكتب :

$$Y = \sum_{i=1}^{n} X_i^2 \sim \chi_n^2$$

خواص:

1- إذا $Y_1 \sim \chi_{n_1}^2$ و $Y_2 \sim \chi_{n_2}^2$ مع Y_1 و Y_2 متغيران عشوائيان مستقلان عندئذ:

$$Y_1 + Y_2 \sim \chi_{n_1+n_2}^2$$

2- إذا $Y \sim \chi_n^2$ عندئذ : $E(Y) = n$ و $Var(Y) = 2n$

وتابع الكثافة:

$$f(y) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{(n-2)}{2}} e^{-\frac{x}{2}}$$

3.7.2 قانون ستودنت: (Student)

ليكن X و Y متغيران عشوائيان مستقلان بحيث: $X\sim\mathcal{N}(0\,,1)$ و $Y\sim\chi_n^2$

عندئذ المتغير العشوائي $T=\frac{X}{\sqrt{Y/n}}$ يتبع قانون مستمر يسمى قانون ستودنت بـ n درجة من الحرية ونكتب

$$T=\frac{X}{\sqrt{Y/n}}\sim t_n$$

خواص:

- $E(T)=0$
- $Var(T)=\frac{n}{n-2}$ إذا $n\geq 2$
- تابع الكثافة:

$$f(t)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\left(\frac{n}{2}\right)\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

8.2 الإختبارات الإحصائية:

الفرضية الإحصائية:

يفترض على الباحث ان يضع الفرضية الإحصائية لاختيارها قبل البدء بتنفيذ التجربة، والفرضية الإحصائية عبارة عن ادعاء او تصريح قد يكون صائباً او خطأ حول معلمة (صفة) او أكثر لمجتمع او مجموعة من المجتمعات، وهي نوعان:

1- **فرضية العدم**: يرمز لها بالرمز H_0 وهي التي تفترض عدم وجود فروق معنوية بين المتوسطات للمعاملات اي ان $M_1= M_2$

2- **الفرضية البديلة**: ويرمز لها بالرمز H_1 وهي التي تنص عن وجود فروقات معنوية بين متوسطات المعاملات اي ان $M_1\neq M_2$

ولذلك فإن الباحث او الإحصائي دائماً يحاول ان يضع الفرضية بشكل يأمل ان يرفضها فمثلاً إذا اراد باحث ان يقارن بين عقار مصنع محلياً مع عقار مصنع خارج الجزائر في فعاليتهما في علاج مرض فإنه يضع فرضية فحواها بأنه لا توجد فروقات جوهرية او معنوية بين العقارين في فعاليتهما في علاج المرض، وهكذا الفرضية التي يضعها الباحث على امل ان يرفضها تدعى فرضية العدم يقودنا إلى قبول فرضية بديلة، وعند رفض فرضية العدم وهي صحيحة تقع في خطأ من النوع الأول ،اما إذا قبلنا فرضية العدم وهي خطأ نقع في خطأ من النوع الثاني ، وان خطأ القبول او الرفض للفرضيات الموضوعة يكون بدرجة احتمال او تسمى مستوى المعنوية والتي يرمز لها بالرمز (α) وهي 1% أو 5% ومستوى المعنوية (هي احتمال ان ترفض فيها فرضية العدم وهي صحيحة) ويكون اتخاذ القرار بدرجة احتمال 1% اقوى وبثقة اكبر وهذا يعني ان إعادة التجربة مئة مرة يكون احتمال ارتكاب الخطأ في اتخاذ القرار مرة واحدة اي اننا نرفض فرضية العدم وهي صحيحة مرة واحدة ، واتخاذ القرار بمستوى 5% يحتمل ان تخطيء خمس مرات برفضنا فرضية العدم وهي صحيحة .

الإختبارات الإحصائية:

تستخدم عدة طرق إحصائية لمعرفة الفروقات بين تأثير معاملة واخرى إضافة إلى طرق التصميم المتبعة ولا تقل هذه الإختبارات الإحصائية في الأهمية في التحليل والإستنتاج عن طريق تصميم التجارب وهي الطرق الإحصائية ذات الإستخدام الواسع في مجال علوم الحياة والعلوم الأخرى.

1.8.2 اختبار -*t* :

كتب احد باحثي الإحصاء في بداية القرن العشرين المدعو *William Gossat* تحت اسم مستعار *Student* احد بحوثه الإحصائية عن هذه الطريقة استنبط فيها طريقة لفحص الإحصائية بإ ستخدام قياسات محسوبة (S^2 و $\overline{X}$) من العينات والمتغيرات وهذه الطريقة عبارة عن اختبار *t* ويقسم اختبار *t* إلى :

1.1.8.2 اختبار *t* يتعلق بمتوسط واحد

$$t = \frac{\overline{X}-M}{S_{\overline{X}}}$$

t = المحسوبة , M = متوسط المجتمع , $\overline{X}$ = متوسط العينة , $S_{\overline{X}}$ الخطأ القياسي للعينة.

$$S_{\overline{X}} = \frac{S}{\sqrt{n}}$$

2.1.8.2 اختبار t يتعلق بمتوسطين

$$t = \frac{\overline{X}_1-\overline{X}_2}{S_{(\overline{X}_1-\overline{X}_2)}}$$

حيث t= المحسوبة , $S_{(\overline{X}_1-\overline{X}_2)} = \sqrt{\frac{2\ mse}{r}}$ الخطأ القياسي للفرق بين متوسطين , r= عدد التكرارات و

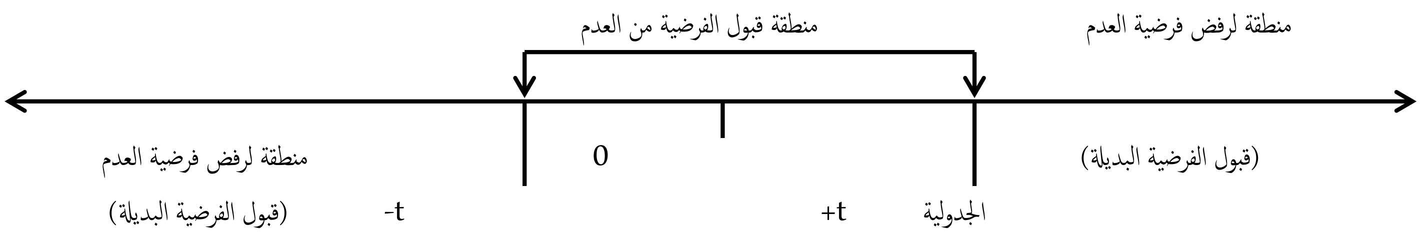
$$mse = S_p^2 = \frac{SS_1 + SS_2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

حيث تمثل t إنحراف معدل العينة عن معدل المجتمع مقسوماً على الإنحراف القياسي او المعياري للمعدلات، ويستخدم للاستدلال فيما إذا كان انحراف معدل العينة عن معدل المجتمع طبيعياً او غير اعتيادي، إذ من المفروض ان المشاهدات تتوزع توزيعاً طبيعياً حول المجتمع الذي اخذت منه، كذلك فإن معدلات العينات لها توزيع طبيعي حول المجتمع.

تحسب قيمة t من العينة بصورة مباشرة وتقارن مع t الجدولية والتي على اساسها يتم قبول او رفض الفرضيات الموضوعة ، فإذا كانت قيمة t المحسوبة اكبر او تساوي قيمتها في الجدول (t الجدولية) لمستوى المعنوية المطلوب للاختبار عليه ودرجة الحرية (n-1) تعتبر في هذه الحالة العينة غير ممثلة للمجتمع , كذلك يستخدم اختبار t لمقارنة معدلين او متوسطين من عينتين إذا كانت هاتان العينتان تعودان لنفس المجتمع ام لا سواء كانت هاتان العينتان متساويتان في عدد المشاهدات مزدوجة متساوية او غير متساوية (غير مزدوجة) وفي هذه الحالة نستخدم :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{(\bar{X}_1 - \bar{X}_2)}}$$

كذلك نستخدم t لاستخراج الفرق المعنوي الأصغر L.S.D $\quad$ $L.S.D = t(\alpha , v)\sqrt{\frac{2mse}{r}}$



1- اختبار يتعلق بمتوسط واحد

مثال$_1$: اشار سجل مستشفى الولادة لمدينة تبسة بأن معدل وزن الأطفال عند الولادة للسنين الماضية هو 5.5 كغم، اخذت عينة عشوائية في الربع الأول من هذه السنة مؤلفة من 30 طفل وكان معدل وزنهم في تلك السنة 5.1 كغم وبانحراف قياس قدره 0.9 كغم. فهل هناك فرق معنوي في وزن الأطفال في هذه السنة عما هو معروف في السنين الماضية؟ اختبر ذلك تحت مستوى احتمال 0.01 علماً ان قيمة t الجدولية = 2.756؟

خطوات الاختبار

1- وضع الفرضيات $\quad H_0 : M_1 = 5.5$

$H_1 : M_1 \neq 5.5$

2- اختبار الفرضية

$$t=\frac{\bar{X}-M}{\frac{s}{\sqrt{n}}}$$

$$t_{cal}=\frac{5.1-5.5}{\frac{0.9}{\sqrt{30}}}=-2.434$$

3- استخراج قيمة t الجدولية لمستوى معنوية $\alpha = 0.01$ ودرجة حرية =29

t_{abl} الجدولية = 2.756

4- الإستنتاج: - بما ان القيمة المطلقة ل (t_{cal} المحسوبة = 2.434 -) اقل من t_{tabl} الجدولية لذا نقبل فرضية العدم H_0 أي ان لا يوجد فرق معنوي بين اوزان الأطفال عند الولادة في هذه السنة عما هو في السنين الأخرى.

مثال$_2$: كان متوسط الزيادة لوزن 12 فأرة بعد تغذيتها بغذاء يحتوي على 1% مضاد حيوي هو 148.54 غم وبانحراف قياسي للوسط الحسابي 2.3غم. ففي مستوى احتمال 0.05 هل يمكن القول بأن الزيادة في الوزن نتيجة التغذية على هذا الغذاء لا تقل عن 150غم علماً ان قيمة t الجدولية 2.201.

1- وضع الفرضيات

$$H_0 : M_1 \geq 150$$

$$H_1 : M_1 < 150$$

2- اختبار الفرضية

$$t=\frac{\bar{X}-M}{\frac{s}{\sqrt{n}}}$$

$$t_{cal}=\frac{148.54-150}{\frac{2.3}{\sqrt{12}}}=-2.199$$

استخراج t الجدولية من اجل $\alpha = 0.05$ ودرجة حرية11

$2.201=t_{tabl}$

4- الإستنتاج: بما ان القيمة المطلقة ل t_{cal} المحسوبة (2.199-) اقل من قيمتها الجدولية 2.201 (t_{tabl} الجدولية) نقبل فرضية العدم ونرفض الفرضية البديلة اي معدل الزيادة في الوزن لا تقل عن 150غم.

مثال$_3$: ادعت احدى شركات انتاج السجائر بأن نسبة النيكوتين في انتاجها من السجائر لا يتجاوز 17.5ملغم، اخذت عينة عشوائية مؤلفة من 9 سجائر وقيست نسبة النيكوتين فيها فكانت كالاتي:

X_i= 18, 18 ,16, 20 ,19, 19 ,18, 18, 17

فهل ادعاء الشركة صحيح تحت مستوى 0.05 ؟علماً ان t الجدولية تحت مستوى 0.05 تساوي 2.306 .

1- ضع الفرضيات

$$H_0 : M \leq 17.5$$

$$H_1 : M > 17.5$$

2- اختبار الفرضية

$$t=\frac{\overline{X}-M}{S_{\overline{X}}}$$

$$\sum_{i=1}^{9} X_i = 163 \qquad \sum_{i=1}^{9} X_i^2 = 2963$$

$$\overline{X} = \frac{\sum_{i=1}^{9} X_i}{9} = \frac{163}{9} = 18.1$$

$$SS = \sum_{i=1}^{9}\left(X_i - \overline{X}\right)^2 = \sum_{i=1}^{9} X_i^2 - \frac{(\sum_{i=1}^{9} X_i)^2}{9}$$

$$SS= 2963 - \frac{(163)^2}{9} = 2963 - 2959.11 = 10.89$$

$$S^2=\frac{SS}{d.f} = \frac{10.89}{9-1} = \frac{10.89}{8} = 1.36$$

$$S = \sqrt{S^2} = \sqrt{1.36} = 1.17$$

$$t_{cal}= \frac{18.1-17.5}{\frac{1.17}{\sqrt{9}}}=1.5385$$

3- استخراج قيمة t الجدولية لمستوى معنوية $\alpha = 0.05$ ودرجة حرية 8

$2.306=T_{tabl}$

3- الإستنتاج: بما ان قيمة t_{cal} المحسوبة 1.5385 اقل من t_{tabl} الجدولية 2.306 لذا نقبل فرضية العدم ونرفض الفرضية البديلة اي ان ادعاء الشركة صحيح.

2- اختبار يتعلق بمتوسطين

أ- **اختبار t للعينات المستقلة:** هناك العديد من الإفتراضات التي يقوم عليها اختبار t للعينات المستقلة:

1- ان العينتين تم اختيارها بشكل عشوائي من المجتمع الخاص لكل عينة.
2- ان المجتمعان يتصفان بالسواء (التوزيع الطبيعي).
3- الملاحظات، البيانات، المشاهدات، ضمن كل عينة مستقلة عن بعضها.
4- العينات تم توزيعها بشكل عشوائي إلى المجموعتين.
5- لغرض تحديد العينتان متجانسة او غير متجانسة يجري استخدام فحص التجانس وعلى النحو التالي:

$$\text{المحسوبة } F_{cal} = \frac{S^2\ larges}{S^2\ smallest}$$

$$\text{المحسوبة } F_{cal} = \frac{S^2\ L}{S^2\ s}$$

$$\text{المحسوبة } F_{cal} = \frac{\text{التباين الاكبر}}{\text{التباين الاصغر}}$$

تقارن F_{cal} المحسوبة مع F_{tabl} الجدولية بدرجة حرية التباين الأكبر بالاتجاه الأفقي وبدرجة حرية التباين الأصغر بالاتجاه العمودي، فإذا كانت F_{cal} المحسوبة اصغر من F_{tabl} الجدولية فهناك تجانس العينتان وإذا كانت F_{cal} المحسوبة اكبر من F_{tabl} الجدولية فهناك عدم وجود تجانس العينتان .

اولاً: في حالة التجانس

مثال: في تجربة لمقارنة نسبة المواد الفعالة التي تستخدم في صنع العقاقير في صنفين من نبات الكزبرة الصنف المحلي والصنف الباكستاني تم اختيار 12 نباتاً من كل صنف وقدرت نسبة المواد الفعالة فيها وكانت النتائج كما يأتي فهل يختلف الصنفان تبعاً لنسبة المادة الفعالة تحت مستوى احتمال 0.05 ؟ علماً ان قيمة t_{tabl} الجدولية تساوي 2.075 تحت مستوى احتمال 0.05 ودرجة حرية 22.

الصنف المحلي ملغم	الصنف الباكستاني ملغم
12.5	9.4
9.4	8.4
11.7	11.6

7.2	11.3
9.7	9.9
7.0	8.7
10.4	9.6
8.2	11.5
6.9	10.5
12.7	10.6
7.3	9.6
9.2	9.7
$\sum_{i=1}^{12} X_i = 108$ $\bar{X}=9$	$\sum_{i=1}^{12} X_i = 124.8$ $\bar{X}= 10.4$

الحل:

1- اجراء اختبار التجانس

$$F_{cal} = \frac{S^2\ L}{S^2\ s} = \frac{\text{التباين الاكبر}}{\text{التباين الاصغر}}$$

$$S^2 = \frac{SS}{d.f}$$

$$S_1^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}{n-1} = \frac{1316.2 - \frac{(124.8)^2}{12}}{12-1}$$

$$S_1^2 = \frac{1316.2-1297.92}{11} = \frac{18.3}{11} = 1.66$$

$$S_2^2 = \frac{1010.64-972}{11} = \frac{38.64}{11} = 3.51$$

المحسوبة $F_{cal} = \frac{3.51}{1.66} = 2.11$

قيمة F_{tabl} الجدولية = 2.82

بدرجة حرية بالاتجاه الأفقي = 11

بدرجة حرية بالاتجاه العمودي = 11

بما ان F_{cal} المحسوبة أصغر من F_{tabl} الجدولية: العينتان متجانستان

2- وضع الفرضيات $H_0: M_1 - M_2 = 0$

$H_1: M_1 - M_2 \neq 0$

اختبار الفرضيات

$$t_{cal} = \frac{\bar{X}_1 - \bar{X}_2}{S_{(\bar{X}_1 - \bar{X}_2)}}$$

$\bar{X}_1$= الوسط الحسابي للعينة الأولى

$\bar{X}_2$ = الوسط الحسابي للعينة الثانية

الخطأ القياسي للفرق = $S_{(\bar{X}_1 - \bar{X}_2)}$

متوسط معاملتين

$$S_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\frac{2\,mse}{r}}$$

التباين المشترك $= S_p^2 = mse$

$$S_p^2 = \frac{SS_1 + SS_2}{n_1 + n_2 - 2}$$

او $$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$

حسب القانون الأول

$$S_p^2 = \frac{14.08+38.64}{12+12-2} = \frac{52.72}{22} = 2.40$$

$$S_{(\bar{X}_1-\bar{X}_2)} = \sqrt{\frac{2\ mse}{r}} = \sqrt{\frac{2 \times 2.40}{12}} = 0.63$$

$$t = \frac{\bar{X}_1-\bar{X}_2}{S_{(\bar{X}_1-\bar{X}_2)}} = \frac{10.4-9.0}{0.63} = 2.2$$

ويمكن استخدام القانون التالي لإيجاد t_{cal} المحسوبة

$$t_{cal} = \frac{\bar{X}_1-\bar{X}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

S_p هوالإنحراف المعياري المشترك للفرق بين معاملتين

$$S_p = \sqrt{S_p^2} = \sqrt{mse}$$

$$S_p = \sqrt{2.40} = 1.55$$

$$t_{cal} = \frac{10.4-9}{1.55\sqrt{\frac{1}{12}+\frac{1}{12}}} = 2.2$$

3- استخراج t_{tabl} الجدولية لمستوى معنوية مطلوبة ودرجة حرية $n_1 + n_2 - 2$ ومستوى المعنوية المطلوب 0.05

حيث كانت t_{tabl} الجدولية = 2.074

4- الإستنتاج: بما ان t_{cal} المحسوبة أكبر من t_{tabl} الجدولية لذا نرفض فرضية العدم ونقبل الفرضية البديلة وبما ان معدل المواد الفعالة في الصنف المحلي اعلى من الصنف الباكستاني لذا نوصي بإستخدام الصنف المحلي لأستخلاص المواد الفعالة.

ثانياً: في حالة عدم التجانس

مثال: اراد احد المدرسين في كلية الصيدلة ان يدرس اثر طريقتين من التدريس هما طريقة A و B في التعليم المختبري على التحصيل عند عينة من الطلبة اللذين يعانون من مشكلات تحصيلية في درس كيمياء الأدوية فأختار عينة مؤلفة من 20 طالباً قام بتوزيعهم بشكل عشوائي الى مجموعتين 10 طلاب لكل مجموعة ثم عرض المجموعة الأولى للطريقة A وعرض المجموعة الثانية للطريقة B وبعد ذلك طبق عليهم امتحان تحصيلياً في التعليم المختبري وحصل على البيانات التالية علماً ان النقطة القصوى للامتحان 30 نقطة , فهل هناك فرق معنوي بين الطريقتين على امتحان التحصيل للمادة العلمية اختبر ذلك تحت مستوى احتمال 0.01 ؟

المجموعة B	المجموعة A
12	06
13	05
28	04
05	04
10	07
18	04
23	05
06	05
04	07
30	06
149	53

الحل:

$$\overline{X}_A = \frac{53}{10} = 5.3$$

$$\overline{X}_B = \frac{149}{10} = 14.9$$

$$S_A^2 = \frac{293 - 280.9}{9} = \frac{12.1}{9} = 1.34$$

$$S_B^2 = \frac{3027 - 2220.1}{9} = 89.7$$

$$F_{cal} \text{ المحسوبة} = \frac{89.7}{1.34} = 66.94$$

$F_{tabl}=3.35$ الجدولية

بما ان F_{cal} المحسوبة اكبر من F_{tabl} الجدولية : العينتين غير متجانستين

1- وضع الفرضيات $H_0: M_1 - M_2 = 0$

$H_1 : M_1 - M_2 \neq 0$

2- اختبار الفرضية

$$t_{cal} = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} = \frac{5.3-14.9}{\sqrt{\frac{1.34}{10}+\frac{89.7}{10}}} = -3.18$$

$$d.f = \frac{\left(\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}\right)^2}{\frac{\left(\frac{S_A^2}{n_A}\right)^2}{n_A-1} + \frac{\left(\frac{S_B^2}{n_B}\right)^2}{n_B-1}}$$

$$d.f = \frac{\left(\frac{1.34}{10} + \frac{89.7}{10}\right)^2}{\frac{\left(\frac{1.34}{10}\right)^2}{9} + \frac{\left(\frac{89.7}{10}\right)^2}{9}} = \frac{(9.104)^2}{0.001+8.940} = \frac{82.88}{8.94}$$

$$d.f = 9.27 \simeq 9$$

t_{tabl} الجدولية عند مستوى 0.01 ودرجة حرية 9

$3.25 = T_{tabl}$

بما ان t_{cal} المحسوبة اصغر من t_{tabl} الجدولية نقبل فرضية العدم ونرفض البديلة اي لا يوجد فرق معنوي بين الطريقتين على امتحان التحصيل للمادة العلمية لمادة كيمياء الأدوية .

ب - اختبار t للعينات المرتبطة

يستخدم اختبار t في هذه الحالة لاختبار الفروقات بين معاملتين (صفتين) مطبقة على نفس العينة او نفس الوحدة التجريبية وتشكل البيانات بشكل ازواج تسجل قبل وبعد تطبيق المعاملة، غالباً ما تستخدم في الدراسات الطبية وفي هذه الحالة نستخرج الفروقات بين ازواج المشاهدات وتعامل كعينة واحدة.

مثال$_1$: في تجربة لدراسة تأثير غذاء معين مع برنامج لإجراء بعض التمارين الرياضية لتقليل مستويات الكوليسترول بالدم، طبق هذا البرنامج وكانت النتائج كالآتي، اختبر ذلك تحت مستوى احتمال 0.01

طالب	نسبة الكوليسترول قبل البرنامج	نسبة الكوليسترول بعد البرنامج	الفرق d_i
1	201	200	1
2	236	231	5
3	221	216	5
4	260	233	27
5	228	224	4
6	237	216	21
7	326	296	30
8	235	195	40
9	240	207	33
10	267	247	20
11	284	210	74
12	218	210	8
	2953	2685	268

الحل:

1- وضع الفرضيات

أ- ان الفرق بين مستوى الكولسترول قبل وبعد تطبيق البرنامج أكبر او يساوي صفر

$H_0: M \geq 0$

ب- ان الفرق بين مستوى الكولسترول قبل وبعد تطبيق البرنامج اصغرمن الصفر

$H_1: M < 0$

2- اختبار الفرضيات

$$t = \frac{\acute{d}}{S_d / \sqrt{n}}$$

$$\acute{d} = \frac{\sum d_i}{n} = \frac{268}{12} = 22.33$$

$$S_d^2 = \frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n-1} = \frac{(1)^2+(5)^2+\cdots+(8)^2 - \frac{(268)^2}{12}}{11}$$

$$S_d^2 = \frac{4780.67}{11} = 434.61$$

$$S_d = \sqrt{434.61} = 20.85$$

$$t_{cal} = \frac{\acute{d}}{S_d / \sqrt{n}} = \frac{22.33}{20.85 / \sqrt{12}} = 3.70$$

3- استخراج t_{tabl} الجدولية لمستوى المعنوية 0.01 ودرجة حرية 11 والتي تساوي 3.11

4- الإستنتاج: بما ان t_{cal} المحسوبة 3.70 أكبر من t_{tabl} الجدولية 3.11 لذا نقبل الفرضية البديلة ونرفض فرضية العدم اي ان الفرق بين مستوى الكولسترول قبل وبعد تطبيق البرنامج أصغر من صفر لذا كان البرنامج فعالا في تقليل مستوى الكولسترول بالدم.

مثال$_2$: اراد احد الباحثين الأطباء ان يعرف فيما إذا كان متوسط ضغط الدم في الإنسان يختلف في حالة قياسه والشخص معتدل القامة عنه في حالة استلقاء الشخص نفسه على ظهره. فأخذ عينة عشوائية مؤلفة من 12 شخص والنتائج التالية تبين الفرق بين ضغط الدم وهو في حالة وقوفه وضغطه وهو في حالة استلقاء على ظهره فماذا كان قراره تحت مستوى 0.05؟

d_i= -4, 1, 1, -5, -6, -3, 2, -9, 1, -4, -7, -7

الحل:

1- وضع الفرضيات

$H_0: M_1 - M_2 = 0$

$$H_1 : M_1 - M_2 \neq 0$$

2- اختبار الفرضية

$$t_{cal} = \frac{\acute{\bar{d}}}{S_d / \sqrt{n}}$$

$$\acute{\bar{d}} = \frac{\sum d_i}{n} = \frac{-40}{12} = -3.33$$

$$S_d^2 = \frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n-1} = 14.06$$

$$S_d = \sqrt{S_d^2} = \sqrt{14.06} = 3.57$$

$$t_{cal} = \frac{-3.33}{3.57/\sqrt{12}} = -3.23$$

3- استخراج قيمة t_{tabl} الجدولية لمستوى 0.05 ودرجة حرية 11 وهي 2.201

4- الإستنتاج: بما ان القيمة المطلقة ل t المحسوبة أكبر من t الجدولية لذا نرفض فرضية العدم ونقبل الفرضية البديلة اي ان معدل ضغط الفرد يكون اعلى والشخص مستلقي على ظهره من ضغطه وهو في حالة الإعتدال والوقوف.

2.8.2 اختبار كاي دو (χ^2):

اختبار χ^2 شائع الإستخدام مع البيانات العددية المتقطعة والتي تكون نوعية اكثر منها كمية , كعدد الذكور والإناث في عينة ،عدد الأصحاء وعدد المرضى في مجتمع او عدد الأحياء والأموات او الإجابة بنعم او بدون نعم .

يعتبر توزيع χ^2 من التوزيعات المستمرة ويعتمد على التوزيع الطبيعي في حين تكون التوزيعات التكرارية غير مستمرة لذا يكون اختبار التكرارات المشاهدة مع التكرارات النظرية (المتوافقة) ذات دقة تقريبية .

$$\chi^2 = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i}$$

حيث o_i قيم المشاهدات المشاهدة

وحيث e_i قيم المشاهدات النظرية (المتوافقة)

ملاحظة: تكون قيم χ^2 صغيرة عندما تكون قيم المشاهدات المتوافقة قريبة جداً من قيم المشاهدات المشاهدة او الواقعة وكذلك لا تكون قيم χ^2 سالبة .

وتقارن قيم χ^2 الجدولية والتي تستخرج على اساس درجات الحرية ومستوى المعنوية المطلوبة فإذا كانت قيمة χ^2 المحسوبة اكبر او تساوي χ^2 الجدولية نرفض فرضية العدم ونقبل الفرضية البديلة اي هناك فروقات معنوية .

استخدامات اختبار χ^2 :

1.2.8.2 استخدام (χ^2) لجودة المطابقة :

اي المطابقة بين القيم المشاهدة (او الملاحظة) والقيم النظرية (المتوافقة) وهذا الإختبار يقوم على اساس ان القيم المشاهدة (او الملاحظة) لها نفس توزيع القيم النظرية كما ويفيد بصورة خاصة لاختبار البيانات الوراثية لجودة تطابق انعزالات الجيل الثاني.

مثال$_1$: إذا كان عدد الذكور في مرحلة معينة من مراحل الدراسة في كلية الصيدلة 70 طالباً وعدد الإناث 90 طالبة هل ان عدد الذكور إلى عدد الإناث متساوية ؟ , اختبر ذلك تحت مستوى المعنوية 0.05 علماً ان قيمة χ^2 الجدولية = 3.84 ؟

الحل:

1- وضع الفرضيات: يتوزع الطلاب حسب الجنس بالتساوي: H_0

لا يتوزع الطلاب حسب الجنس بالتساوي: H_1

حساب التكرار المتوقع؟

70+90 = 160 (عدد افراد العينة)

التكرار المتوقع : $\frac{1}{2} \times 160 = 80$

القيمة المتوقعة للذكور = 80 $(e_1=80)$

القيمة المتوقعة للإناث = 80 $(e_2=80)$

2- ايجاد قيمة χ^2

$$\chi^2 = \sum_{i=1}^{2} \frac{(o_i-e_i)^2}{e_i}$$

$$\chi^2 = \frac{[70-80]^2}{80} + \frac{[90-80]^2}{80}$$

$$\chi^2 = 2.5$$

3- استخراج قيمة χ^2 الجدولية لدرجة حرية n-1 هي ان n تساوي 2 ذكور واناث ودرجة الحرية 1 ومستوى معنوية 0.05 حيث قيمة χ^2 الجدولية تساوي 3.84.

4- القرار : لماكانت قيمة χ^2 المحسوبة اقل من قيمتها في الجدول لذا يمكن الإستنتاج بأن القيم النظرية(المتوافقة) لا تختلف عن المشاهدة اي ان عدد الذكور مشابهة لعدد الإناث وبدون فرق معنوي وما موجود من فرق بينهم يرجع الى عامل الصدفة .

تمرين: في تزاوج بين نباتين أحدهما قصير والآخر طويل ظهرت نتائج الجيل الثاني 30 نبات طويل و20 نبات قصير، بين الى اي نسبة تنتسب هذه النباتات؟ اختبر تحت مستوى 0.05 علماً ان χ^2 الجدولية 3.84 ؟

2.2.8.2 اختبار χ^2 للاستقلال:

يستخدم χ^2 لأختبار الفرضيات الموضوعة على اساس وجود معيارين من التصنيف لمكونات المجموعة لتحديد فيما إذا كان هناك ارتباط بين الصفتين او المعيارين ام انهما مستقلان .

حيث r هي عدد الصفوف تمثل مستويات مختلفة لاحد معايير التصنيف

حيث c هي تمثل الأعمدة وتمثل مستويات مختلفة للمعيار الآخر

درجات الحرية $d.f = (r\text{-}1)\ (c\text{-}1)$

وتستخدم المجاميع الحرية للفئات التي توزع عليها الصفات في تحديد التكرارات المتوقعة فإن هذه المجاميع يجب ان تعتبر ثوابت

فمثلاً : درجة الحرية لجدول التوافق χ^2

$$(2\text{-}1)\ (2\text{-}1)=1$$

كما يعتبر اختبار الإستقلال مقيداً لزوجين من العوامل صفات اختبار χ^2 للاستقلال والتي تميزه عن الإختبارات الأخرى.

1- تسحب العينة من المجتمع موضوع الدراسة والتي تصنف فيه الفئات وفق المتغير التابع والمتغير المستقل على اساس اهمية المتغيرين

2- يعتمد حساب نسب التكرارات المتوقعة لكل فئة على قانون الإحتمال الذي ينص على انه إذا وقع حدثان والحادث هنا معيار التصنيف بصورة مستقلة الواحد عن الآخر فإن احتمال حدوثهما معاً يكون مساوياً لحاصل ضرب احتمال كل منهما على انفراد.

قانون ضرب الإحتمال للأحداث المستقلة $P(A \cap B) = P(A) \times P(B)$

$$\text{الإحتمال} = \frac{\text{عدد الحالات المؤاتية}}{\text{عدد الحالات الممكنة}}$$

3- توضع الفرضية على اساس المتغيرين مستقلين عن بعضهما.

مثال: درست مجموعة من الباحثين العلاقة بين زمرة الدم وشدة الإصابة بحالة مرضية معينة لمجتمع ،جمعت بيانات من (1500) شخصاً وكانت النتائج كما مبين ادناه , هل الحالتين مرتبطتان ؟, اختبر ذلك تحت مستوى احتمال 0.05 علماً ان χ^2 الجدولية = 12.592 .

الحالة المرضية					*Total*
	A	*B*	*AB*	0	
غير مصاب	543 (541.2)	211 (212.96)	90 (92.40)	476 (473.44)	1320
متوسط الاصابة	44 (43.05)	22 (16.94)	8 (7.33)	31 (73.66)	105
شديد الاصابة	28 (30.75)	9 (12.10)	7 (5.25)	31 (26.90)	75
Total	615	242	105	538	1500

الحل:

وضع الفرضيات

الصفتان مستقلتان: H_0

الصفتان مرتبطتان: H_1

القيم المتوقعة (النظرية)

قانون ضرب الإحتمال للأحداث المستقلة × المجموع العام

$$e_{11} = \frac{1320}{1500} \times \frac{615}{1500} \times 1500$$

$$e_{11} = \frac{1320}{1500} \times 615 = 541.20$$

$$e_{12} = \frac{1320}{1500} \times 242 = 212.96$$

وهكذا لبقية القيم

$$\chi^2 = \frac{[543-541.2]^2}{541.2} + \cdots + \frac{[31-26.90]^2}{26.90} = 5.12$$

استخراج قيمة χ^2 الجدولية لدرجة حرية $(r\text{-}1)\ (c\text{-}1)$

$(3-1)\ (4-1)=6$

وهي 12.592.

قيمة χ^2المحسوبة اقل من قيمة χ^2 الجدولية لذا نقبل فرضية العدم ونرفض الفرضية البديلة التي تنص على ان شدة الإصابات وزمرة الدم مستقلة عن بعضها. .

3.2.8.2 استخدام اختبار(χ^2) لاختبار التجانس:

يتصف اختبار الإستقلال اننا نسحب العينة من المجتمع قبل تصنيف الفئات وفقاً لمعياري التصنيف وهذا يعني ان العدد المشاهد للفئات يحدد بقدر سحب العينة ولذلك فإن مجاميع الصفوف والأعمدة تعتبر مقادير محتملة وليس تحت سيطرة الباحث وان العينة المحسوبة تحت هذه الظروف هي عينة مفردة تسحب من مجتمع واحد.

اما في حالة اختبار التجانس فإن الباحث قد يحدد العينات المستقلة تسحب من مجتمعات عديدة وفي هذه الحالة تكون واحدة من المجاميع الحدية ثابتة بينما تكون المجموعة الأخرى وفق معيار التصنيف المستخدم غير ثابتة.

واختبار χ^2 للتجانس تكون فرضية العدم H_0 ان العينات المسحوبة من المجتمعات متجانسة والفرضية البديلة ان المجتمعات غير متجانسة .

مثال: درس باحث مدى استخدام عقار معين بين طلبة كلية الصيدلة الذين اعلنو عن استخدام الأدوية , واختار من هذه المجموعة عينة مكونة من 150 طالباً من الصف الأول و 135 طالباً من الصف الثاني و 125 طالباً من الصف الثالث و 100 طالباً من الصف الرابع واجاب كل طالب الإستفتاء عن مدى استخدام العقار هل هذه البيانات مطابقة او موافقة للفرضية بأن المجتمعات الأربعة متجانسة فيما يخص تناول العقار وكانت النتائج كما في الجدول التالي :

Total	استخدام العقار			المراحل الدراسية
	متقطع كثيراً	متقطع	اختيار احياناً	
150	43 (35.27)	50 (51.47)	57 (63.24)	صف اول
135	20 (31.76)	58 (46.33)	57 (56.91)	صف ثاني
125	24 (29.41)	45 (42.89)	56 (52.70)	صف ثالث
100	33 (23.53)	22 (34.31)	45 (42.16)	صف رابع
510	120	175	215	*Total*

الحل:

1- وضع الفرضيات

H_0: المجتمعات متجانسة

H_1: المجتمعات غير متجانسة

2- استخراج قيمة χ^2 المحسوبة

$$\chi^2 = \sum_{i,j} \frac{(o_{ij}-e_{ij})^2}{e_{ij}}$$

$$\chi^2 = \frac{(57-63.24)^2}{63.24} + \dots + \frac{(33-23.53)^2}{23.53} = 19.4$$

3- استخراج χ^2 الجدولية بدرجة حرية (r-1)(c-1)

(4-1) (3-1) =6

ومستوى معنوية 0.01

لما كانت قيمة χ^2 المحسوبة اكبر من χ^2 الجدولية نرفض فرضية العدم ونقبل الفرضية البديلة اي المجتمعات الأربعة غير متجانسة

امثلة:

1- في دراسة فيما إذا كان هناك ارتباط بين الإصابة بالملاريا وتضخم الطحال وجدت البيانات التالية، فهل هناك علاقة ارتباط بين الإصابة بالملاريا وتضخم الطحال ؟ اختبر ذلك تحت مستوى احتمال 0.01 علماً ان قيمة χ^2 الجدولية تساوي 5.41 .

الإصابة بالملاريا	تضخم الطحال		*Total*
	+	-	
+	740 (546.45)	743 (936.55)	1483
—	1287 (1480.55)	2731 (2537.45)	4018
Total	2027	3474	5501

الصفتان مستقلتان عن بعضهما :H_0

الصفتان مرتبطتان :H_1

$$\chi^2 = \sum_{i,j} \frac{(o_{ij}-e_{ij})^2}{e_{ij}}$$

$$e_{11} = \frac{2027}{5501} \times 1483 = 546.45$$

$$e_{12} = \frac{3474}{5501} \times 1483 = 936.55$$

$$e_{21} = \frac{2027}{5501} \times 4018 = 1480.55$$

$$e_{22} = \frac{3474}{5501} \times 4018 = 2537.45$$

$$\chi^2 = \frac{(740-546.45)^2}{546.45}\frac{(743-936.55)^2}{936.55} + \frac{(1287-1480.55)^2}{1480.55} + \frac{(2731-2537.45)^2}{2537.45}$$

$$\chi^2 = 68.55 + 40 + 25.30 + 14.76 = 148.6$$

بما ان قيمة χ^2 المحسوبة اكبر من χ^2 الجدولية نرفض فرضية العدم ونقبل الفرضية البديلة اي الصفتان مرتبطتان هناك علاقة بين الإصابة بالملاريا وتضخم الطحال .

2-على فرض ان احد الباحثين اراد إيجاد العلاقة بين الجنسين والإصابة بالسرطان فأختار عينة مؤلفة (15) فرداً (8 ذكور , 7 اناث) وقد حصل على البيانات التالية , اختبر ذلك تحت مستوى احتمال 0.05 علماً ان قيمة χ^2 الجدولية تساوي 3.84

الجنس	الإصابة بالسرطان		*Total*
	غير مصاب	مصاب	
ذكور	6 (4.77)	2 (3.20)	8
اناث	3 (4.2)	4 (2.80)	7
Total	9	6	15

H_0: الصفتان مستقلتان

H_1: الصفتان مرتبطتان

$$e_{11} = \frac{8}{15} \times 9 = 4.77$$

$$e_{12} = \frac{8}{15} \times 6 = 3.20$$

$$e_{21} = \frac{7}{15} \times 9 = 4.2$$

$$e_{22} = \frac{7}{15} \times 6 = 2.80$$

$$\chi^2 = \frac{[6-4.77]^2}{4.77}\frac{[2-3.20]^2}{3.20} + \frac{[3-4.2]^2}{4.2} + \frac{[4-2.80]^2}{2.80}$$

$$\chi^2 = 0.32 + 0.45 + 0.34 + 0.51 = 1.62$$

بما ان χ^2 المحسوبة اقل من χ^2 الجدولية نقبل فرضية العدم ونرفض الفرضية البديلة اي الصفتان مستقلتان عن بعضهما .

9.2 الإرتباط والإنحدار:

لقد كان في اهتمامنا في الإختبارات السابقة حول قضايا الإحصاء الإستنتاجي التي تعود الى متغير واحد اما الآن سوف نتحول الى القضايا التي تخص توزيع ذو متغيرين وسنرمز لهذين المتغيرين بالرمز y , x

1.9.2 الإرتباط:

هو الأسلوب الذي يفسر درجة وقوة واتجاه العلاقة بين المتغيرين y , x دون النظر إلى السببية بينهما فقط ، يرتبط هذين المتغيرين بعلاقة خطية او غير خطية وقد لا تكون بينهما اي علاقة على وجه الإطلاق فمثلاً لا نتوقع ان تكون هناك علاقة بين طول الفرد (x) وعمر والده بينما نتوقع ان تكون هناك علاقة بين طول الفرد (x) ووزنه (y) وسوف نتناول الإرتباط البسيط وان كلا المتغيرين (y , x) هما متغيرين مستقلين وان كلاهما يتبع التوزيع الطبيعي وتوضح احد الفرضيات التالية عندما تكون المشاهدات مزدوجة اي قيم y , x :

1- عدم وجود علاقة بين المشاهدات وتحلل بصورة منفردة او منفصلة اي نقصد عدم وجود علاقة بين مشاهدات y , x

2- وجود علاقة بينهما وتحدد هذه العلاقة باستخدام الإرتباط

3- تقدير مقدار هذه العلاقة يستخدم تحليل الإنحدار

ولمعرفة فيما هناك علاقة بين متغيرين نحسب بما يسمى بمعامل الإرتباط وسنرمز له بالرمز r .

ان معامل الإرتباط يوضح العلاقة الخطية بين متغيرين، وكمثال على وجود الإرتباط الخطي البسيط بين متغيرين مستقلين هو عند دراسة العلاقة بين طول الأخ والأخت في عدة عوائل ففي هذه الحالة لا توجد علاقة بين المتغيرين لان التغير في طول الأخ لا يسبب تغير في طول الأخت لان كلا المتغيرين مستقلين، ولكون طول الأخ والأخت يتغيران سوية تبعاً لتغير طول الآباء، هذا يوجب التأكيد ان يكون هناك ترابط بسيط بين المتغيرين. ان قيمة معامل الإرتباط تتراوح بين (-1) و (+1)

$$-1 \leq r \leq 1$$

$$r = \frac{\sum_i x_i y_i - \frac{(\sum_i x_i)(\sum_i y_i)}{n}}{\sqrt{\left(\sum_i x_i^2 - \frac{(\sum_i x_i)^2}{n}\right)\left(\sum_i y_i^2 - \frac{(\sum_i y_i)^2}{n}\right)}}$$

وعندما يكون الإرتباط الخطي ضعيفاً فان معامل الإرتباط r يقترب من الصفر وعندما لا يوجد ارتباط تكون قيمة r= صفر وعندما يكون هناك ارتباط موجب فأن r تقترب من +1 وهذا يعني الزيادة والنقصان في أحد المتغيرين يتبعها زيادة او نقصان في المتغير الآخر يعني ارتباط طردي.

وعندما تقترب r من 1- تعني الزيادة او النقصان في احد المتغيرين يصاحبها نقصان او زيادة في المتغير الآخر على التوالي علاقة الإرتباط عكسي.

يدعى مربع r معامل التحديد r^2 او ما يسمى بالقدرة التنبؤية وهو نسبة مربعات الإنحدار SSR الى مجموع المربعات الكلية

$$r^2 = \frac{SSR}{SST}$$

مثال₁: فيما يلي اوزان واطوال عشرة اشخاص، هل توجد علاقة بين اوزانهم واطوالهم؟ اختبر ذلك تحت مستوى 0.05 علماً r الجدولية تساوي 0.63

$x_i y_i$	x_i^2	y_i^2	الطول سم x_i	الوزن كغم y_i
8448	16384	4356	128	66
9588	19881	4624	141	68
7552	13924	4096	118	64
10710	23409	4900	153	70
9522	19044	4761	138	69
12410	28900	5329	170	73
9180	18225	4624	135	68
8710	16900	4489	130	67
8125	15625	4225	125	65
12024	27889	5184	167	72
$\sum x_i y_i = 96269$	$\sum x_i^2 = 200181$	$\sum y_i^2 = 46588$	$\sum x_i = 1405$	$\sum y_i = 682$

1- وضع الفرضيات $H_0: r = 0$

$$H_1 : r \neq 0$$

2- حساب معامل الإرتباط

$$r = \frac{96269 - \frac{(1405)(682)}{10}}{\sqrt{(200181 - \frac{(1405)^2}{10})(46588 - \frac{(682)^2}{10})}}$$

$$r = \frac{96269 - 95821}{\sqrt{(200181 - 1974025)(46588 - 46512.4)}}$$

$$r = \frac{448}{\sqrt{27785 \times 75.6}}$$

$$r = \frac{448}{\sqrt{2100546}} = \frac{448}{1449.33} = 0.31$$

3-الإستنتاج

لما كانت r المحسوبة اقل من r الجدولية لذا نقبل فرضية العدم ونرفض الفرضية البديلة اي توجد علاقة ارتباط بين وزن الطلاب واطوالهم ولكن هذه العلاقة لم تصل إلى المستوى المعنوي وهي علاقة طردية ضعيفة.

مثال$_2$: الجدول التالي يحوي على بيانات الكميات الناتجة في سلسلة من التفاعلات اجريت في درجة حرارة مختلفة المطلوب ايجاد معامل الإرتباط بين درجات الحرارة وكمية الناتج من التفاعل بغية التعرف على مدى ارتباط الكميات الناتجة بدرجة الحرارة اختبر ذلك تحت مستوى احتمال 0.01 علما r الجدولية تحت مستوى المعنوية 0.01 ودرجة حرية 5 تساوي 0.87؟

درجة الحرارة x_i	الكمية الناتجة من التفاعل y_i	x_i^2	y_i^2	$x_i y_i$
15	41	225	1681	615
20	40	400	1600	800
25	38	625	1444	950
30	35	900	1225	1050
35	32	1225	1024	1120
40	30	1600	900	1200
50	20	3500	400	1000

6735	8274	7475	236	215

1- وضع الفرضيات $H_0: r= 0$

$H_1 : r \neq 0$

2- حساب معامل الارتباط

$$r = \frac{6735-\frac{(215)(236)}{7}}{\sqrt{(7475-\frac{(215)^2}{7})(8274-\frac{(236)^2}{7})}}$$

$$r = \frac{6735-7248.57}{\sqrt{(7475-6603.57)(8274-7956.57)}}$$

$$r = \frac{-513.57}{\sqrt{871.43 \times 317.43}}$$

$$r = \frac{-513.57}{525.94} = -0.98$$

3-الإستنتاج:

بما ان القيمة المطلقة r المحسوبة 0.98 اكبر من r الجدولية هناك ارتباط معنوي قدره 0.98 بين درجة الحرارة والكمية الناتجة من التفاعل وهذه العلاقة عكسية اي كلما زادت درجة الحرارة قلت كمية المواد الناتجة من التفاعل.

2.9.2 الإنحدار:

يعرف الإنحدار مقدار التغير في المتغير y والذي يسمى المتغير التابع نتيجة تغير وحدة واحدة في المتغير x الذي يسمى المتغير المستقل.

ففي حالة اجراء بحث نعين قيم المتغير المستقل مسبقاً فمثلاً تأثير نقطة الإمتحان الفصلي للطالب في مادة الإحصاء على نقطة الإمتحان النهائي , تأثير سرعة نبضات القلب على القلق عند الكبار وتأثير عدد اطفال العائلة على معيار ذكاء الإطفال وتأثير الوزن على مستوى الغلوكوز بالدم وتأثير انقباض ضغط الدم على كمية الدم المفقودة خلال العمليات الجراحية وتأثير مكونات النظام الغذائي على مقياس الشحوم في البلازما وتأثير العمر على ضغط الدم وتأثير عقار معين على الإنخفاض في النبض (ضربة/دقيقة) والإنحدار يختلف عن الإرتباط لأن الإنحدار يدرس فيما إذا المتغير سوف يؤثر في المتغير الآخر بينما الإرتباط يدرس فيما اذا كانت هناك علاقة بين المتغيرين.

إن قيمة معامل الإنحدار يعبر عنها بنفس الوحدات المستخدمة للصفة وتأخذ قيم سالبة او موجبة وقيمة الإنحدار تأخذ قيمة غير محدودة وعندما تكون:

أ- قيمة الإنحدار موجبة يعني كل زيادة في قيم x يتبعها زيادة في قيم y او كل نقصان في قيم x يتبعها نقصان في قيم y

ب- قيمة الإنحدار سالبة فإن كل زيادة في x يتبعها انخفاض في قيمة y

ومن خلال معادلة الخط المستقيم يمكن ان نتنبأ بالمتغير التابع y بدلالة المتغير المستقل x مثلاً التنبأ بضغط الدم المتغير التابع من خلال العمر المتغير المستقل.

$$y^{\sqcap} = a + bx$$

حيث ان $y^{\sqcap}$ هي القيمة المتوقعة للمتغير التابع

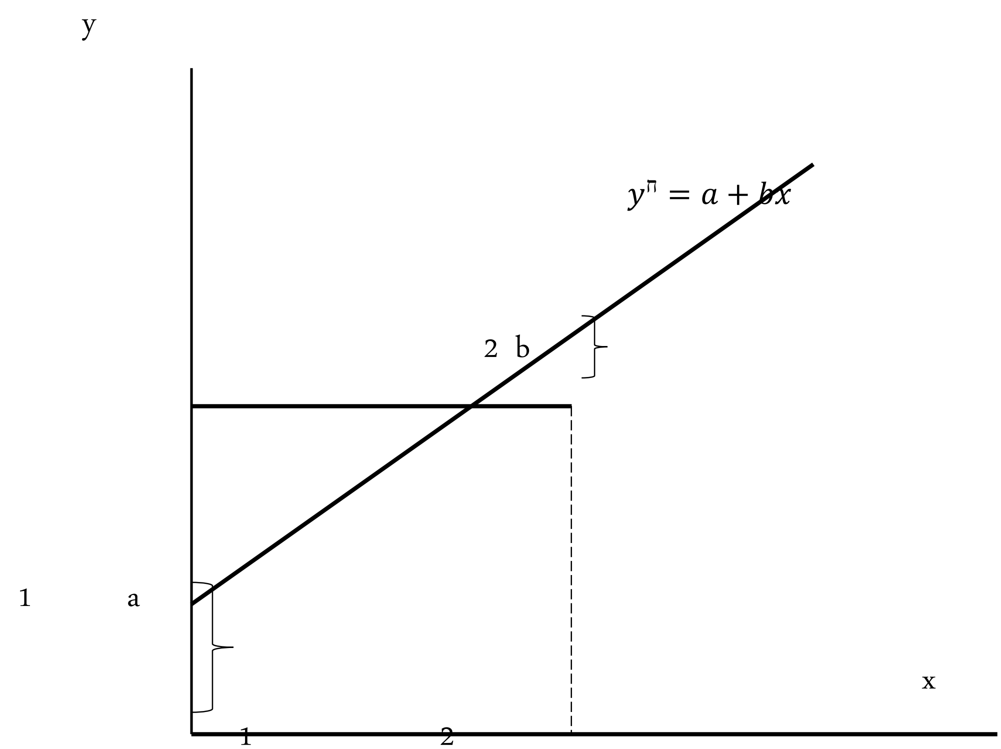
a هي نقطة تقاطع خط الإنحدار مع المحور التصاعدي

b معامل الانحدار

x المتغير المستقل $\qquad \hat{a} = \bar{y} - \hat{b}\bar{x}$

معامل الانحدار $\hat{b} = \dfrac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$

$\hat{a}$ و $\hat{b}$ هي مقدرات جيدة منصفة لكل من a و b على الترتيب.



مثال: من البيانات التالية التي تمثل النقطة الفصلية والنقطة النهائية لأثنى عشر طالباً في مادة الإحصاء من طلبة كلية الصيدلة.

1-اوجد معامل الإرتباط واختبر معنويته على مستوى احتمال 0.01 علماً ان r الجدولية = 71.0

2- اوجد معامل الإنحدار واختبر معنويته على مستوى احتمال 0.01 .

3- تنبأ بنقطة الطالب النهائية في مادة الإحصاء علماً ان النقطة الفصلية x=65 ، x= 50 ، x= 70

x_i النقطة الفصلية	y_i النقطة النهائية	x_i^2	y_i^2	$x_i y_i$	$y^{\sqcap}$ المقدرة y
65	85	4225	7225	5525	88.4
50	74	2500	5476	3700	74.9
55	76	3025	5776	4180	79.4
65	90	4225	8100	5850	88.4
55	85	3025	7225	4675	79.4
70	87	4900	7569	6090	92.8
65	94	4225	8836	6110	88.4
70	98	4900	9604	6860	92.8
55	81	3025	6561	4455	79.4
70	91	4900	8281	6370	92.8
50	76	2500	5776	3800	74.90
55	74	3025	5476	4070	79.4
725	1011	44475	85905	61685	

1- معامل الإرتباط

$$r = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sqrt{(\sum x_i^2 - \frac{(\sum x_i)^2}{n})(\sum y_i^2 - \frac{(\sum y_i)^2}{n})}}$$

$$r = \frac{61685 - \frac{(725)(1011)}{12}}{\sqrt{(44475 - \frac{(725)^2}{12})(85905 - \frac{(1011)^2}{12})}}$$

$$r = \frac{61685 - 61081.25}{\sqrt{(44475 - 43802.08)(85905 - 85176.75)}}$$

$$r = \frac{603.75}{\sqrt{672.92 \times 723.25}}$$

$$r = \frac{603.75}{700.4} = 0.86$$

الإستنتاج

بما ان r المحسوبة أكبر من r الجدولية هناك ارتباط معنوي وطردي.

-2

$$b = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

$$b = \frac{61685 - \frac{(725)(1011)}{12}}{44475 - \frac{(725)^2}{12}}$$

$$b = \frac{61685 - 61081.25}{44475 - 43802.08}$$

$$b = \frac{603.75}{672.92} = 0.897$$

وبالتعويض في معادلة خط الإنحدار نحصل على قيمة $y^{\sqcap}$

$$y^{\sqcap} = a + bx$$

$\bar{x}$ =60.417 $\bar{y} = 84.220$

$a=\bar{y}-b\bar{x}$

a= 84.220 – (0.897)(60.417)

a= 30.056

اوجد قيمة y عندما تكون $65=x$

$50=x$

$70=x$

$$\hat{y}(65)=30.05+0.897\times 65=88.40$$

$$\hat{y}(50)=30.05+0.897\times 50=74.9$$

$$\hat{y}(70)=30.05+0.897\times 70=92.8$$

وعند رسم خط الإنحدار بتحديد نقطتين ولتكن احدهما $(\bar{x},\bar{y})$ ونحدد نقطة اخرى على ان تكون قيمة x محددة ونستخرج قيمة y حسب المعادلة

$$.\hat{y}=a+bx$$

تقدير معامل الانحدار بمجال ثقة

10.2 مجال الثقة :

إلى جانب التقدير النقطي لوسائط المجتمع (المتوسط ، الإنحراف المعياري،...) هناك تقدير آخر وهو التقدير بمجال الثقة.

تعريف1: نسمي المجال $[a\,,b]$ بمجال الثقة بمعدل α إذا أمكن كتابة

$$P(a\le m\le b)\ge 1-\alpha$$

حيث α عدد حقيقي مسبقا و m وسيط للمجتمع، و α تأخذ القيم مثلا 1% و 2% ، .....

فمثلا من أجل المعدل 5% نكتب :

$$P(a\le m\le b)\ge 0.95$$

(هذا التعريف غير واضح)

تعريف2: وينتج من كون التقدير بواسطة مجال الثقة يتمثل في البحث عن مجال $[a\,,b]$ بحيث يكون

$$P(a < \theta < b) = \alpha$$

حيث α مختار مسبقا و θ وسيط، ونعين مجال الثقة في الحالتين الأولى:

- **إذا كانت العينات كبيرة الحجم $n > 30$**

نعلم أن $\overline{x} = (x_1 + x_2 + \cdots + x_n)/n$ ينتهي إلى توزيع طبيعي حيث (x_i, m_i, σ_i) موزعة توزيع طبيعي بوسيطين، غيرانه نفضل عادة اعتبار هذه المتغيرات x_i كلها بوسيطين $\sigma\,, m$ لأنها مستخلصة من مجتمع واحد موزع توزيع طبيعي، وعليه فإن $\overline{x}$ يكون في النهاية موزع توزيع طبيعي بوسيطين m و $\sigma/\sqrt{n}$ في حالة سحب غير شامل و m و $\frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$ في حالة سحب شامل . وبالتالي فإن المتغير المتمركز المختزل $u = \frac{\overline{x}-m}{\sigma/\sqrt{n}}$ أو $u = \frac{\overline{x}-m}{\frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}}$ وهو متغير عشوائي موزع توزيع طبيعي متمركز مختزل $\mathcal{N}(0,1)$ ، ويحدث هذا مهما كان التوزيع الأصلي وذلك استنادا إلى نظرية النهاية المركزية (Théorème centrale limite) . لأننا بصدد عينات كبيرة الحجم، وإذا كان u_α بحيث $\phi(u_\alpha) = \frac{\alpha+1}{2}$ فإننا نأخذ أو نعين من جدول قيم التوزيع الطبيعي $\phi(x)$ بحيث نكتب :

$$P(-u_\alpha < u < u_\alpha) = \alpha$$

وبوضع $r = \sigma/\sqrt{n}$ نحصل على $r.u = \overline{x} - m$ وعليه:

$$P(|u| < u_\alpha) = P(|\overline{x} - m| < r.u_\alpha) = \alpha$$
$$\Rightarrow \; P(-r.u_\alpha + \overline{x} < m < r.u_\alpha + \overline{x}) = \alpha$$

وهذا من أجل سحب غير شامل ، ومن أجل سحب شامل يكفي ان نضرب في عامل الشمولية $\left(\sqrt{\frac{N-n}{N-1}}\right)$.

ونلاحظ في الحالة العامة σ غير معروف فعندئذ في العبارة السابقة

$$P\left(-\frac{\sigma}{\sqrt{n}}.u_\alpha + \overline{x} < m < \frac{\sigma}{\sqrt{n}}.u_\alpha + \overline{x}\right) = \alpha$$

نختار S بدل σ وهو مقدر جيد وبدون حيد لها.

2- من أجل عينات صغيرة الحجم:

(في حالة n صغيرة لا نستخدم معامل الشمولية)

$(\sigma^2 = S^2)$ في حالة n كبيرة ، أما في حالة n صغيرة لا يكون صحيح ،

نطبق (student) $\chi^2_{(n-1)}$ في حالة $n \leq 30$.

$\chi^2_{(n-1)}$ متغير عشوائي يأخذ نقاطه:

- من أجل عينات صغيرة المقاس

في هذه الحالة يمكن اهمال عامل الشمولية حيث $\frac{N-n}{N-1}$ يتقارب نحو 1 ويكون هذا إذا كان N كبيرا بقدر كاف بالنسبة إلى n صغيرة أو العكس .

وعندئذ يمكن اعتبار $x_1, x_2, \dots, x_n$ متغيرات مستقلة وحيث ان n صغيرة فإنه لا يمكن تطبيق نظرية النهاية المركزة (T.C.L) ، وبالتالي لا يمكن استنتاج أي شيء بشأن طبيعة توزيع مميزات العينة $(\sigma^2, \overline{x})$ ، لذا نعتبر P مجتمع طبيعي وعندئذ $\overline{x}$ موزعا وفق قانون طبيعي أو غوص بوسطين m و $\frac{\sigma}{\sqrt{n}}$ ، وهنا نميز حالتين :

1- إذا علمنا σ : فإن مجال الثقة ينشأ بالطريقة السابقة أي ان:

$$P\left(-\frac{\sigma}{\sqrt{n}}.u_\alpha + \overline{x} < m < \frac{\sigma}{\sqrt{n}}.u_\alpha + \overline{x}\right) = \alpha$$

2- إذا كان σ غير معلوم : وهي الحالة الشائعة التي يكون فيها تقارب S^2 غير اكيد وبالتالي المتغير $u = \frac{\overline{x}-m}{\sigma/\sqrt{n}}$ أو $u = \frac{\overline{x}-m}{S/\sqrt{n}}$ (استبدلنا σ بـ S^2) لا يمكن اعتباره غوصي او طبيعي .

إلا أن المتغير $v = \sum_{i=1}^{n}\left(\frac{x_i-\overline{x}}{\sigma}\right)^2$ الذي يساوي $v = \frac{(n-1)S^2}{\sigma^2}$ أو $v = \frac{n\sigma_e^2}{\sigma^2}$ بـ $(n-1)$ درجة من الحرية.

يتبع قانون $\chi^2_{(n-1)}$ بـ $(n-1)$ درجة من الحرية.

وبالتالي فإن الكمية أو المتغير $t = \frac{\overline{x}-m}{S/\sqrt{n}}$ يتبع قانون Student-ficher

أيضا بـ $(n-1)$ درجة من الحرية (لوجود S).

وعليه فإذا اعتبرنا المجال المفتوح $]-u_\alpha , u_\alpha[$ فإننا نكتب :

$$P(-u_\alpha < t < u_\alpha) = \alpha$$
$$\Rightarrow P\left(|\overline{x} - m| < \frac{S}{\sqrt{n}}u_\alpha\right)$$

مثال: جد مجال الثقة لمعامل الإنحدار (في المثال السابق)

نعلم ان المعادلة $y = a + bx$ ناتجة من $E(Y) = f(X) = a + bX$ أو $Y = a + bX + \varepsilon$ حيث X و Y متغيرات عشوائية و ε الخطأ العشوائي مركز وذو تباين σ^2 ثابت ، وهذا الخطأ مستقل عن المتغير X .

هذه المعادلة بالضبط $y = E(Y/X = x) = a + bx$

و $\varepsilon \sim \mathcal{N}(0 , \sigma^2)$ فرضا.

لتقدير كل من a و b و σ^2 نستعمل طريقة المربعات الصغرى وذلك بأخذ عينة من المشاهدات $\{(x_i\,,\,y_i)\,,\; i=1,2,\ldots,n\;\}$ وحساب

$$\min_{a,b}\sum_{i=1}^{n}(y_i-a-bx_i)^2 \qquad (*)$$

نضع $\overline{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$ ، $\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$

$$S_y^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\overline{y})^2$$

$$S_x^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

$$r=\frac{S_{xy}}{S_xS_y}S_{xy}=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})\,(y_i-\overline{y}) \quad ,$$

المقدرات $\hat{b}=\frac{S_{xy}}{S_x^2}$ و $\hat{a}=\overline{y}-\hat{b}\overline{x}$ تجعل العبارة (*) اصغر مايمكن وهي مقدرات جيدة ومنصفة أي ان $E\hat{b}=b$ و $E\hat{a}=a$.

أما مقدر σ^2 هو S^2 حيث :

$$S^2=\frac{1}{n-2}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \quad ; \qquad \hat{y}_i=\hat{a}+\hat{b}x_i$$

بالنسبة لقوانين الوسائط:

نعلم $\hat{a}$ و $\hat{b}$ هي متغيرات عشوائية حقيقية بمصفوفة التغاير (Matrice de covariance) :

$$\sigma^2\begin{bmatrix}\frac{1}{n}+\frac{\overline{x}^2}{(n-1)S_x^2} & -\frac{\overline{x}}{(n-1)S_x^2}\\ -\frac{\overline{x}}{(n-1)S_x^2} & \frac{1}{(n-1)S_x^2}\end{bmatrix}$$

والذي تم تقديرها باستبدال σ^2 بـ S^2 وتحت الفرض $(y_i-\hat{y}_i)$ يتبع قانون غوص (قانون طبيعي) ينتج لنا :

$$\frac{(n-2)S^2}{\sigma^2}\sim\chi^2_{(n-2)}$$

وعندئذ الإحصائيات:

$$\left.(\hat{a}-a)\middle/ s\left(\frac{1}{n}+\frac{\overline{x}^2}{(n-1)S_x^2}\right)^{\frac{1}{2}}\right.$$

و

$$\left.(\hat{b}-b)\middle/ s\left(\frac{1}{(n-1)S_x^2}\right)^{\frac{1}{2}}\right.$$

تتبع قانون ستودنت $(student)$ بـ $(n-2)$ درجة من الحرية ، وهذا ما يسمح بإنشاء مجالات ثقة لكل من a و b وهي:

$$\left]\hat{a}-t_{\alpha/2,(n-2)}s\left(\frac{1}{n}+\frac{\overline{x}^2}{(n-1)S_x^2}\right)^{\frac{1}{2}} \quad , \quad \hat{a}+t_{\alpha/2,(n-2)}s\left(\frac{1}{n}+\frac{\overline{x}^2}{(n-1)S_x^2}\right)^{\frac{1}{2}}\right[$$

و

$$\left]\hat{b}-t_{\alpha/2,(n-2)}s\left(\frac{1}{(n-1)S_x^2}\right)^{\frac{1}{2}} \quad , \quad \hat{b}+t_{\alpha/2,(n-2)}s\left(\frac{1}{(n-1)S_x^2}\right)^{\frac{1}{2}}\right[$$

ملاحظة: أي استنتاج يرافق كل من a وb لا يمكن الحصول عليه من خلال اعتبار مجالات الثقة منفصلة، منطقة الثقة في الأصل هي القطع الناقص للمعادلة:

$$n(\hat{a}-a)^2+2(\hat{a}-a)(\hat{b}-b)\sum_{i=1}^{n}x_i+(\hat{b}-b)^2\sum_{i=1}^{n}x_i^2=2S^2F_{\alpha,2,(n-2)}$$

وهو محتوى في المستطيل المحدد بمجالات الثقة وهناك جزء كبير من قيم الثنائية $(a\,,b)$ يستبعد من منطقة الثقة وخاصة a وb مترابطان .

ولاختبار معنوية معامل الإنحدار يستخدم تحليل البيانات

$$H_0: b = 0$$

$$H_1: b \neq 0$$

ثم نستعين في الحل بجدول Nova.

مصدر التباين	$d.f$ درجات الحرية	مجموع المربعات SS	MS التباين	F_{cal}	F_{tabl}
Regression	1 ونقصد 1 هنا عدد المتغيرات وهي x فقط	SSR= $b^2 \times (\sum x_i^2 - \frac{(\sum x_i)^2}{n})$	$\frac{SSR}{d.f}$	$\frac{MSR}{Mse}$	من جدول F بدرجة حرية R بالاتجاه الأفقي ودرجة حرية الخطأ بالاتجاه العمودي
Residual الخطأ التجريبي error	n-1-1 ونقصد n هنا عدد المشاهدات لكل متغير	SSe = SST - SSR	$\frac{SSe}{d.f}$		
Total	n-1 n هنا هي عدد المشاهدات	SST = $\sum y_i^2 - \frac{(\sum y_i)^2}{n}$			

$$SSR = b^2 \times SSx = (0.897)^2 \times (672.92) = 541.44$$

$$SSx = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$SSx = 44475 - \frac{(725)^2}{12} = 672.92$$

$$SST = SSy = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$SST = 85905 - \frac{(1011)^2}{12} = 728.25$$

$$SSe = SST - SSR$$

$$SSe = 728.25 - 541.44 = 186.81$$

<table>
<tr><th>S.o .v</th><th>d.f</th><th>SS</th><th>MS</th><th>F_{cal}</th><th>F_{tabl}</th></tr>
<tr><td>Regression
الإنحدار</td><td>1</td><td>541.44</td><td>541.44</td><td rowspan="2">28.98</td><td rowspan="3">10.04
F الجدولية عند مستوى 0.01 ودرجة حرية 10 للخطأ و 1 للانحدار</td></tr>
<tr><td>Residual
الخطأ التجريبي error</td><td>10</td><td>186.81</td><td>18.681</td></tr>
<tr><td>Total</td><td>11</td><td>728.25</td><td></td><td></td></tr>
</table>

بما ان F المحسوبة اكبر من F الجدولية لذا نرفض فرضية العدم أي معامل الإنحدار معنوي يختلف عن الصفر وهنا تجدر الإشارة ان نسبة مجموع المربعات للانحدار إلى مجموع المربعات الكلية التي يمكن تفسيرها بالعلاقة الخطية الموجودة بين المتغيرين y,x وعادة يرمز لها بالرمز r^2 يسمى معامل التحديد او القدرة التنبؤية اما نسبة مجموع المربعات التي لا تفسر العلاقة الخطية بين المتغيرين y,x والتي تعزى إلى الخطأ العشوائي ويطلق عليها بمعامل عدم الإرتباط ويرمز لها بالرمز K

$$K = 1 - r^2$$

وفي المثال السابق $$r^2 = \frac{SSR}{SST} = \frac{541.44}{728.25}$$

$$r^2 = 0.74$$

اي ان 75% من الإختلافات الكلية في قيم y تعود الى وجود العلاقة الخطية بين المتغيرين y , x وان $1 - 0.74 =$ 26% من الإختلافات الكلية تعود الى الخطأ العشوائي:

$$r^2 = (r)^2$$

$$r^2 = (0.86)^2 = 0.74$$

11.2 بعض المفاهيم الأساسية في تصميم التجارب:

1- التجربة: هي الخطط التي ترسم مقدماً لتشكل اساس جيد ومأمون لكي يتم الحصول على معلومات جديدة لرفض او تأكيد فرضيات سابقة او استنتاج قواعد او قوانين جديدة او هي الأسلوب العلمي لاختيار الفرضيات واستكشاف العلاقات بين المتغيرات ويمكن تلخيص التجربة بالنقاط التالية :

أ- تحديد المشكلة

ب- تحديد المتغير المتأثر او ما يسمى بمتغير الإستجابة او المتغير التابع وهي القيمة او المستوى الذي يتم التوصل للمتغير او المتغيرات المدروسة لكل وحدة تجريبية على انفراد وعليه كمية المحصول للوحدة التجريبية هو استجابة للمعاملة المعنية على سبيل المثال عند دراسة تأثير مستويات عقار معين على معايير الدم في الإنسان فمعايير الدم هي بمتغير الإستجابة.

جـ- تحديد العامل او العوامل التي سيجري تغيرها ويطلق عليها العوامل المؤثرة حيث يدرس تأثير عامل بعدة صيغ او مستويات مثال العامل المدروس او المؤثر الأصناف، مواعيد الزراعة، مستوى العقار حيث يدرس تأثير عامل واحد فقط مع ثبوت العوامل الأخرى تسمى التجارب البسيطة وحيث يدرس تأثير أكثر من عامل وتسمى بالتجارب العاملية.

2- التصميم:

تصميم التجربة ببساطة يعني تخطيطها، بحيث يصبح بالإمكان جمع المعلومات المتعلقة بالمشكلة المراد دراستها لكي تضمن امكانية الحصول على البيانات المناسبة التي تسمح بتحليلها تحليلاً سليماً وموضوعياً للوصول إلى استنتاجات صحيحة فيما يتعلق بالمشكلة.

وعند اجراء تصميم التجربة يجب تحديد ما يأتي:

أ- عدد المشاهدات المطلوب تسجيلها وحجم العينة يجب ان يكون ملائم اي يجب ان يكون حجم العينة أكبر كي يؤدي إلى قلة الخطأ.

ب- الأسلوب التجريبي الذي ستجري عليه التجربة

يجب ان يكون الأسلوب العشوائي (بدون تدخل شخص) من حيث ان العشوائية تميل لموازنة العوامل وبالأسلوب العشوائي الذي يقلل الخطأ العشوائي.

3- النموذج الرياضي:

يمكن للباحث بعد ان يتخذ الأسلوب التجريبي العشوائي من وضع نموذج يصف التجربة بحيث يظهر هذا النموذج المتغير المتأثر كدالة لكل العوامل التي سنذكرها وهي فروض او قيود فرضية على التجربة كنتيجة لتطبيق الأسلوب العشوائي.

4- التحليل الإحصائي:

يعتبر التحليل المرحلة الأخيرة وفي هذه المرحلة يتم جمع البيانات ثم جدولتها ثم اجراء الإختبارات الإحصائية والرياضية ثم الوصول إلى قرارات مفيدة لاختبار فرضيات متعلقة بالنموذج الرياضي يصف التجربة.

ويمكن تلخيص التحليل بثلاث نقاط:

1- جمع البيانات وجدولتها.

2- اجراء الإختبارات الإحصائية.

3- مناقشة النتائج وتفسيرها.

الوحدة التجريبية

هي أصغر وحدة اساسية او أصغر جزء في مواد التجربة تطبق عليها المعاملات او المعالجات وتستخدم في قياس المتغيرات تحت الدراسة وقد تكون الوحدة التجريبية حيواناً او نباتاً.

المعالجات او المعاملات

تمثل مجموعة الظروف التجريبية المتغيرة التي تخضع تحت سيطرة الباحث والذي يقوم بتوزيعها على الوحدات التجريبية حسب التصميم وقد تكون المعالجات عدة مستويات لعامل واحد تسمى بالتجارب البسيطة او تكون عدة مستويات لاكثر من عامل كما هو الحال في التجارب العاملية مثلاً في النبات نوع السماد (سماد عضوي، سماد اليوريا، سماد فوسفاتي) وكذلك كمية السماد، عمق البذور، ميعاد الزراعة وفي الحيوان تكون المعاملة نوع الحيوان، جنس الحيوان، سلالة الحيوان، نوع الغذاء.

معاملة المقارنة

عادة ما تشمل التجربة على معاملة او اكثر بهدف اتخاذها اساساً في جدوى اضافة عامل (مبيد , تسميد ......الخ) فقد تشمل تجربة مخصصة لدراسة تأثير مبيد معين بتراكيز (2% , 4% , 6% , 8%) اضافة الى معاملة اضافية تمثل عدم استخدام المبيد كي تكون اساساً لمعرفة جدوى اضافة المبيد من حيث المبدأ , وكذلك عندما لا تعطي بعض الوحدات التجريبية كمية من السماد (اي صفر سماد) بينما تعطي مجموعات اخرى من الوحدات التجريبية مستويات محدودة من السماد قيد الدراسة .

التكرارات

تظهر المعاملة اكثر من مرة في التجربة ويطلق على ذلك التكرارات ويجب تكرار المعاملة الواحدة اكثر من مرة في التجربة كي يمكن تقدير قيمة الخطأ التجريبي وبالتالي فصله عن تأثير المعاملة حيث أن تمثيل كل معاملة بوحدة تجريبية واحدة لا يعطي فكرة صحيحة عن تأثير المعاملة.

كما ان زيادة تكرار المعاملات يؤدي الى زيادة دقة كفاءة التجربة وذلك كنتيجة مباشرة لتقليل قيمة الخطأ لمتوسط المعاملة وان تكرار المعاملة يفيد في حالة فقد او تلف احدى القطع التجريبية.

تحليل التباين:

ان مصطلح تحليل التباين يطلق على مدى واسع من الأساليب الإحصائية الفنية ويكاد اغلب الإحصائيين في المواضيع السابقة مثل اختبار t التعرف على معنوية الإختلافات بين الوسطين الحسابيين لمجموعتين تجريبيتين ، وبالتالي التباين هو اسلوب احصائي يتم بواسطته مقارنة الإختلافات بين اكثر من متوسطين حسابيين تعود لأكثر من مجموعتين، وبالتالي فإن تحليل التباين يتم بتجزئة التباين بين مجموعة من المشاهدات يتم بواسطته تجزئة الإختلافات الكلية إلى مكوناتها المختلفة وحسب مصادرها المعروفة وغير المعروفة وبالتالي الكشف عن وجود او عدم وجود فروق معنوية احصائية بين عدد من المتوسطات الحسابية ذلك من خلال اختبار المعنوية والتي اطلق Snedecor اسم اختبار t نسبة إلى العالم Fisher , وفي تحليل التباين يتم اختبار عدة متوسطات لعوامل او مجموعة عينات دفعة واحدة مما يسهل على الباحثين في ميادين البحوث التجريبية كالتجارب الزراعية والبيولوجية والصناعية مما يسهل العمل كثيراً .

فعلى سبيل المثال لو أردنا اختبار 4 متوسطات عينات دفعة واحدة باستخدام اختبار t فإن ذلك يتطلب إختباركل زوج من متوسطات العينات على حدة بصورة عامة

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$ توافيق r من n (طرق الاختيار غير المرتب)

$$\binom{4}{2} = \frac{4!}{2!\,(4-2)!} = \frac{4 \times 3 \times 2 \times 1}{2 \times 1 \times 2 \times 1} = 6$$

اي 6 اختبارات t للتعرف على الإختلافات بين العينات، ولكن باستخدام تحليل التباين يعني استخدام اختبار واحد لجميع البيانات او المتوسطات في نفس الوقت للحصول على استنتاج عام بوجود او عدم وجود اختلاف معنوي بين العينات وهذا يعتبر ادق وأفضل من الأسلوب السابق، اي ان التحليل يعني تجزئة إلى مركباته، وتحليل التباين عبارة عن عملية رياضية يقسم فيها التباين الكلي إلى مكوناته المحتملة.

خطوات تحليل التباين

y_{ij} = قيمة اي مشاهدة y في المعاملة i في التكرار j

حيث

$i= 1 ,2 , 3 , ... t$ (من المعاملات)

$j= 1 ,2 , 3....... r$ (من التكرارات)

خطوات تحليل التباين:

1- نستخرج قيمة معامل التصحيح $C.F$

وذلك من خلال

$$C.F = \frac{\text{مربع المجموع الكلي}}{\text{عدد المشاهدات الكلية}} = \frac{y_{..}}{tr}$$

2- استخراج SST

$$SST = \sum_{i,j} y_{ij}^2 - \frac{(y_{..})^2}{tr}$$

3- استخراج SSt traitement

$$SSt = \frac{\sum_i y_i^2}{r} - \frac{(y_{..})^2}{tr}$$

4- استخراج مجموع المربعات للخطأ التجريبي SSe

$$SSe = SST - SSt$$

5- عمل جدول تحليل التباين

6- استخراج قيمة F المحسوبة

7- استخراج قيمة F الجدولية من جدول F اعتماداً على درجات الحرية للمعاملات بالاتجاه الأفقي ودرجات حرية الخطأ التجريبي بالاتجاه العمودي ومستوى المعنوية 0.01 او أكثر 0.05

8- إذا كانت F المحسوبة أكبر من F الجدولية أو تساويها نستخدم أحد طرق اختبار المعنوية للمتوسطات للتعرف على المعامل ذو التأثير الأكبر لان هناك تأثير معنوي للمعامل في الصفة المدروسة.

الصيغة العامة لجدول تحليل التباين

Analyse de la variance

Une table Nova

مصادرالإختلاف Source de variation	درجات الحرية Degré de liberté d.f	مجموع المربعات Somme des carrés SS	متوسطات المربعات التباين Moyenne des carrés MS	F_{cal}	F_{tabl}
يشمل جميع مسببات الإختلافات بين مواد التجربة للتصميم المستعمل في التجربة	وهي عدد المقارنات المستقلة يمكن اجراءها في كل مصدر من مصادر الإختلاف	مجموع مربعات الإنحرافات المسؤول عنها كل مصدر في مصادر الإختلاف	هو التباين لكل مصدر من مصادر الإختلاف ويحسب بقسمة مجموع المربعات على درجات الحرية لكل مصدر	تحسب بقسمة تباين كل مصدر على تباين الخطأ التجريبي اي انها نسبة التباين	تستخرج من جدول F لكل مصدر اختلاف اعتماداً على مستوى المعنوية ودرجة حرية البسط بالاتجاه الأفقي ودرجات حرية الخطأ بالاتجاه العمودي

اختيار التصميم التجريبي المناسب يعتمد على عدة امور هي:

أ- طبيعة تجانس الوحدات التجريبية في المجال المخصص لتنفيذ التجربة.

ب- عدد المعاملات المشمولة بالتجربة.

جـ- عدد العوامل المراد دراستها والأهمية النسبية لكل تصميم.

التصميم العشوائي الكامل مع تسجيل أكثر من مشاهدة

في بعض تجارب التصميم العشوائي الكامل قد نسجل عدة مشاهدات في كل وحدة تجريبية طبق عليها معامل ، وعليه فإن المعامل يحتوي على عدد من الوحدات التجريبية ، وقد يأخذ من كل وحدة تجريبية عدد من العينات العشوائية ، ونتيجة لذلك نلاحظ إضافة الى تباين المعاملات وتباين الخطأ التجريبي سوف يظهر تباين العينات حيث ان عدم حساب تباين العينات سوف يؤدي الى خطا في التحليل الإحصائي حيث ان تباين المعاملات سوف يزداد وتباين الخطأ كذلك .وفي مثل هذه التجارب الحقلية فقد لا يتوفر الوقت الكافي للباحث مثلا كي يحصد كلا من القطع التجريبية بأكملها ولهذا فقد يلجأ الباحث الى اختيار قطعة عشوائية من كل وحدة تجريبية ثم يحصر بيانات كل من الأجزاء المختارة . وكذلك في تجارب تغذية الحيوانات قد تكون الوحدات التجريبية عدد من حظائر الحيوانات ويوجد في كل حظيرة عدد من الحيوانات التي نطلق عليها العينات يوزع على كل منها أحد العلائق (المعاملات) حيث ان الحيوانات تمثل مشاهدات داخل الوحدة التجريبية التي تمثل الحظيرة.

إذا افترضنا t من المعاملات وكررت على عدد r من الوحدات التجريبية ثم سجلت المشاهدات على عدد s من العينات داخل كل وحدة تجريبية

$$\begin{aligned} i &= 1,2,\ldots\ldots\ldots\ldots t \\ j &= 1,2,\ldots\ldots\ldots\ldots r \\ k &= 1,2,\ldots\ldots\ldots\ldots s \end{aligned}$$

معادلة النموذج الرياضي

$$\boxed{y_{ijk} = M + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}}$$

حيث y_{ijk} قيمة المشاهدة y في المعامل i في الوحدة التجريبية j في العينة k

M المتوسط العام للتجربة

α_i تأثير المعاملi

β_j تأثير المعامل j (الوحدة التجريبية)

γ_{ij} تأثير الخطأ العشوائي للوحدة التجريبية j في المعامل i

ε_{ijk} تأثير الخطأ العشوائي للعينة k في الوحدة التجريبية j في المعامل i وهو يتوزع وفق التوزيع الطبيعي $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$

لتقدير وسائط النموذج نستعمل الطريقة التالية:

أولا نفرض أن : $\sum_{i=1}^{t} \alpha_i = 0$ ، $\sum_{j=1}^{r} \beta_j = 0$ ، $\sum_{i=1}^{t} \gamma_{ij} = 0$ ، $\sum_{j=1}^{r} \gamma_{ij} = 0$

لأن التأثير ليس بالضرورة إيجابي ممكن أن يكون سلبي.

ثانيا نضع

$$\Delta = \sum_{i,j,k} \left(y_{ijk} - M - \alpha_i - \beta_j - \gamma_{ij} \right)^2$$

ونبحث عن قيم الوسائط التي تجعل الكمية Δ أصغر ما يمكن أي أن

$$\frac{\partial \Delta}{\partial \alpha_i} = 0 \Rightarrow -2 \sum_{i,j,k} \left(y_{ijk} - M - \alpha_i - \beta_j - \gamma_{ij} \right) = 0$$

$$\Rightarrow \sum_{i,j,k} y_{ijk} - trsM - rs \sum_{i=1}^{t} \alpha_i - ts \sum_{j=1}^{r} \beta_j - s \sum_{i,j} \gamma_{ij} = 0$$

ومنه تقدير المتوسط العام للتجربة M والذي نرمز له بالرمز $\widehat{M}$ هو:

$$\boxed{\widehat{M} = \frac{1}{trs} \sum_{i,j,k} y_{ijk} = y_{...}}$$

بالنسبة لتقدير α_i لدينا

$$\sum_{i} \left(\sum_{j,k} \left(y_{ijk} - M - \alpha_i - \beta_j - \gamma_{ij} \right) \right) = 0$$

$$\sum_{j,k} y_{ijk} - rsM - rs\alpha_i = 0 \Rightarrow \alpha_i = \frac{1}{rs} \sum_{j,k} y_{ijk} - M$$

$$\Rightarrow \boxed{\hat{\alpha}_i = y_{i..} - y_{...}}$$

بالنسبة لتقدير β_j لدينا

$$\sum_{j}\left(\sum_{i,k}\left(y_{ijk}-M-\alpha_i-\beta_j-\gamma_{ij}\right)\right)=0$$

$$\sum_{i,k} y_{ijk}-tsM-ts\beta_j=0 \ \Rightarrow \ \beta_j=\frac{1}{ts}\sum_{i,k} y_{ijk}-M$$

$$\Rightarrow \ \boxed{\hat{\beta}_j=y_{.j.}-y_{...}}$$

وبالنسبة لتقدير γ_{ij} لدينا

$$\sum_{i,j}\left(\sum_{k}\left(y_{ijk}-M-\alpha_i-\beta_j-\gamma_{ij}\right)\right)=0$$

$$\sum_{k}\left(y_{ijk}-M-\alpha_i-\beta_j-\gamma_{ij}\right)=0 \ \Rightarrow \ \gamma_{ij}=\frac{1}{s}\sum_{k} y_{ijk}-M-\alpha_i-\beta_j$$

$$\Rightarrow \ \hat{\gamma}_{ij}=y_{ij.}-y_{...}-(y_{i..}-y_{...})-\left(y_{.j.}-y_{...}\right)$$

$$\Rightarrow \ \boxed{\hat{\gamma}_{ij}=y_{ij.}-y_{i..}-y_{.j.}+y_{...}}$$

ملاحظة: $\widehat{M}$ و $\hat{\alpha}_i$ و $\hat{\beta}_j$ و $\hat{\gamma}_{ij}$ تقديرات منصفة أي أن:

$$E\left(\widehat{M}\right)=M \ , \quad E(\hat{\alpha}_i)=\alpha_i \ , E\left(\hat{\beta}_j\right)=\beta_j \ , \ \ E\left(\hat{\gamma}_{ij}\right)=\gamma_{ij}$$

بالفعل

$$E\left(\widehat{M}\right)=E(y_{...})=E\left(\frac{1}{trs}\sum_{i,j,k} y_{ijk}\right)=\frac{1}{trs}E\left(\sum_{i,j,k} y_{ijk}\right)$$

$$=\frac{1}{trs}E\left(\sum_{i,j,k}\left(M+\alpha_i+\beta_j+\gamma_{ij}+\varepsilon_{ijk}\right)\right)$$

$$=\frac{1}{trs}E\left(trsM+rs\sum_{i}\alpha_i+ts\sum_{j}\beta_j+s\sum_{i,j}\gamma_{ij}+\sum_{i,j,k}\varepsilon_{ijk}\right)$$

$$=\frac{1}{trs}E\left(trsM+rs.0+ts.0+s.0+\sum_{i,j,k}\varepsilon_{ijk}\right)$$

$$= \frac{1}{trs}\left(trsM + \sum_{i,j,k} \underbrace{E\varepsilon_{ijk}}_{=0}\right) = \frac{1}{trs}\left(trsM + \sum_{i,j,k} 0\right) = M$$

وأيضا

$$E\hat{\alpha}_i = E(y_{i..} - y_{...}) = \frac{1}{rs}\sum_{j,k} Ey_{ijk} - Ey_{...}$$

$$= \frac{1}{rs}E\left(rsM + rs\alpha_i + s.0 + 0 + \sum_{j,k} \varepsilon_{ijk}\right) - M$$

$$= M + \alpha_i + \frac{s.0}{rs} + \frac{0}{rs} + \frac{1}{rs}\sum_{j,k} \underbrace{E\varepsilon_{ijk}}_{=0} - M = \alpha_i$$

وأيضا

$$E\hat{\beta}_j = E\left(y_{.j.} - y_{...}\right) = \frac{1}{ts}\sum_{i,k} Ey_{ijk} - Ey_{...}$$

$$= \frac{1}{ts}E\left(tsM + ts.0 + ts.\beta_j + 0 + \sum_{i,k} \varepsilon_{ijk}\right) - M$$

$$= M + 0 + \beta_j + \frac{0}{ts} + \frac{1}{ts}\sum_{i,k} \underbrace{E\varepsilon_{ijk}}_{=0} - M = \beta_j$$

وكذلك

$$E\hat{\gamma}_{ij} = E\left(y_{ij.} - y_{i..} - y_{.j.} + y_{...}\right)$$

$$E\hat{\gamma}_{ij} = E\left(\left(y_{ij.} - y_{...}\right) - (y_{i..} - y_{...}) - \left(y_{.j.} - y_{...}\right)\right)$$

$$= E\left(y_{ij.} - y_{...}\right) - E\hat{\alpha}_i - E\hat{\beta}_j$$

$$= Ey_{ij.} - M - \alpha_i - \beta_j = E\left(\frac{1}{k}\sum_{k} y_{ijk}\right) - M - \alpha_i - \beta_j$$

$$= \frac{1}{s}\sum_{k} Ey_{ijk} - M - \alpha_i - \beta_j$$

$$= \frac{1}{s} \sum_{k} \left(M + \alpha_i + \beta_j + \gamma_{ij} + \underbrace{E \varepsilon_{ijk}}_{=0} \right) - M - \alpha_i - \beta_j$$

$$= \frac{1}{s} (sM + s\alpha_i + s\beta_j + s\gamma_{ij}) - M - \alpha_i - \beta_j = \gamma_{ij}$$

تذكير: العبارة $E(X)$ تعني التوقع الرياضي للمتغير العشوائي X .

∴ الآن في كل خانة (ij) نعرف

$$\forall\, i,j \;\; ; \;\; \omega_{ij} = \frac{1}{s-1} \sum_{k=1}^{s} \left(y_{ijk} - y_{ij\cdot} \right)^2$$

لدينا

$\frac{y_{ijk} - y_{ij\cdot}}{\sigma} \sim \mathcal{N}(0\,,1)$ ومنه $\sum_{k=1}^{s} \frac{(y_{ijk} - y_{ij\cdot})^2}{\sigma^2} \sim \chi^2_{(s-1)}$ (لوجود علاقة)

وبالتالي

$$\frac{(s-1)}{\sigma^2} \omega_{ij} \sim \chi^2_{(s-1)}$$

نتيجة: المقدر $\hat{\sigma}^2 = \frac{1}{rt} \sum_{i,j} \omega_{ij}$ هو مقدر منصف للتباين σ^2 .

بالفعل: لدينا $\sum_{i,j} \frac{s-1}{\sigma^2} \omega_{ij} \sim \chi^2_{rt(s-1)}$ ومنه $\frac{rt(s-1)}{\sigma^2} \left(\frac{1}{rt} \sum_{i,j} \omega_{ij} \right) \sim \chi^2_{rt(s-1)}$ أي أن

$$\frac{rt(s-1)}{\sigma^2} \hat{\sigma}^2 \sim \chi^2_{rt(s-1)}$$

وبالتالي $E\left(\frac{rt(s-1)}{\sigma^2} \hat{\sigma}^2 \right) = rt(s-1) \Leftarrow$ $\boxed{E\hat{\sigma}^2 = \sigma^2}$

نعلم أن حسب نظرية كوكران:

$$\boxed{\sum_{i,j,k} y_{ijk}^2 = trsy_{\cdots}^2 + rs \sum_{i=1}^{t} \hat{\alpha}_i^2 + ts \sum_{j=1}^{r} \hat{\beta}_j^2 + s \sum_{i,j} \hat{\gamma}_{ij}^2 + \sum_{i,j,k} \left(y_{ijk} - y_{ij\cdot} \right)^2}$$

بالفعل لدينا

$$\sum_{i,j,k}(y_{ijk}-y_{\cdots})^2=\sum_{i,j,k}y_{ijk}^2-trsy_{\cdots}^2$$

$$\sum_{i,j,k}(y_{ijk}-y_{\cdots})^2=\sum_{i,j,k}(y_{ijk}-y_{ij\cdot}+y_{i\cdot\cdot}-y_{\cdots}+y_{\cdot j\cdot}-y_{\cdots}+y_{ij\cdot}-y_{\cdots}+y_{\cdots})^2$$

$$=\sum_{i,j,k}\Big((y_{ijk}-y_{ij\cdot})+(y_{i\cdot\cdot}-y_{\cdots})+(y_{\cdot j\cdot}-y_{\cdots})+(y_{ij\cdot}-y_{i\cdot\cdot}-y_{\cdot j\cdot}+y_{\cdots})\quad\Big)^2$$

$$=\sum_{i,j,k}\Big((y_{ijk}-y_{ij\cdot})+\hat{\alpha}_i+\hat{\beta}_j+\hat{\gamma}_{ij}\quad\Big)^2$$

$$=\sum_{i,j,k}(y_{ijk}-y_{ij\cdot})^2+rs\sum_i\hat{\alpha}_i^2+ts\sum_j\hat{\beta}_j^2+s\sum_{i,j}\hat{\gamma}_{ij}^2$$

ومنه درجة الحرية للإحصائية $\sum_{i,j}\hat{\gamma}_{ij}^2$ تعطى بالشكل

$$trs-1-(t-1)-(r-1)-tr(s-1)=(t-1)(r-1)$$

الآن ليكن الإختبار التالي :

$$H_0:\forall\, i,j\ ;\ \gamma_{ij}=0$$

$$H_1:\exists\, i,j\ ;\ \gamma_{ij}\neq 0$$

H_0 معناه لا يوجد تفاعل بين المعاملين.

لدينا $\frac{s}{\sigma^2}\sum_{i,j}\hat{\gamma}_{ij}^2\sim\chi^2_{(t-1)(r-1)}$

$\chi^2_{(t-1)(r-1)}$ لا نستطيع أن نحصل عليه لأن σ مجهول ولهذا يجب تقديره ثم أخذ الإحصائية التالية:

$$\mathcal{U}=\frac{s\sum_{i,j}(y_{ij\cdot}-y_{i\cdot\cdot}-y_{\cdot j\cdot}+y_{\cdots})^2/(t-1)(r-1)}{\sum_{i,j,k}(y_{ijk}-y_{ij\cdot})^2/rt(s-1)}\sim\mathcal{F}\big((t-1)(r-1)\,,rt(s-1)\big)$$

حيث $\mathcal{F}$ هو توزيع فيشر وهي القيمة الجدولية، و $\mathcal{U}$ هي القيمة المحسوبة .

عندئذ يكون القرار: نقبل H_0 إذا كانت القيمة الجدولية أصغر من القيمة المحسوبة.

مثال: في إحدى تجارب تغذية الحيوانات وضعت 4 حيوانات اختيرت عشوائياً في كل 18 قفص وكانت المعاملات هي 3 انواع من العلائق تختلف في ما تحتويه من الذرة وقد وزعت هذه المعاملات الثلاث بطريقة عشوائية على الأقفاص (الوحدات التجريبية) حيث خصصت لكل عليقة ست حظائر وبعد فترة محددة سجلت الزيادة في اوزان الحيوانات وكانت موضحة كما في الجدول التالي، هل هناك تأثير معنوي للعلائق في الزيادة الوزنية للحيوانات اختبر ذلك تحت مستوى احتمال 0.05؟

Treat المعاملات	الوحدات التجريبية الأقفاص		مجموع الوحدات التجريبية $y_{ij\cdot}$	مجموع المعاملات $y_{i\cdot\cdot}$
T1 38% corn	1	2.9 3.1 3.0 3.0	3.0	3.2
	2	3.2 3.3 3.2 3.1	3.2	
	3	3.2 3.3 3.4 3.3	3.3	
	4	3.2 2.8 2.9 3.0	3.0	
	5	3.3 3.5 3.4 3.4	3.4	
	6	3.4 3.5 3.5 3.6	3.5	
T2 48% corn	1	3.1 3.4 3.4 3.3	3.3	3.1
	2	2.9 2.9 3.0 3.1	3.0	
	3	3.3 3.4 3.2 3.2	2.3	
	4	3.7 3.5 3.6 3.5	3.6	
	5	2.9 3.1 2.8 3.0	3.0	
	6	3.7 3.6 3.5 3.6	3.6	
T3 50% corn	1	2.7 2.7 2.8 2.9	2. 8	2.9
	2	3.1 2.9 3.0 3.1	3.0	
	3	2.7 2.7 2.6 2.7	2.7	
	4	2.8 2.9 2.9 3.0	2.9	

	5	3.2 3.0 3.0 3.2	3.1	
	6	2.8 2.7 2.6 2.6	2.7	

$y_{\cdots} = 3.0$

$\boxed{i = 1}$	$y_{1j\cdot} - y_{1\cdot\cdot} - y_{\cdot j\cdot} + y_{\cdots}$	$\left(y_{1j\cdot} - y_{1\cdot\cdot} - y_{\cdot j\cdot} + y_{\cdots}\right)^2$
$j = 1$	-0.2	0.1
$j = 2$	-0.1	0.0
$j = 3$	0.3	0.1
$j = 4$	-0.4	0.1
$j = 5$	0.0	0.0
$j = 6$	0.0	0.0
		0.3

$\boxed{i = 2}$	$y_{2j\cdot} - y_{2\cdot\cdot} - y_{\cdot j\cdot} + y_{\cdots}$	$\left(y_{2j\cdot} - y_{2\cdot\cdot} - y_{\cdot j\cdot} + y_{\cdots}\right)^2$
$j = 1$	0.2	0.0
$j = 2$	-0.2	0.0
$j = 3$	-0.6	0.3
$j = 4$	0.3	0.1
$j = 5$	-0.3	0.0
$j = 6$	0.2	0.0
		0.4

$\boxed{i=3}$	$y_{3j\cdot}-y_{3\cdot\cdot}-y_{\cdot j\cdot}+y_{\cdots}$	$(y_{3j\cdot}-y_{3\cdot\cdot}-y_{\cdot j\cdot}+y_{\cdots})^2$
$j=1$	-0.1	0.0
$j=2$	0.0	0.0
$j=3$	0.0	0.0
$j=4$	0.0	0.0
$j=5$	0.0	0.0
$j=6$	-0.5	0.2
		0.2

$$\sum_{i,j}(y_{ij\cdot}-y_{i\cdot\cdot}-y_{\cdot j\cdot}+y_{\cdots})^2=0.3+0.4+0.2=0.9$$

$$s\sum_{i,j}\left(y_{ij\cdot}-y_{i\cdot\cdot}-y_{\cdot j\cdot}+y_{\cdots}\right)^2/(t-1)(r-1)=4\times 0.9/(3-1)\,(6-1)$$

$$=0.36$$

$$\sum_{i,j,k}(y_{ijk}-y_{ij\cdot})^2=\sum_{j,k}(y_{1jk}-y_{1j\cdot})^2+\sum_{j,k}(y_{2jk}-y_{2j\cdot})^2+\sum_{j,k}(y_{3jk}-y_{3j\cdot})^2$$

$$\Sigma_{i,j,k}(y_{ijk}-y_{ij\cdot})^2=0.19+4.03+0.16=4.38$$

$$\Rightarrow\sum_{i,j,k}\left(y_{ijk}-y_{ij\cdot}\right)^2/rt(s-1)=4.38/6\times 3\times(4-1)=0.08$$

عندئذ قيمة $\mathcal{U}$ المحسوبة هي:

$$\mathcal{U}=\frac{0.36}{0.08}=4.5$$

يبقى حساب القيمة الجدولية $\mathcal{F}(10\,,54)$ عند مستوى المعنوية 0.05 . وهي 2.0095

وجدنا القيمة المحسوبة أكبر من القيمة الجدولية وهذا يعني رفض الفرضية H_1.

الفصل الثالث

تمارين

تمارين الفصل الأول

تمرين1: أكمل جدول التوزيع التكراري التالي

الفئات	التكرار n_i	مركز الفئات x_i	الحدود الحقيقية	التكرار النسبي	التكرار المئوي	التكرار المتجمع الصاعد	التكرار المتجمع النازل
-14	4	12					
-	8	17					
-	5	22					
-	3	27					

بعد اكمالك جدول التوزيع التكراري السابق اوجد ما يأتي:

1-الوسط الحسابي   3- المنوال   5- التباين

2- الوسيط   4- الإنحراف المتوسط   6- الإنحراف المعياري

تمرين2: لتكن لدينا بيانات في جدول التوزيع التكراري التالي:

الفئات	التكرار n_i
16 – 12	15
22 – 27	16
28 – 33	25
34 – 39	20
40 - 45	10
	86

اوجد ما يأتي :

1-الوسط الحسابي   3- المنوال   5- التباين

2- الوسيط   4- الإنحراف المتوسط   6- الإنحراف المعياري

تمرين3: لتكن البيانات التالية

x_i=15, 20, 23, 26, 27, 29, 30, 33

اوجد مايلي:

1-الوسط الحسابي   3- الانحراف المتوسط   5- الانحراف المعياري

2- الوسيط   4- التباين   6- الخطأ القياسي

تمرين4: أكمل جدول التوزيع التكراري التالي

الفئات	التكرار n_i	مركز الفئاتx_i	الحدود الحقيقية	التكرار النسبي	التكرار المئوي
-10	3	12			
-	7	17			

-	10	22			
-	15	27			
-	12	32			
-	7	37			
-	6	42			

بعد اكمال جدول التوزيع التكراري السابق اوجد ما يأتي:

1-الوسط الحسابي 3-المنوال 5- التباين

2- الوسيط 4- الإنحراف المتوسط 6- الإنحراف المعياري

تمرين5: لتكن البيانات التالية :

x_i=2 , 4 , 6 , 8 , 10 , 12 , 14 , 17 , 20 , 2

اوجد ما يأتي:

الوسط الحسابي , الوسيط , المنوال , الإنحراف المتوسط , التباين , الإنحراف المعياري, الخطأ القياسي.

تمرين6: البيانات التالية تمثل كمية الهيموغلوبين في دم 9 رجال مقاس بالغرام لكل 100 مل؟ اوجد ما يأتي :1- الوسط الحسابي 2- الوسيط 3- المنوال

x_i=12 , 9 , 10 , 11 , 13 , 8 , 14 , 15 , 10

تمرين7: أكمل جدول التوزيع التكراري التالي علماً ان اطوال الفئات متساوية وأنها ارقام صحيحة؟

الفئات	التكرار n_i	الحدود الحقيقة
1-	4	
-		
-		

		-
		-

تمرين8: إذا علمت بأن مراكز الفئات لأعمار عدد من طلبة كلية العلوم الدقيقة وعلوم الطبيعة والحياة هي :18 , 21 , 24 , 27 , 30 فما هي:

1- طول الفئة 2- الحدود الحقيقية للفئات 3- حدود الفئات لهذا التوزيع؟

تمرين9: البيانات التالية تمثل كمية المطر الساقطة سنوياً (بالمليمترات) على ولاية تبسة خلال فترة خمس سنوات فما هو متوسط سقوط المطر خلال هذه الفترة؟

x_i= 400 , 380 , 450 ,350 , 520

تمرين10: البيانات التالية تمثل اطوال 80 نبات من نباتات الذرة الصفراء، اكمل جدول التوزيع التكراري واوجد ما يأتي 1-الوسط الحسابي 2-الوسيط 3-المنوال 4-الإنحراف المتوسط 5- التباين 6- الإنحراف المعياري 7- ارسم المضلع التكراري لأطوال هذه البيانات 8- ارسم المدرج التكراري لأطوال هذه البيانات 9- التكرار النسبي لهذه البيانات 10- التكرار المئوي لهذه البيانات 11- التكرار المتجمع الصاعد 12- التكرار المتجمع النازل.

حدود الفئات	التكرار n_i	مركز الفئات x_i	الحدود الحقيقية
-40	1	*35.5*	
-50	2	*45.5*	
-60	5	55.5	
-70	15	65.5	
-80	25	75.5	
-90	20	85.5	
-100	12	95.5	

تمرين11: إذا علمنا البيانات المدونة أدناه اوجد ما يلي:

1-الوسط الحسابي 2- الوسيط 3-المنوال 4-الإنحراف المتوسط 5- التباين 6- الإنحراف المعياري 7- الخطأ القياسي ؟

1- x_i= 3 , 5 ,2 ,6 ,9 ,5 ,2 ,8 ,6

2- x_i= 51.6 , 48.7 , 50.3, 49.5 , 48.9

تمرين12: لكل من المجموعات التالية اوجد ما يأتي:

1-التباين 2- المدى 3-الإنحراف المعياري 4-الخطأ القياسي 5- الإنحراف المعياري 6- معامل الإختلاف

(a) x_i= 2 , 5 , 9 ,11 ,13

(b) x_i= 4 , 10 , 2 ,8 ,4 ,14 ,10 ,12 ,8

(c) x_i= -4 , 2 , -6 , 0 , -4 , 6 , 2 , 4 , 0

(d) x_i=16 , 2 , 22 , 8 , 6 , 20 ,24 , 14

تمرين13: في كل من الأفرع التالية اوجد القيمة المفقودة:

الفرع	$\bar{y}$	S	$C.V$
(a)		10	%20
(b)	50	30	
(c)	25		%5

تمارين الفصل الثاني:

تمرين1: اراد أحد الباحثين مقارنة لقاحين في علاج مرض داء الكلب، فقسم مجموعة مصابة بهذا المرض إلى مجموعتين، اعطيت المجموعة الأولى جرعة من لقاح رقم 1 واعطيت الثانية جرعة من لقاح رقم 2، سجلت الإستجابة لهذين اللقاحين بعد اسبوعين، وكانت المجموعتان مستقلتان غير متجانستان، كان الوسط الحسابي والإنحراف المعياري وحجم المجموعتين كما مبين ادناه: هل هناك فروقات معنوية في فعالية اللقاحين المستخدمين اختبر ذلك تحت مستوى احتمال 0.05؟

المجموعة	حجم العينة	$\bar{y}$	S
1	10	4.5	2.5
2	9	2.5	2.0

الحل:

نجري فحص التجانس (في الصفحة 57 من المحاضرة)

في حالة التجانس (انظر المثال في الصفحة 57 من المحاضرة)

في حالة عدم التجانس (انظر المثال في الصفحة 60 من المحاضرة)

تمرين2: زرع صنفان من القمح لمقارنة الفرق في عدد التفرعات وقد تم حساب هذه التفرعات لعشر نباتات من الصنف الأول وثمان نباتات من الصنف الثاني كان عدد التفرعات لكل نبات ضمن كل صنف كما في الجدول، هل هناك فرق معنوي في عدد التفرعات بين الصنفين؟ اختبر ذلك تحت مستوى احتمال 0.05 علماً ان العينتان مستقلتان متجانستان.

ت	عدد التفرعات للصنف الأول	عدد التفرعات للصنف الثاني
1	6	6
2	4	5
3	0	2
4	6	7

5	3	5
6	5	6
7	2	4
8	1	4
9	5	
10	4	

تمرين3: عرضت عشرة حيوانات إلى الإصابة بمرض معين وقد تم قياس عدد ضربات القلب لكل دقيقة قبل وبعد الإصابة بالمرض، الإصابة ادت الى زيادة عدد ضربات القلب لكل دقيقة؟ اختبر ذلك تحت مستوى احتمال 0.05؟

عدد ضربات القلب قبل الاصابة بالمرض	عدد ضربات القلب بعد الاصابة بالمرض
70	115
84	148
88	176
110	191
105	158
100	178
110	179
67	140
79	161
86	157

تمرين4: ادعت احدى الشركات بأنها انتجت غذاءاً يستطيع تخفيض الوزن بمقدار 10 كغم خلال اربعة اسابيع وقد استخدم هذا الغذاء 6 اشخاص وكان مقدار النقص في اوزانهم كالآتي علماً ان النقص مقاس بالكيلوغرام

5.7 , 11.9 , 11.4 , 6.4 , 10.1

فهل ادعاء الشركة صحيح؟ اختبر ذلك تحت مستوى 0.05 ؟

الحل: خطوات الإختبار

الفرضيات

$$H_0 \ : M = 10$$

$$H_1 \ : M \neq 10$$

اختبار الفرضية

$$t = \frac{\overline{X} - M}{\frac{S}{\sqrt{n}}}$$

$$\overline{X} = \frac{5.7 + 11.9 + 11.4 + 6.4 + 10.1}{5} = 9.1$$

$$S^2 = \frac{(5.7 - 9.1)^2 + (11.9 - 9.1)^2 + (11.4 - 9.1)^2 + (6.4 - 9.1)^2 + (10.1 - 9.1)^2}{5 - 1} = 8.245$$

$$\Rightarrow S = \sqrt{8.245} = 2.871$$

$$t = \frac{9.1 - 10}{\frac{2.871}{\sqrt{5}}} = -0.70096$$

إستخراج قيمة t الجدولية لمستوى معنوية 0.05 ودرجة حرية 4

وهي 2.1318

الإستنتاج:

بمان القيمة المطلقة للقيمة المحسوبة اقل من الجدولية تقبل فرضية العدم ونرفض الفرضية البديلة أي ان ادعاء الشركة صحيح.

تمرين5: اراد باحث ان يدرس تأثير معالجة معينة على تحسين الأداء عند مجموعة من الأفراد يعانون من ضعف مهارات الحاسوب، فأختار عينة مؤلفة من 10 افراد وطبق عليهم اختبار قبل تعرضهم للبرنامج والمعالجة ثم طبق عليهم اختباراً بعد المعالجة وحصل على البيانات التالية، هل هناك تأثير للبرنامج على تحسين اداء الأفراد اختبر ذلك تحت مستوى 0.05؟

ت	درجات الإختبار القبلي	درجات الإختبار البعدي
1	5	8
2	4	6
3	6	8
4	3	5
5	5	7
6	6	6
7	2	4
8	5	6
9	6	8
10	5	7

تمرين6: اراد احد الباحثين ان يدرس طريقة الحاسوب وطريقة النقاش على التحصيل في اللغة الإنجليزية عند عينة من طلبة الصف الثاني متوسط، فأختار عينة عشوائية من 50 طالباً قام بتوزيعهم على الطريقتين بالتساوي اي بمعدل 25 طالباً لكل مجموعة وبعد ان تم تعريض كل مجموعة لطريقة من الطرق، طبق عليهم اختباراً تحصيلياً في اللغة الإنجليزية وحصل على البيانات الآتية اختبر ذلك تحت مستوى احتمال 0.05 , 0.01 ؟

الطريقة	حجم العينة	متوسط التحصيل $\bar{y}$	s
الحاسوب	25	75	8
المناقشة	25	70	6

ملاحظة: العينتان مستقلتان متجانستان.

تمرين7: لدراسة صنف الطول في نبات البزاليا اخذت عينة مؤلفة من 40 نبات من نباتات الجيل الثاني الناتجة من تزاوج نبات قصير مع نبات طويل نقي فوجد ان بينها 31 نباتاً طويلاً فهل تتفق هذه النتائج مع النسبة 3 طويل :1 قصير اختبر ذلك تحت مستوى احتمال 0.01؟

تمرين8: في الجيل الثاني للتهجين ما بين نبات ذو بذور لونها اصفر وشكلها مستدير مع نبات ذو بذور لونها اخضر وشكلها مدور حصل مندل على ما يلي، هل ان الصفتان مستقلتان، اختبر ذلك تحت مستوى 0.05؟

لون البذور	شكل البذور		*Total*
	مستديرة	مجعدة	
صفراء	320	100	420
خضراء	110	30	140
Total	430	130	560

تمرين9: عوملت بذور القمح بنوعين من المبيدات الفطرية لعلاج مرض فطري معين وكانت النتائج كالآتي، فهل هناك علاقة بين المبيدات وحالة البذور اختبر ذلك تحت مستوى احتمال 0.05؟

نوع المبيد	حالة البذور		Total
	بذور حية	بذور ميتة	
المبيد الاول	120	25	145
المبيد الثاني	80	50	130
Total	200	75	275

تمرين10: اراد باحث ان يعرف الفرق بين نسبة المؤيدين لفكرة اعطاء المرأة حق الترشيح للانتخابات النيابية ونسبة المؤيدات من النساء للفكرة نفسها، فقام باستطلاع اراء 200 رجلاً و150 امرأة وحسب نسبة المؤيدين للفكرة من الجنسين ووجدها (60%) عند الرجال و (80%) عند النساء هل توجد علاقة بين جنس الفرد ومسألة تأييد حق المرأة في الترشيح للانتخابات النيابية اختبر ذلك تحت مستوى احتمال 0.05؟