مبادئ الإحصاء الطبي

Principles of Biostatistics

المستوى الثاني / صيدلة

مدرس المادة:

د. جميل مجلي

الفصل الأول

1-1: مقدمة: "Introduction"،

يعد الإحصاء من المواضيع الرياضياتية التطبيقية، حيث يستخدم في كل التخصصات، كما تبرز الحاجة إليه في الكثير من الدراسات والتجارب والأبحاث وما إلى ذلك...،
يمكن لنا وبكل اختصار أن نعرف الإحصاء على أنه قواعد وأساليب وتقنيات تستخدم لدراسة حالات أو ظواهر ما من أجل الوصول للنتائج المرسومة...،
لقد ظهرت تسميات مختلفة لعلم الإحصاء، غير أن القواعد الإحصائية لا تختلف عنها في أي جانب إلا من خلال فقط نوعية البيانات المتمثلة في تلك الجوانب والتخصصات...،

2-1: الطريقة & المتغيرات الإحصائية: "Statistical Method & Variable"،

وهي قواعد علمية دقيقة تقود إلى ما يصبو إليه المعنيين، كما أنه من خلال الالتزام بها يكون عدم التحيز وتجنب الأخطاء...، كما أن لها مراحل هي:

1- **تحديد الهدف**، "ذلك أنه من خلال تحديد الهدف تتضح أوجه الاستفادة المرجوة..."،
2- **جمع البيانات**، "حيث يستلزم هنا تحديد أسلوب الحصول على البيانات والمعلومات المطلوبة للدراسة"، ومما تجدر الإشارة إليه هو أن عملية جمع البيانات لها أسلوبين هما:

- **"التسجيل الشامل"** وهنا يتم جمع البيانات من مجتمع الدراسة ككل، ولهذا الأسلوب ميزات وعيوب سيكون التعرف عليها تباعا...،
- **"العينات"** وهنا يتم جمع البيانات من مجموعة ما من مفردات المجتمع كبديل عن المجتمع ككل مع مراعاة أن يكون لكل مفردة من مفردات المجتمع نفس الفرصة من التمثيل في العينة، وأيضا لهذا الأسلوب ميزات وعيوب سيتم ملاحظتها تباعا...، إضافة إلى أن العينة تكون على وجهين هما:

أ: "عشوائية" وذلك في حال تجانس مفردات المجتمع وفي هذه الحالة إما أن تُأخذ العينة العشوائية:

بسيطة: وذلك عندما يكون لكل مفردة من مفردات المجتمع نفس فرصة الظهور...،
أسلوبية: وهنا يقسم مجتمع الدراسة إلى وحدات ومن ثم يتم وبطريقة عشوائية اختيار عدد معين ومتساوي من كل الوحدات التي مجموعها العينة المطلوبة...،

كما أن أساليب جمع البيانات تتنوع حسب طبيعة الدراسة والبيانات المطلوبة، من ذلك: "الاستمارة الإحصائية أو الإستبانة، المقابلات الشخصية، العودة للسجلات والمعلومات من المؤسسة المعنية،..."،

إضافة إلى أن البيانات الإحصائية تختلف من حيث الميزة والنوع، فقد تكون لها ميزة كمية "تترجم عدديا"، أو أن لها ميزة كيفية "لا تترجم إلى عدد"، وقد تكون البيانات متقطعة "مفرداتها كأرقام منفصلة"، أو متصلة "مفرداتها يعبر عنها في مجالات متصلة"، إضافة إلى أن المتغيرات قد تكون اسمية أو رتبيه وقد تكون نسبية أو فئوية...، وكل ذلك سنتعرض له تباعا...،

3- **تبويب وعرض وتحليل البيانات**، "حيث أنه بعد الحصول على البيانات المطلوبة يستلزم معالجتها بطرق علمية للاستفادة منها وإظهارها بدقة ووضوح ذلك ما سنقف عليه خلال المواضيع القادمة..."،

4- حساب المؤشرات الإحصائية، "أيضا سنقف على ذلك تباعا..."،

5- التفسير والتنبؤ،...،

3-1: تبويب وعرض وتحليل البيانات: " Classification & Showing & Processing of Data"،

- **طرق تبويب البيانات:** نبدأ بالمثال التالي،

مثال1: البيانات التالية جمعت حول تلوث الهواء (ميكرو/م^3) في 50 مدينة،

27	79	32	28	36	30	27	42	63	68	51	74	25	43	65	44	25
24	23	22	51	57	12	45	25	28	31	28	42	36	21	39	50	31	27
42	38	49	32	12	28	49	27	43	22	23	47	69	24	16		

من أجل تدوين البيانات في جدول نحتاج إلى تبسيطها من خلال فئات متساوية كل منها تضم عددا معين البيانات المندرجة داخل تلك الفئة، وفي ضوءه لدينا العلاقات الأساسية التالية:

المدى=القيمة الأعلى- القيمة الأدنى

$$\text{طول الفئة} = \frac{\text{القيمة الأعلى} - \text{القيمة الأدنى}}{\text{عدد الفئات}} \quad ، \quad \text{مركز الفئة} = \frac{\text{الحد الأعلى للفئة} + \text{الحد الأدنى للفئة}}{2}$$

ووفقا للبيانات المعطى أعلاه والتي يبلغ مداها القيمة "79-12=67"، يمكن لنا توزيع البيانات عبر سبع فئات طول كل منها 10،

وفي ضوءه يمكن لنا معالجة البيانات السابقة في جدول تكراري كما يلي:

الفئات	10-	20-	30-	40-	50-	60-	70- 80	المجموع
التكرار	\\\	~~\\\\~~~~\\\\~~~~\\\\~~ \\\	~~\\\\~~ \\\\	~~\\\\~~~~\\\\~~	\\\\	\\\\	\\	50
	3	18	9	10	4	4	2	

والجدول السابق يمثل جدول تكراري بسيط،

ملاحظة1: تجدر الإشارة إلى أن الكيفية التي يتم بها توزيع الفئات لا تؤثر في النتائج، ذلك أن عدد الفئات وأطوالها يمكن تمثيله بالطرق التي يتم من خلالها تخزين وترتيب الأشياء والتي هي بالطبع تختلف من شخص لآخر، غير أن ذلك الاختلاف لا يؤثر على المقادير الكمية أو العددية...، إضافة إلى أننا وخلال الأمثلة اللاحقة سنحاول تبويب وتصنيف بيانات معينة بأكثر من أسلوب وذلك بالتغيير في عدد وأطوال الفئات وذلك من أجل التحقق من أننا سنصل لنفس النتائج...،

وحول جدول تكراري مزدوج نعطي المثال التالي،

مثال2: البيانات التالية تبين الوزن "X" والطول "Y" لعينة ما،

"X": 60 55 65 70 75 60 90 83 65 56 46 65 49 48 69 60 72 45 56 74،

"Y": 160 159 175 170 175 165 180 175 173 160 159 170 157 165 170 165 170 154 170 154،

وسيكون الجدول التكراري المزدوج كما يلي:

"X" / "Y"	-40	-50	-60	-70	90-80	المجموع
-150	\\\		\	\		5
-160	\	\\	\\			5
180-170		\	\\\\	\\\	\\	10
المجموع	4	3	7	4	2	20

إضافة لما سبق، يمكن لنا تكوين ما يسمى بالجدول التكراري المتجمع الصاعد" ك. م. ص"، أو الجدول التكراري المتجمع النازل" ك. م. ز"، فباعتبار الجدول التكراري المتكون من البيانات المعطى في المثال1،

سيكون الجدول التكراري المتجمع الصاعد كما يلي:

الفئات	-10	-20	-30	-40	-50	-60	80 -70	المجموع
التكرار	3	18	9	10	4	4	2	50
ك. م. ص	3	21	30	40	44	48	50	

وسيكون الجدول التكراري المتجمع النازل كما يلي:

الفئات	-10	-20	-30	-40	-50	-60	70- 80	المجموع
التكرار	3	18	9	10	4	4	2	50
ك. م. ز	50	47	29	20	10	6	2	

نشاط/ 1) ناقش بالأمثلة الفقرات من 1- 3) الواردة الفقرة بعنوان "الطريقة الإحصائية"،

2) أوجد الجدول التكراري التجمع الصاعد وكذلك المتجمع النازل للبيانات الواردة في مثال2،

- أساليب عرض البيانات الإحصائية:

سنتطرق فيما يلي لبعض أساليب عرض البيانات،

أولاً: إذا كانت البيانات غير مبوبة، فإن من الأساليب المستخدمة لعرض البيانات: "الرسوم، الشريط البياني، المستطيلات، الدوائر البيانية،..."،

مثال3: البيانات التالية تمثل عدد المستشفيات بعشر محافظات في الجمهورية اليمنية حتى عام 2006،

المحافظة	صنعاء	عدن	تعز	حضرموت	إب	الحديدة	ذمار	لحج	مأرب	صعدة	المجموع
ع. المستشفيات	24	12	13	9	4	19	9	4	2	2	98

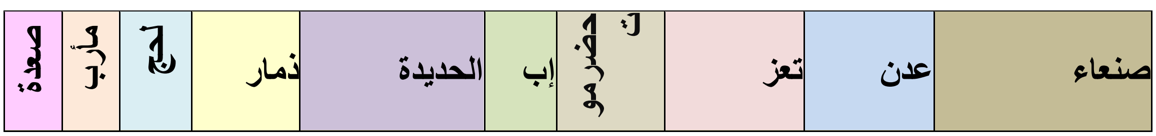
ولعرض تلك البيانات في مستطيل بياني نحتاج للقاعدة: القاعدة الجزئية = القاعدة الكلية × $\frac{\text{البيانات الجزئية}}{\text{البيانات الكلية}}$

وباعتبار القاعدة الكلية للبيانات السابقة 10 سم، سيكون لدينا:

ق$_1$= 10 $\frac{24}{98}$×=2.4 سم، ق$_2$= 10 $\frac{12}{98}$×=1.2 سم، ق$_3$= 10 $\frac{13}{98}$×=1.32 سم، وبنفس الكيفية سيكون:

ق$_4$=0.9 سم، ق$_5$=0.4 سم، ق$_6$=1.94سم، ق$_7$=0.9 سم، ق$_8$=0.4 سم، ق$_9$=0.2 سم، ق$_{10}$=0.2 سم،

والمستطيل التالي هو التمثيل المطلوب:



مثال4: البيانات التالية تمثل عدد الأطباء في خمسة مستشفيات حكومية:

المستشفى	الجمهوري	الكويت	الثورة	السبعين	العسكري	المجموع
ع. الأطباء	117	82	115	72	112	498

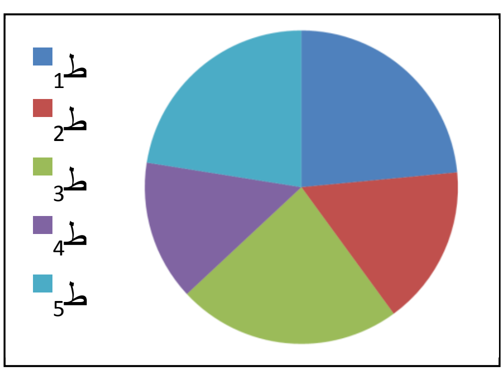
ولعرض البيانات في دائرة بيانية، نحتاج للقاعدة: زاوية القطاع الجزئي= 360 × $\frac{\text{البيانات الجزئية}}{\text{البيانات الكلية}}$

وعليه سيكون: ط$_1$= 360 × $\frac{117}{498}$=85، ط$_2$= 360 × $\frac{82}{498}$=59،

وبنفس الكيفية: ط$_3$=83، ط$_4$=52، ط$_5$=81،

حيث ط$_1$- ط$_5$ ترمز لزاوية القطاعات التي تمثل المشافي حسب ترتيبها في الجدول أعلاه،

والشكل المقابل هو المطلوب،

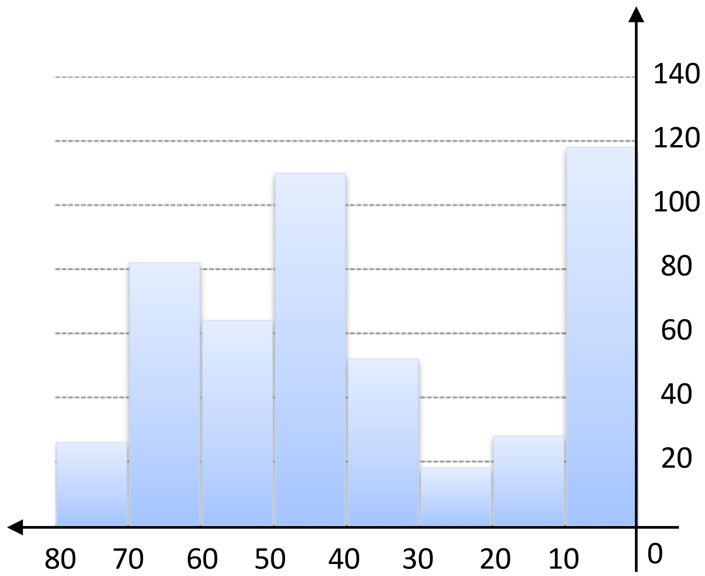


ثانياً: إذا كانت البيانات مبوبة في جدول تكراري فإن ما يعرف بالمدرج التكراري أو المضلع التكراري يمثلان أشهر طرق عرض مثل تلك النتائج المبوبة.. .،

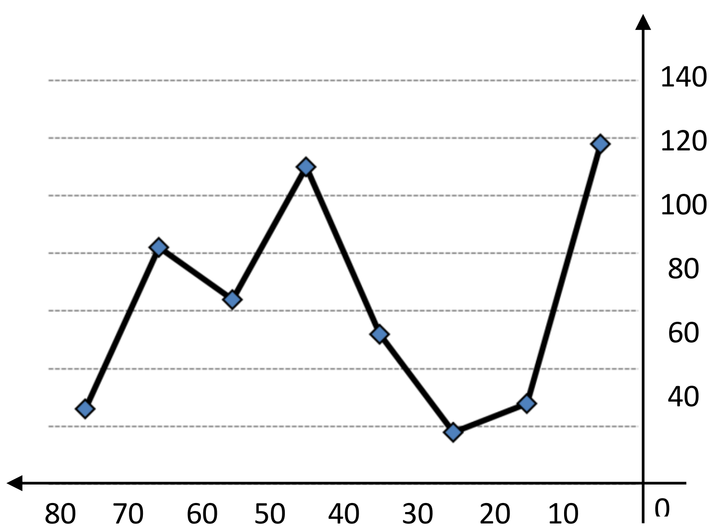
مثال5: الجدول التكراري التالي يمثل عدد المراجعين بمستشفى الكويت للفترة من 8-2008 إلى 2-2009،

فئات العمر	0-	10-	20-	30-	40-	50-	60-	70- 79	المجموع
التكرار	118	28	18	52	110	64	82	26	498

والشكل المقابل يمثل المدرج التكراري،



هذا ومن أجل المضلع التكراري نقوم بالتوصيل بين مراكز الفترات بمستقيمات حسب ما هو مبين في التالي،



ملاحظة2: هناك ما يسمى بالمنحنى التكراري ويستخدم أيضا لعرض البيانات المبوبة، كما أن عملية إنشائه تتم باستبدال الخطوط المستقيمة التي تصل بين مراكز الفئات في الشكل السابق بمنحنيات تصل بين مراكز الفئات...،

نشاط/ 1) إذا طلب منك عمل بحث حول ارتفاع معدل الوفيات في اليمن، حدد العوامل المؤثرة في ذلك ومصادر جمع البيانات.

2) أذكر مزايا وعيوب استخدام أسلوب العينات.

3) أجريت تجربة على مجتمع من سيقان النبات تتألف من 50 نبتة، ووجد أن عدد الجذور هو:

7 1 3 3 6 2 4 3 2 5 6 5 7 7 4 0 4 9 3 6 2 3 8 6 5 3 5
3 2 3 3 4 0 5 3 4 1 4 3 4 1 3 5 2 2 3 0 6 2 7

أ) بين نوع المتغيرات، ب) بوب البيانات في جدول تكراري، ج) إعرض البيانات بالشكل البياني المناسب،

4) البيانات التالية عدد المراكز الصحية بأربع محافظات يمنية حتى عام 2009:

المحافظة	لحج	الحديدة	المحويت	ذمار	المجموع
عدد المراكز	31	47	30	41	149

مثل ذلك بمستطيل بياني ودائرة بيانية.

5) صمم استمارة بحث لتحديد العوامل المؤثرة في مرض السكر.

6) البيانات التالية تمثل توزيع الوزن "كغم" لعينة من 20 مريض:

الوزن	40-	50-	60-	70-	80-90	المجموع
عدد المرضى	4	3	7	4	2	149

مثل ذلك بدرج تكراري ومنحنى تكراري.

4-1: المقاييس الإحصائية: "Statistical Measurements"،

تبعا لما سبق نتعرض لموضوع **"تحليل البيانات الإحصائية"** من خلال مقاييس إحصائية كما يلي،

1- مقاييس النزعة المركزية: وهي مقاييس توضيحية لتمثيل البيانات بقيمة واحدة تكون أنموذجا لها، قد تكون تلك القيمة من بين البيانات أو قريبة منها، كما أن لها نزعة نحو وسط البيانات حيث سميت بذلك...،

من تلك المقاييس: "الوسط الحسابي (Mean)، الوسيط (Median)، المنوال (Mode)، الوسط الهندسي (Geometric mean)،..."، ونشير إلى أن لكل مقياس مزايا وعيوب حسب ما سنقف عليه تباعا...،

الوسط الحسابي (Arithmetic Mean): وهو من أشهر مقاييس النزعة المركزية، كما أنه يستخدم لبيانات غير مبوبة وفقا للقاعدة $\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ ، أو لبيانات مبوبة وفقا للقاعدة $\overline{X} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$ ، حيث أن:

$\overline{X}$ يمثل المتوسط، x_i يمثل التغيرات، f_i يمثل التكرارات، $\sum$ يمثل رمز المجموع، n عدد المتغيرات،

من مزايا المتوسط: الوضوح والسهولة، **من عيوب المتوسط:** التحيز للقيم الشاذة،

مثال5: البيانات التالية تمثل عدد المراجعين للمستشفيات بخمسة عشر محافظة في العام 2008م،

المحافظة	صنعاء	عدن	تعز	حضرموت	إب	الحديدة	ذمار	لحج	مأرب	حجة	ريمة	صعدة	شبوة	عمران	البيضاء	م
العدد	41	16	20	4	1	59	21	22	33	43	12	0	15	44	15	346

لذا سيكون: $\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{346}{15} = 23$ "معدل عدد المرضى المراجعين من تلك المحافظات"،

مثال6: الجدول التالي يمثل توزيع مرضى السكر بالوراثة حسب فئات العمر بأمانة العاصمة للعام 2007م،

الفئات	10-	20-	30-	40-	50-	70-60	المجموع
العدد	10	5	9	19	11	1	55

وعليه سنكون الجدول المساعد التالي:

الفئات	10-	20-	30-	40-	50-	70-60	المجموع
f_i	10	5	9	19	11	1	55
x_i	15	25	35	45	55	65	
$f_i x_i$	150	125	315	855	605	65	2115

أي أن: $\overline{X} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{2115}{55} \approx 38$ "معدل عمر الشخص المصاب بمرض السكر وراثة"،

الوسيط (Median): وهو القيمة الواقعة في وسط بيانات مرتبة تصاعديا أو تنازليا،

أي أن الوسيط من بيانات "غير مبوبة" هو كما يلي:

- ترتيب الوسيط في عدد فردي من البيانات "$\frac{n+1}{2}$"،

- ترتيب الوسيط في عدد زوجي من البيانات هو معدل القيمتين التي ترتيبها هما "$\frac{n}{2}, \frac{n}{2}+1$"،

أما الوسيط من "بيانات مبوبة" فيحسب بالقاعدة: $m_e = L_i + \frac{\frac{\sum f_i}{2} - u_i - 1}{f_i} W$ ، حيث أن:

L_i بداية الفئة الوسطية، f_i تكرار الفئة الوسطية، $u_i - 1$ التكرار المتجمع الصاعد السابق، W طول الفئة،

$\frac{\sum f_i}{2}$ ترتيب الفئة الوسطية، كما أن m_e يمثل رمز الوسيط...،

من مزايا الوسيط: السهولة وعدم التأثر بالقيم الشاذة، **من عيوب الوسيط:** عدم إعطاء معلومات كافية،

مثال7: البيانات التالية تمثل الولادات المسجلة حسب السنوات كما يلي:

السنة	1991	1992	1993	1994	1995	1996	1997	1998	1999
العدد	79	83	89	91	103	135	151	172	186

وبترتيب البيانات تصاعديا: 79، 83، 89، 91، 103، 136، 151، 172، 186
وعدد البيانات فردي،

أي أن ترتيب الوسيط هو: $\frac{n+1}{2} = \frac{9+1}{2} = 5$ ،
بمعنى أن الوسيط هو 103، "معدل عدد الولادات"،

أما إذا كان لدينا:

السنة	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
العدد	79	83	89	91	103	135	151	172	186	152

وبترتيب البيانات تصاعديا: 79، 83، 89، 91، 103، 136، 151، 152، 172، 186
وعدد البيانات زوجي،

أي أن الوسيط هو معدل القيمتين التي ترتيبها: "$\frac{n}{2} = 5, \frac{n}{2}+1 = 6$"،

بمعنى أن الوسيط= $\frac{135+103}{2}=119$، "معدل عدد الولادات"،

مثال8: الجدول التالي يبين توزيع عدد من شملتهم حملة مكافحة الملاريا للفترة 1990- 1995،

ف.العمر	10-	20-	30-	40-	50-	70-60	المجموع
العدد	20	22	13	11	17	7	90

سنكون الجدول المساعد التالي:

ف.العمر	10-	20-	30-	40-	50-	70-60	المجموع
f_i	20	22	13	11	17	7	90
ك. م. ص.:	20	42	55	66	83	90	

ترتيب الفئة الوسطية "$\frac{\sum_{i=1}^{n} f_i}{2}=45$"، بداية الفئة الوسطية "30"، وطولها "10"، أيضا التكرار المتجمع الصاعد السابق 42، لذا سيكون: "$m_e = L_i + \frac{\frac{\sum f_i}{2} - u_i - 1}{f_i} W = 30 + \frac{45-42}{13} \times 10 \approx 31$"، معدل عمر المشمول،

المنوال (Mode): وهو القيمة الأكثر تكرر من بين البيانات المعطى، ويلاحظ أنه إذا كانت البيانات جميعها مختلفة فلا وجود للمنوال، وقد يكون هناك أكثر من قيمة للمنوال...، وفي حالة البيانات المبوبة فإن مركز الفئة التي تقابل أكبر تكرار هو ما يمثل قيمة المنوال،

مثال9: البيانات التالية تمثل الأشخاص من فصيلة الدم B بمنطقة السبعين للفترة 1995- 1999،

السنة	1995	1996	1997	1998	1999
العدد	35	39	30	32	30

أي أن المنوال هو القيمة 30، "معدل عدد الأشخاص من فصيلة الدم B"،

أما إذا كانت لدينا البيانات بالصورة:

السنة	1995	1996	1997	1998	1999
العدد	34	32	28	39	31

فإنه لا يوجد منوال...،

مثال10: الجدول التالي يمثل توزيع المكفولين بالضمان الاجتماعي حسب فئة العمر للعام 2000،

ف.العمر	-10	-20	-30	-40	-50	70-60
f_i	21	27	47	85	144	75

ويمثل المنوال مركز الفئة المقابلة لأكبر تكرار، أي أن المنوال هو 55،

الوسط الهندسي (Geometric mean): ويستخدم لإيجاد نسبة الزيادة أو النقص في قيم الظواهر كالولادات والوفيات والسكان...،

لإيجاد الوسط الهندسي من بيانات غير مبوبة، فإذا كان لدينا **n**-مفردة مثل $x_1, x_2, ..., x_n$، فإن الوسط الهندسي لها هو ما يتمثل بالجذر النوني لحاصل ضربها، أي أن: $G.m = \sqrt[n]{x_1 \times x_2 \times ... \times x_n}$، ولصعوبة الحسابات في بعض المسائل نستخدم اللوغاريتمات كما يلي: $\boldsymbol{log}\ G.m = \frac{\sum \boldsymbol{log}\ \boldsymbol{x_i}}{\boldsymbol{n}}$،

مثال11: لإيجاد الوسط الهندسي للقيم 12، 45، 50، سيكون:

$\boldsymbol{log}\ G.m = \frac{\sum \boldsymbol{log}\ \boldsymbol{x_i}}{\boldsymbol{n}} = \frac{\boldsymbol{log}\ 12 + \boldsymbol{log}\ 45 + \boldsymbol{log}\ 50}{3} = \frac{1.0792 + 1.6532 + 1.6990}{3} = 1.4771$، أي أن $G.m = 30$،

أما في حال البيانات المبوبة فإن القاعدة هي على الصورة $\boldsymbol{log}\ G.m = \frac{\sum f_i\ \boldsymbol{log}\ \boldsymbol{x_i}}{\sum f_i}$، حيث أن x_i هو ما يمثل مركز الفئة المقابلة للتكرار f_i لكل $\boldsymbol{i}$،

مثال12: الجدول التالي يمثل أجور الفحوصات الطبية اليومية لدى عينة مكونة من 65 طبيب،

ف.بالريال	-50	-60	-70	-80	-90	-100	120-110	المجموع
ع.الأطباء	8	10	16	14	10	5	2	65

سنكون الجدول المساعد التالي:

ف.بالريال	-50	-60	-70	-80	-90	-100	120-110	المجموع
ع.الأطباء	8	10	16	14	10	5	2	65
x_i	55	65	75	85	95	105	115	
$\boldsymbol{log}\ \boldsymbol{x_i}$	1.7404	1.8129	1.8751	1.9294	1.9777	2.0221	2.0607	
$f_i\ \boldsymbol{log}\ \boldsymbol{x_i}$	13.9232	18.1290	30.0016	27.0116	19.7770	10.7000	4.1214	123.0698

$$log\ G.m = \frac{\sum f_i\ log\ x_i}{\sum f_i} = \frac{123.0698}{65} = 1.8938$$ ، أي أن $G.m = 78.23$ ، "معدل أجر الفحص اليومي"،

نشاط/ 1) الجدول التالي يمثل توزيع مرضى السكر "إناث" حسب فئات العمر بأمانة العاصمة،

ف.العمر	-10	-20	-30	-40	-50	70-60
العدد	4	13	24	49	95	48

أوجد كل من: "الوسط الحسابي، الوسيط، المنوال، الوسط الهندسي".

2) البيانات التالية تمثل الأشخاص من فصيلة الدم ، بأربع مناطق في أمانة العاصمة:

المنطقة	التحرير	الصافية	معين	الثورة
عدد المراكز	91	202	25	1207

أوجد كل من: "الوسط الحسابي، الوسيط، المنوال، الوسط الهندسي".

3) البيانات التالية تمثل توزيع المرضى في قسم الباطنية حسب فئة العمر ونوع العلاج:

ف.العمر	-10	-20	-30	-40	-50	70-60
نوع العلاج	20	22	13	11	17	7

أوجد كل من: "الوسط الحسابي كمعدل لنوع العلاج، المنوال كمعدل للعمر".

2- مقاييس التشتت: مقاييس النزعة المركزية تعطي فكرة عن طبيعة البيانات ولا تعطي الوصف الكامل عن تمثيل البيانات، وللوقوف على درجات التفاوت بين قيم البيانات سنعطي المقاييس التالية "المدى (Range)، الانحراف الربيعي (Semi-inter-Quartite-Range)، الانحراف المتوسط (Mean deviation)، الانحراف المعياري (Standard deviation)،..." إضافة لما سنتطرق له من المقاييس النسبية مثل "معامل الاختلاف المعياري، معامل الاختلاف الربيعي"،

حيث سيكون الاستطراد أكثر في حال المقاييس المهمة والأكثر استخدام...،

- المدى (Range): وهو أبسط المقاييس وأقلها دقة لاعتماده على قيمتين متطرفتين فقط،

فإذا كان لدينا بيانات غير مبوبة نحسب مداها كالفرق بين أدنى وأعلى قيمة، أما في البيانات المبوبة نحسب المدى كالفرق بين الحد الأدنى للفئة الدنيا والحد الأعلى للفئة العليا...،

مثال13: البيانات التالية تبين معدل الوفيات للرضع بأمانة العاصمة للفترة 1994- 1999،

السنة	1994	1995	1996	1997	1998	1999
المعدل	41.4	76.9	46.2	43.4	47.1	39.8

لذا فإن المدى: $\boldsymbol{R} = \boldsymbol{x}_L - \boldsymbol{x}_S = 76.9 - 39.8 = 37.1$ ،

مثال14: الجدول التالي يبين توزيع المرضى بأحد المستشفيات الخاصة حسب فئة العمر ومعدل الدخل،

ف.العمر	-10	-20	-30	-40	-50	70-60
نوع العلاج	2	8	11	17	16	20

لذا فإن المدى: $\boldsymbol{R} = \boldsymbol{x}_L - \boldsymbol{x}_S = 70 - 10 = 60$ ،

-الانحراف المعياري (Standard deviation): وهو أهم مقاييس التشتت لدقته، وهو الأكثر استخدام،

فإذا كان لدينا بيانات غير مبوبة نحسب الانحراف المعياري لها بالقاعدة: $\boldsymbol{S} = \sqrt{\frac{\sum(\boldsymbol{x}_i - \bar{\mathrm{x}})^2}{\boldsymbol{n}-1}}$ ، حيث:

" $\boldsymbol{x}_i$ المفردة بالترتيب $\boldsymbol{i}$ ، $\bar{\mathrm{x}}$ المتوسط الحسابي، أيضا $\boldsymbol{S}^2 = \frac{\sum(\boldsymbol{x}_i - \bar{\mathrm{x}})^2}{\boldsymbol{n}-1}$ هو ما يعرف بالتباين(Variance)"،

وبنفس الكيفية إذا كانت البيانات مبوبة فإن الانحراف المعياري يحسب بالقاعدة: $S = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}}$ ،

حيث f_i يمثل التكرار في الترتيب i ،

مثال15: البيانات التالية كجزء من المعطى بمثال-5) سابقا، موضحة كما في الجدول المساعد التالي:

المحافظة	صنعاء	عدن	تعز	حضرموت	إب	الحديدة	ذمار	لحج	م
العدد	41	16	20	4	1	59	21	22	184
$(x_i - \bar{x})$	18	-7	-3	-19	-22	36	-2	-1	
$(x_i - \bar{x})^2$	324	49	9	361	484	1296	4	1	2528

حيث $\bar{x} = \frac{\sum x_i}{n} = \frac{184}{8} = 23$ ، أي أن: $S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{2528}{7}} = \sqrt{361.14} \approx 19$ ،

مثال16: الجدول التالي يبين عدد المراجعين بأحد المشافي حسب فئات العمر:

ف.العمر	10-	20-	30-	40-	50-	70-60
العدد	0	1	1	4	13	17

لحساب الانحراف المعياري نكون الجدول المساعد التالي:

ف.العمر	10-	20-	30-	40-	50-	70-60	المجموع
f_i	0	1	1	4	13	17	36
x_i	15	25	35	45	55	65	
$f_i\, x_i$	0	25	35	180	715	1105	
$(x_i - \bar{x})$	-42	-32	-22	-12	-2	8	
$(x_i - \bar{x})^2$	1764	1024	484	144	4	64	
$(x_i - \bar{x})^2 f_i$	0	1044	484	576	52	1088	3224

حيث $\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{2060}{36} \approx 57$،

أي أن: $S = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}} = \sqrt{\frac{3224}{36}} = \sqrt{89.555} = 9.46$،

- الدرجة المعيارية (Standard Scores): حيث أنه لمجموعة قيم $x_1, x_2, \ldots, x_n$، بانحراف معياري S، ووسط حسابي $\bar{x}$، فإن الدرجة المعيارية تحسب بالقاعدة $Z = \frac{x - \bar{x}}{S}$، مما يعني أن انحراف القيم عن المتوسط أصبحت مقاسه بوحدات الانحراف المعياري بعد أن كانت مقاسه بوحدات قياس القيم الأصلية، وتستخدم هذه العلاقة للمقارنة بين قيمتين في مجموعتين مختلفتين...،

مثال17: حصلت طالبة في مادة صحة مجتمع على درجة 60، وفي مادة الإحصاء على درجة 70، فإذا علمت أن الوسط الحسابي لدرجات صحة مجتمع هو 50 بانحراف معياري 2، وأن الوسط الحسابي لدرجات الإحصاء هو 65 بانحراف معياري 2.5، ففي أي المقررين مستوى الطالبة هو الأفضل.

لدينا، الدرجة المعيارية في صحة مجتمع هي: $Z = \frac{x - \bar{x}}{S} = \frac{60 - 50}{2} = 5$

الدرجة المعيارية في الإحصاء هي: $Z = \frac{x - \bar{x}}{S} = \frac{70 - 65}{2.5} = 2$

وفي ضوءه فإن مستوى الطالبة في مادة صحة مجتمع أفضل منه في الإحصاء.

- معامل الاختلاف (Sufficient Variation): وهو مقياس نسبي للمقارنة بين مجموعتين تختلف بوحدات القياس (كمقارنة تشتت أوزان مجموعة مع مستويات الكوليسترول)، وفقا للقاعدة: $C.V = \frac{S}{\bar{X}} \times 100$

مثال18: ليكن الوسط الحسابي لأوزان مجموعة أشخاص هو 75 كغم بانحراف معياري 5 كغم، والوسط الحسابي لأعمارهم هو 24 سنة بانحراف معياري 4 سنة،

لدينا، معامل الاختلاف للأوزان هو: $C.V = \frac{S}{\bar{X}} \times 100 = \frac{5}{75} \times 100 = 6.67\ \%$

معامل الاختلاف للأعمار هو: $C.V = \frac{S}{\bar{X}} \times 100 = \frac{4}{24} \times 100 = 16.6\ \%$

أي أن تشتت المجموعة من حيث العمر أكبر منه من حيث الوزن.

نشاط/ 1) التوزيع التالي يبين مرضى السكر من الذكور المراجعين لأحد المستشفيات، حسب فئات العمر وعدم الاستمرار في العلاج،

ف.العمر	10-	20-	30-	40-	50-	70-60
نوع العلاج	3	4	6	19	33	18

أحسب "المدى، الانحراف المعياري" لهذا التوزيع.

2) البيانات التالية تمثل وقت اشتغال جهاز طبي بالساعات بين عطل وعطل آخر خلال 3 أشهر،
30 32 40 80 75 60 50 55 35 35 47 85 87 95 90 92 88 90 92 91 77 31
33 40 70 82 81 97 91 92 34،
أوجد: "المدى، الانحراف المعياري" لهذه البيانات.

3) الجدول التالي لعدد الأطباء بعدد من المستشفيات،

ع.المستشفيات	10-	20-	30-	40-	50-	70-60
ع.الأطباء	3	4	6	19	33	18

أوجد: "المدى، الانحراف المعياري، التباين" لهذه التوزيع.

4) أوجد المدى والانحراف المعياري للقيم: 9 10 12 15 13 19.

5) حول القيم التالية لقيم معيارية: 6 2 8 7 5.

6) إذا كان لدينا القيم المعيارية: 0.19، 1.75- ، 0.68، 0.29- ، أثبت أن الوسط الحسابي والانحراف المعياري لتلك القيم المعيارية أعلاه على التوالي هو: صفر، وواحد.

الفصل الثاني

الارتباط والانحدار الخطي البسيط

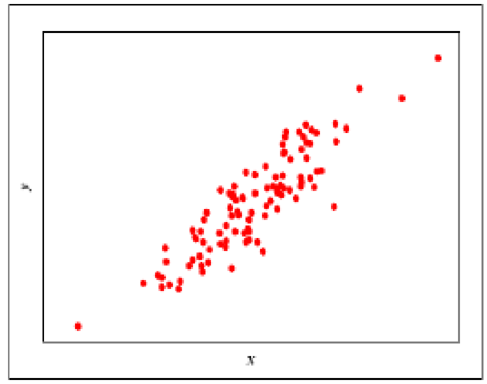
Correlation & Simple Linear Regression

أولا: الارتباط

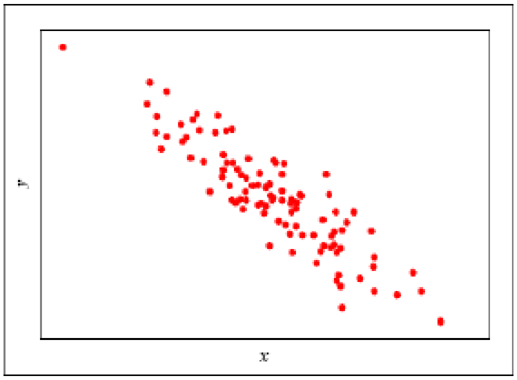
- الارتباط هو تعيين طبيعة وقوة العلاقة بين متغيرين.
- معامل الارتباط Correlation Coefficient هو مؤشر هذه العلاقة.
- اذا كان لدينا متغيران، المتغير X وهو متغير يتم تحديده من قبل الباحث أو الشخص ويسمى بالمتغير المستقل Independent variable .
- يرافق المتغيرX المتغير Y ويسمى بالمتغير التابع dependent variable وهو متغير تابع لأن نتيجته غير محددة وتعتمد على قيم المتغير المستقل.

أنواع الارتباط :

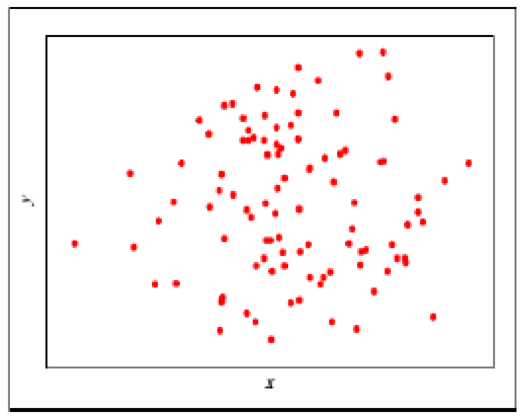
- الارتباط الموجب (الطردي) Positive Ccorrelation : اذا تغير أحد المتغيرين X أو Y فإن المتغير الآخر يتبعه بالاتجاه نفسه.
- الارتباط السالب (العكسي) Negative Correlation : إذا تغير أحد المتغيرين X أو Y فإن المتغير الآخر يتبعه بالاتجاه المضاد.



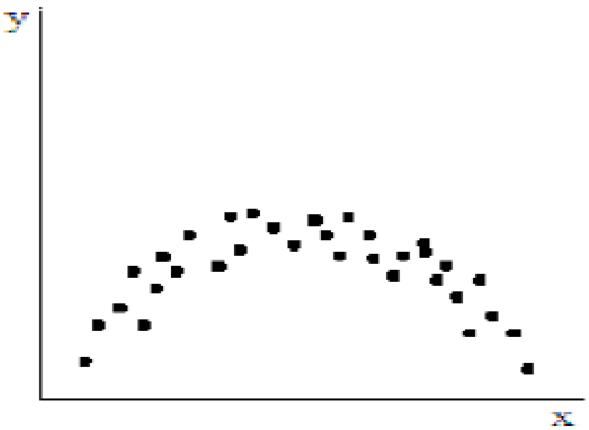
شكل الانتشار الخاص بالارتباط
الموجب (الطردي)



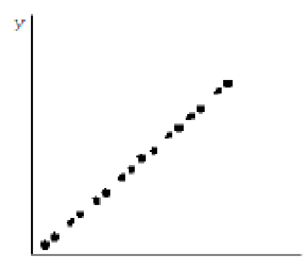
شكل الانتشار الخاص بالارتباط السالب
(العكسي)



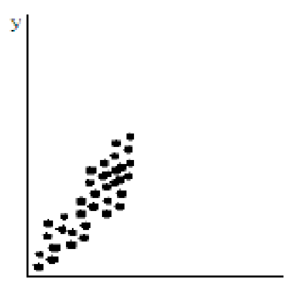
شكل الانتشار الخاص باستقلال
متغيرين (ظاهرتين)



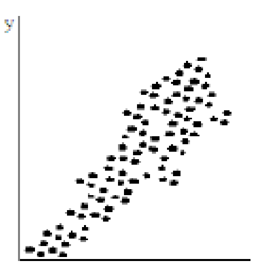
شكل الانتشار الخاص بالعلاقه الغير خطيه
بين متغيرين (ظاهرتين)



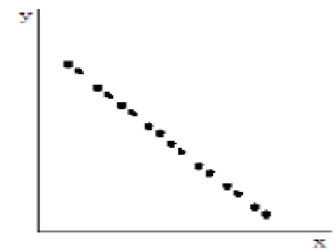
ارتباط طردي تام



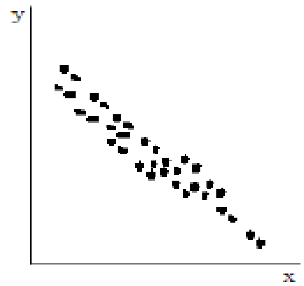
ارتباط طردي قوي



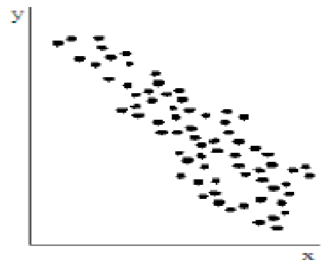
ارتباط طردي



ارتباط عكسي تام



ارتباط عكسي قوي



ارتباط عكسي

قياس الارتباط

- تستخدم معاملات الارتباط لقياس درجة الارتباط بين متغيرين(ظاهرتين).

معامل الارتباط :

- ويعرف معامل الارتباط والذي يرمز له بالرمز r عبارة عن مقياس رقمي يقيس قوة الارتباط بين متغيرين ، حيث تتراوح قيمته بين $(-1 \le r \le +1)$
- وتدل إشارة المعامل الموجبة على العلاقة الطردية ، بينما تدل إشارة المعامل السالبة على العلاقة العكسية .

- الجدول التالي يوضح أنواع الارتباط واتجاه العلاقة وشكل الانتشار لكل نوع

قيمة معامل الارتباط	المعنى
+1	ارتباط طردي تام
من 0.70 إلى 0.99	ارتباط طردي قوي
من 0.50 إلى 0.69	ارتباط طردي متوسط
من 0.01 إلى 0.49	ارتباط طردي ضعيف
0	لا يوجد ارتباط

- وما قيل عن الارتباط الطردي ينطبق على الارتباط العكسي مع عكس الإشارة

- معامل بيرسون للارتباط الخطي من أكثر معاملات الارتباط استخدامًا خاصة في العلوم الإنسانية والاجتماعية .

- ومستوى القياس المطلوب عند تطبيق معامل بيرسون للارتباط هو أن يكون كلا المتغيرين مقياس فترة أو نسبي أو بمعنى آخر أن تكون بيانات كلا المتغيرين (الظاهرتين) بيانات كمية .

- حساب معامل بيرسون للارتباط الخطي :

- يمكن حساب معامل بيرسون بدلالة القراءات لبيانات المتغيرين باستخدام الصيغة التالية:

$$r_p = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

$\sum_{i=1}^{n} x_i y_i$: مجموع حاصل ضرب x في y

$\sum x$: مجموع قيم المتغير x

$\sum y$: مجموع قيم المتغير y

$\sum x^2$: مجموع مربعات قيم المتغير x

$\sum y^2$: مجموع مربعات قيم المتغير y

- مثال 1 : سجلت ست قراءات تقريبية لحجم الإنتاج وحجم صادرات دواء معين بإحدى الدول خلال عدة سنوات كما يلي:

حجم الصادرات (y)	2	2	2	1	1	1
حجم الإنتاج (x)	3	4	2	2	2	2

	x	y	xy	x^2	y^2
	3	2	6	9	4
	4	2	8	16	4
	2	2	4	4	4
	2	1	2	4	1
	2	1	2	4	1
	2	1	2	4	1
$\sum$	15 $=\sum x$	9 $=\sum y$	24 $=\sum xy$	41 $=\sum x^2$	15 $=\sum y^2$

$$r_p = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

$$r_p = \frac{6(24) - (15)(9)}{\sqrt{((6\times 41) - 15^2)((6\times 15) - 9^2)}} =$$

$$\frac{144 - 135}{\sqrt{(246 - 225)(90 - 81)}} = \frac{9}{\sqrt{189}} = \frac{9}{13.75} = 0.65$$

- مثال 2 البيانات التالية تمثل المتغيرين X,Y

Y	1	3	8	7	6	5	7	8	12	12
X	9	11	17	18	19	16	16	19	23	23

احسب معامل الارتباط الخطي، ما مدى قوة العلاقة الخطية؟

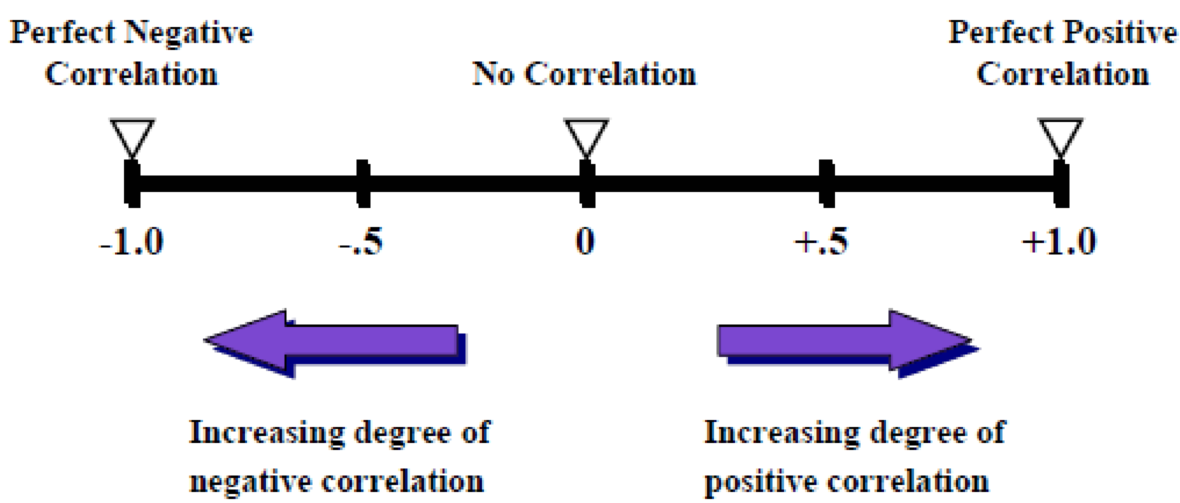
	x	y	xy	x^2	y^2
	9	1	9	81	1
	11	3	33	121	9
	17	8	136	289	64
	18	7	126	324	49
	19	6	114	361	36
	16	5	80	256	25
	16	7	112	256	49
	19	8	152	361	64
	23	12	276	529	144
	23	12	276	529	144
$\sum$	171	69	1314	3107	585
	$=\sum x$	$=\sum y$	$=\sum xy$	$=\sum x^2$	$=\sum y^2$

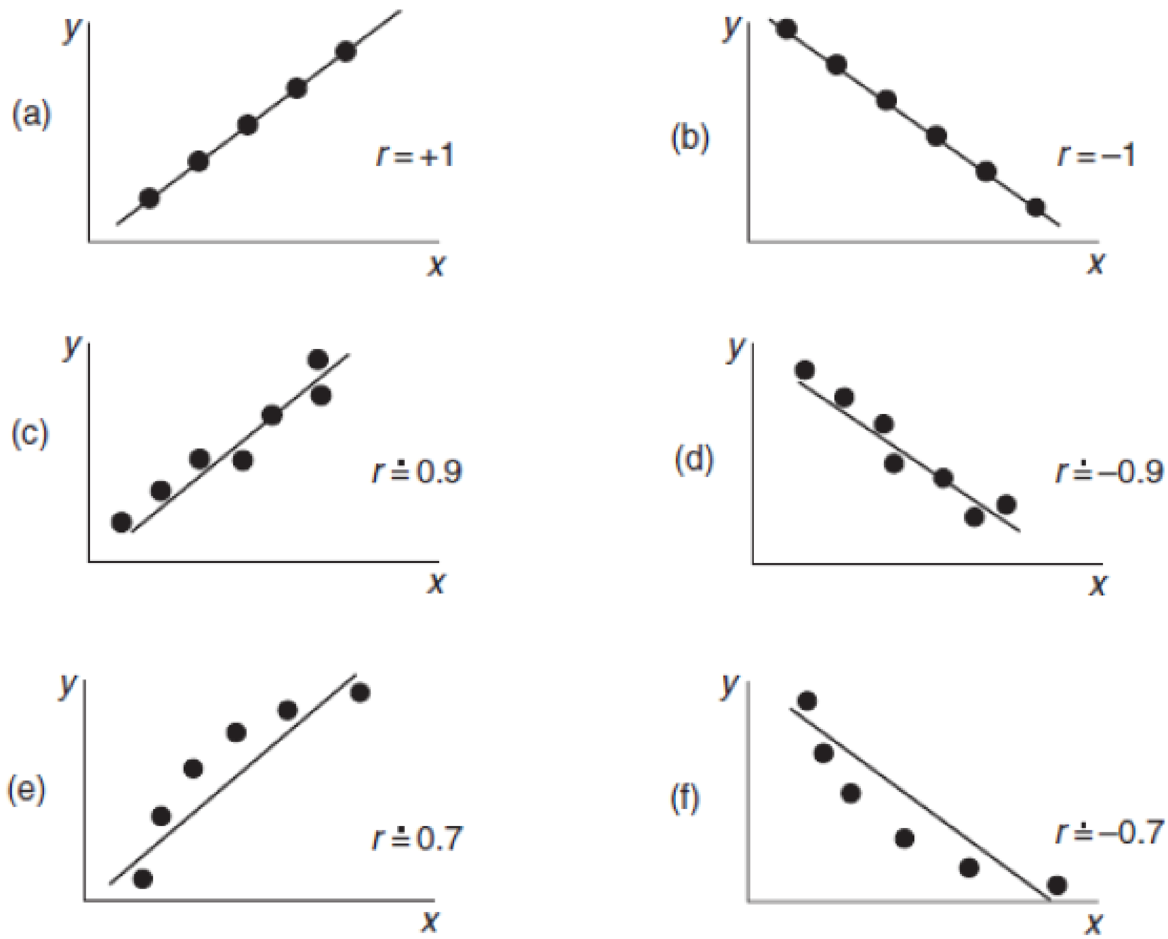
$$r_p = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

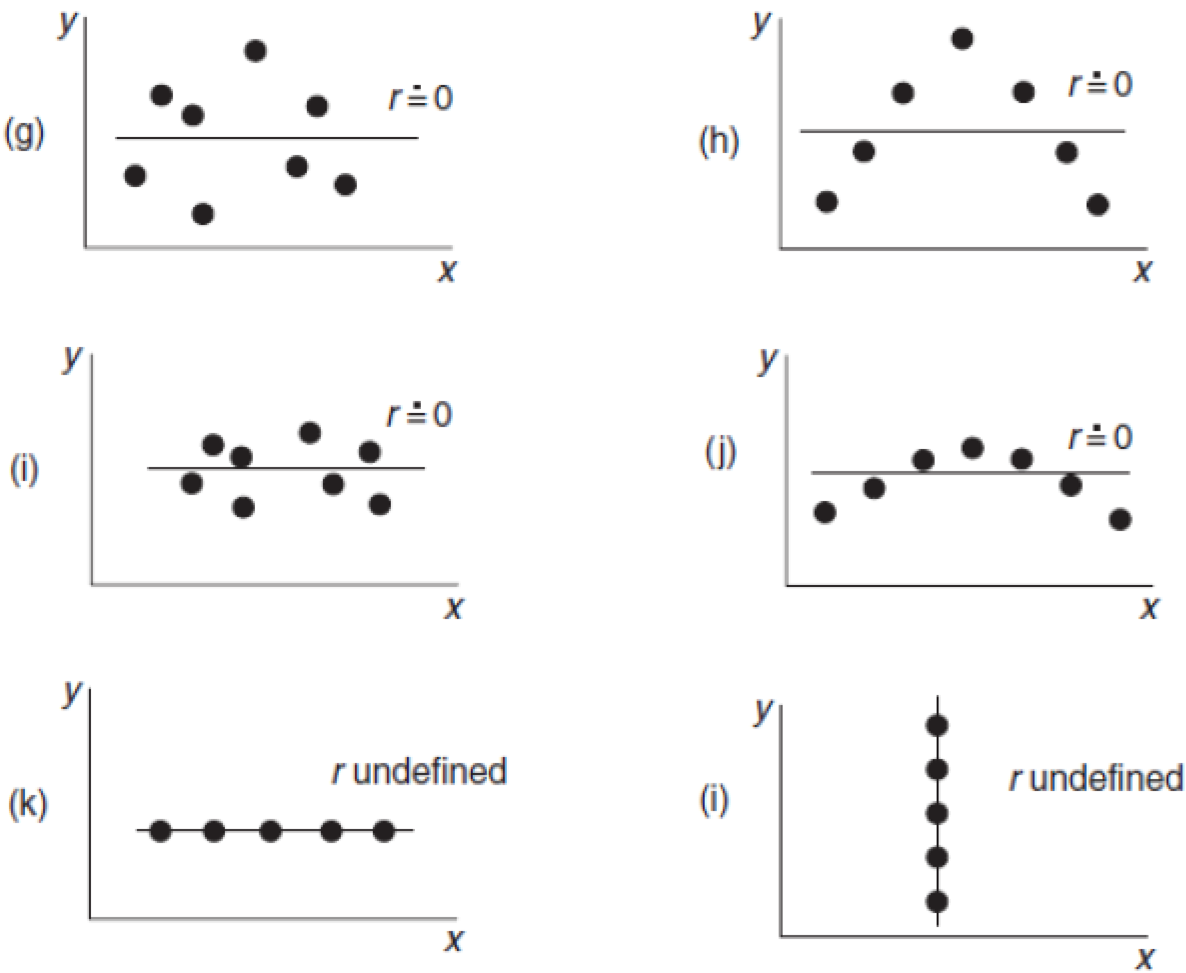
$$r_p = \frac{10(1314) - (171)(69)}{\sqrt{((10\times 3107) - 171^2)((10\times 585) - 69^2)}} =$$

$$\frac{13140 - 11799}{\sqrt{(31070 - 29241)(5850 - 4761)}} = \frac{1341}{\sqrt{18291089}} = \frac{1341}{1411.30} = 0.95$$

<u>علاقة ارتباط طردية قوية</u>




(a)
r = +1
(b)
r = −1
(c)
r ≐ 0.9
(d)
r ≐ −0.9
(e)
r ≐ 0.7
(f)
r ≐ −0.7
y
x


(g)
r ≐ 0
(h)
r ≐ 0
(i)
r ≐ 0
(j)
r ≐ 0
(k)
r undefined
(i)
r undefined
y
x

الانحدار الخطي البسيط

Simple Linear Regression

- الانحدار هو أسلوب يمكن بواسطته تقدير قيمة أحد المتغيرين بمعلومية قيمة المتغير الآخر عن طريق معادلة الانحدار
- تهدف دراسة الانحدار التنبؤ بقيمة متغير (Y) بمعرفة متغير آخر (X)
- ويعرف المتغير الأول بالمتغير التابع (dependent) ويرمز له Y في حين يعرف المتغير الآخر بالمتغير المستقل (Independent) ويرمز لهX
- لذا المتغير X عرف بالمتغير المستقل في حين Y تتعين قيمتها تبعاً لقيمة X لذا عرفت Yبالمتغير التابع) أي تبعاً لقيمةX
- كما أن الانحدار هنا بسيط لوجود متغيرين فقط تابع ومستقل، وعند ذكر كلمة الخط نعني بها خط الانحدار.
- الغرض من استخدام أسلوب تحليل الانحدار الخطي البسيط، هو دراسة وتحليل أثر متغير كمي على متغير كمي آخر، ومن الأمثلة النموذجية على تحليل الانحدار:
- اعتماد ضغط الدم Yعلى عمر الشخص X،
- أو اعتماد الوزن لحيوانات التجربة Yعلى معدل التغذية اليومي X هذا الارتباط والتابعية بين X و Y هي ما ندعوه بالانحدار أو الارتباط فنقول ارتباط Y ب X
- دراسة أثر كمية السماد على إنتاجية الدونم.
- دراسة أثر كمية البروتين التي يتناولها الأبقار على الزيادة في الوزن.
- ويلاحظ من ذلك أن نموذج الانحدار يعتمد دائماً على علاقة السببية بمعنى ان يكون التغير في المتغير المستقل مسبب رئيسي للتغير في المتغير التابع.

• معادلة الخط المستقيم

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

حيث a : ثابت الانحدار أو الجزء المقطوع من محور y

b : ميل الخط المستقيم أو معامل انحدار Y على X (أو Y/X)

• وتحسب القيمتان a و b من العلاقتين التاليتين :

حيث:

$$a = \frac{\sum y - b\sum x}{n} \qquad b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2} =$$

لإيجاد قيمة مقدرة جديدة $\hat{y}_h$ نعوض بقيمة معلومة للمتغير المستقل ولتكن x_h في معادلة تقدير خط الانحدار Y/X

نعوض

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

لدراسة علاقة الاستهلاك المحلي (y)بالإنتاج (x)لمادة الإسفلت (بالمليون برميل) خلال عدة سنوات، أخذنا عشر قراءات تقريبية كما يلي :

y	6	8	9	8	7	6	5	6	5	5
x	10	13	15	14	9	7	6	6	5	5

أوجد معادلة الانحدار الخطي البسيط، وتوقع قيمة الاستهلاك عندما يصل إنتاج 16,000,000برميل .

- الحل :

	x	y	xy	x^2
	10	6	60	100
	13	8	104	169
	15	9	135	225
	14	8	112	196
	9	7	63	81
	7	6	42	49
	6	5	30	36
	6	6	36	36
	5	5	25	25
	5	5	25	25
$\sum$	90	65	632	942
	$=\sum x$	$=\sum y$	$=\sum xy$	$=\sum x^2$

$$b=\frac{n\sum xy-(\sum x)(\sum y)}{n\sum x^2-(\sum x)^2}=\frac{6320-(90)(65)}{9420-90^2}=\frac{6320-5850}{9420-8100}=\frac{470}{1320}=0.36$$

$$a=\frac{\sum y-b\sum x}{n}=\frac{65-(0.36\times 90)}{10}=3.26$$

∴معادلة خط الانحدار البسيط في هذه الحالة :

$$\hat{y}=3.26+0.36x$$

- ولتوقع قيمة الاستهلاك المحلي عندما يصل الإنتاج 16000000برميل، نحول وحدة هذه القيمة من برميل إلى مليون برميل بالقسمة على مليون أي أن القيمة المستخدمة في توقع الاستهلاك هي $x_h=16$
- وبالتعويض في المعادلة السابقة نجد أن:

$$\hat{y}_h=a+bx_h$$
$$=3.26+0.36(16)=9.02$$

أي أن الاستهلاك قد يصل إلى **9.02** مليون برميل، أي ما يعادل **9020000** برميل خلال السنة.

الفصل الثالث

1-2: مبادئ الاحتمالات: "Basics of Probabilities"،

تعد مبادئ الاحتمالات من المحاور الأساسية في الرياضيات، سنتطرق خلال هذا البند لبعض المفاهيم منها ما يلاءم أولي الاهتمامات الصحية، ذلك أنه كثيرا ما نسمع عن نسبة معينة لاحتمال تعافي مريض أو نجاح عملية جراحية...، لذا فإن احتمال أي حادثة ما يمثل النسبة بين عدد النتائج التي تخص الحادثة إلى جميع النتائج الممكنة الوقوع...، أي أن:

$0 \le P(A) \le 1,\ P(A) = \frac{n(A)}{N}$، حيث $P(A)$ يرمز لاحتمال وقوع الحادثة A ، وكذلك $n(A)$ يرمز للنتائج المتعلقة بالحادثة A ، أيضا N يرمز لجميع النتائج الممكنة الوقوع (فراغ العينة)،

هذا وعندما $P(A) = 0$ معنى ذلك أن A حادث مستحيل، وعندما $P(A) = 1$ يعني ذلك أن A حادث مؤكد، وباعتبار أن p يمثل النجاح، وأن q يمثل احتمال الفشل، لذا سيكون $p + q = 1$ ،

مثال1: فيما يلي توزيع أصناف الدم لساكني أحدى المحافظات:

الصنف	A	B	AB	O	المجموع
العدد	199	120	34	151	454

فإذا اختير شخص عشوائيا من تلك المحافظة، أحسب الاحتمالات: $P(A),\ P(B),\ P(AB),\ P(O)$ ،

لدينا: $P(A) = \frac{n(A)}{N} = \frac{199}{454} = 0.328$ ، كما سيكون:

$P(B)=\frac{120}{454}=0.264,\ P(AB)=\frac{34}{454}=0.075,\ P(O)=\frac{151}{454}=0.333$ ،

وعلاوة على ما سبق فإن الدارس لهذا المقرر سيواجه بعض المصطلحات المندرجة تحت موضوع "نظرية المجموعات"، لذا يستلزم استعادة ما يلزم من ذلك مثل معرفة مدلول مصطلح "مجموعة" مع بعض الأساسيات كالمجموعة الخالية والشاملة والجزئية والاتحاد والتقاطع والمكملة...الخ،

إضافة إلى أنه من المناسب استعادة الأساسيات المتعلقة بموضوع العد وما يتعلق بدلك من مفاهيم التباديل والتوافيق...، ذلك أن الدارس سيواجه خلال دراسته لهذا المقرر الكثير من المسائل والتطبيقات التي تستلزم تمكنه من تلك المفاهيم الأساسية...،

مثال2: الجدول التالي يمثل تكرارات أصناف الدم" X " لبعض من ذوي أمراض معينة" Y ":

$Y \quad X$	A	B	AB	O	المجموع
قرحة معدة C_1	47	21	4	20	92
قرحة معوية C_2	13	7	6	28	54
التهاب كبد C_3	70	44	9	62	185
التهاب الزائدة C_4	31	2	4	32	91
المجموع	161	96	23	142	422

فإذا اختير شخص عشوائيا، أوجد ما يلي:

أ) احتمال أن يكون الشخص المختار لديه قرحة معدة،

ب) احتمال أن يكون الشخص المختار إما يحمل صنف الدم B أو لديه التهاب كبد،

ج) احتمال أن يكون الشخص المختار يحمل الصنف AB ولديه قرحة معوية،

د) احتمال أن يكون الشخص المختار غير مصاب بقرحة معدة،

هـ) احتمال أن يكون الشخص المختار يحمل صنف الدم A أو O ،

و) احتمال أن يكون الشخص المختار مصاب بالتهاب الكبد وغير مصاب بالتهاب الزائدة،

الحل: أ) $P(C_1)=\frac{n(C_1)}{N}=\frac{92}{422}=0.218$ ، ب) $P(B\cup C_3)=\frac{n(B\cup C_3)}{N}=\frac{96+185-44}{422}=0.561$ ،

جـ) $P(AB\cap C_2)=\frac{n(AB\cap C_2)}{N}=\frac{6}{422}=0.014$ ، د) $P(C_1^{-})=\frac{n(C_1^{-})}{N}=\frac{422-92}{422}=0.782$ ،

هـ) $P(A\cup O)=\frac{n(A\cup O)}{N}=\frac{161+142}{422}=0.72$ ، و) $P(C_3\cap C_4^{-})=\frac{n(C_3\cap C_4^{-})}{N}=\frac{185}{422}=0.42$ ،

ملاحظة1: في فقرتي أ، د) لدينا C_1^{-} يمثل المكملة لـ C_1، ذلك أن $C_1\cup C_1^{-}=422$، أيضا $P(C_1)+P(C_1^{-})=1$ ،

كما أن من المفاهيم الهامة في موضوع الاحتمالات ما يلي:

- **الحوادث المتنافية:** وتعرف بأن وقوع أحدهما يمنع وقوع الآخر،

- قانون الجمع: لأي حادثين متنافيين A, B ، سيكون: $P(A\cup B)=P(A)+P(B)$ ، "أي أن $A\cap B=\phi$ "،

ملاحظة2: بالعودة للفقرات ب، هـ) في المثال السابق، الحالة **ب)** ولأن "الحادثين غير متنافيين" سيكون:

$P(B\cup C_3)=P(B)+P(C_3)-P(B\cap C_3)=\frac{96}{422}+\frac{185}{422}-\frac{44}{422}=0.561$ ،

أما في **هـ)** ولأن "الحادثين متنافيين" نستخدم قانون الجمع: $P(A\cup O)=P(A)+P(O)=\frac{161}{422}+\frac{142}{422}=0.72$ ،

- **الحوادث الشرطية:** حيث تعرف بأن وقوع أحدها يؤثر على الأخر، ولأي حادثين معتمدين A, B سيكون:

$P(A|B)=\frac{P(A\cap B)}{P(B)}$ ، "حيث $P(A|B)$ تقرأ احتمال وقوع A بشرط وقوع B"،

- **الحوادث المستقلة:** وتعرف بأن وقوع أحدهما لا يمنع وقوع الآخر، أي أن $P(A|B)=\frac{P(A\cap B)}{P(B)}=P(A)$ ،

- قانون الضرب: لأي حادثين مستقلين A, B ، لدينا: $P(A\cap B)=P(A).P(B)$ ، ذلك أنه من شرط الاستقلال لدينا: $P(B|A)=P(B)$ ،

ولأن الحوادث في المثال السابق غير مستقلة، فعلى سبيل المثال في الفقرة **ج)** سنجد أن:

$P(AB\cap C_2)=P(AB).P(C_2|AB)=\frac{23}{422}\times\frac{6}{23}=0.014$ ،

بديهيات وملاحظات: لأي حدثين A, B ، لدينا:

1) $P(A)=P(A-B)=P(A)-P(A\cap B)=P(A\cap B^{C})$ ، أيضا $P(B)=P(B-A)=P(B)-P(A\cap B)=P(A^{C}\cap B)$ ،

2) هناك ما يعرف بقانون "بيز" والذي يعتبر تعميما لمبدأ الاحتمال الشرطي،

ذلك أنه لأي حوادث متنافية $A_1, A_2, ..., A_n$ ، ولأي حادث مثل B سيكون:

$$P(A_k|B)=\frac{P(A_k\cap B)}{P(B)}=\frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n}P(B|A_i)P(A_i)}$$ ،

3) $P(B|A)=\frac{P(A|B)\,P(B)}{P(A|B)\,P(B)+P(A|B^{-})P(B^{-})}$, $P(A|B)=\frac{P(B|A)\,P(A)}{P(B|A)\,P(A)+P(B|A^{-})P(A^{-})}$ ،

4) $P(A^{-})=1-P(A),\quad P(A^{-}|B)=1-P(A|B),\quad P(A|B^{-})=1-P(A|B)^{\neg},$ ،

مثال3: في اختبار تشخيص مرض السرطان، إذا كان A يمثل شخص أجري عليه الاختبار وتبين أنه مصاب، وليكن B يمثل بأن الاختبار يشير إلى أن الشخص مصاب،

فإذا علم أن: $P(A\,|\,B)^{\neg}=0.95,\quad P(A\,|\,B)=0.95,\quad P(B)=0.005$ ،

ما احتمال أن A حقيقة مصاب والاختبار يعطي نفس الأعراض,

الحل: لدينا $P(A|B^{-})=1-P(A|B)^{\neg}=1-0.95=0.05$, $P(B^{-})=1-P(B)=1-0.005=0.995$ ،

لذا سيكون: $P(B|A)=\frac{P(A|B)\,P(B)}{P(A|B)\,P(B)+P(A|B^{-})P(B^{-})}=\frac{(0.95)(0.005)}{(0.95)(0.005)+(0.05)(0.995)}=0.087$ ،

نشاط/ 1) في عينة من 400 شخص، موزعين حسب فصيلة الدم:

الصنف	A	B	AB	O	المجموع
العدد	180	40	20	160	400

فإذا لكل صنف من نوعين "- ,+"، كل منهما يمثل 85% ,15% على الترتيب من بين أفراد العينة، والمعلوم أن أخذ شخص للدم يرتبط بالصنف حسب الجدول المقابل:

الصنف	A	B	AB	O
معطي	AB	A, AB	B, AB	للكل
آخذ	من	A, O	B, O	O

وبفرض أن "+" يأخذ من "- ,+"، وأن "-" يأخذ فقط من "-"،

اوجد ما يلي:

أ) احتمال أن الشخص يحمل الصنف B ، ب) احتمال أن الشخص يحمل الصنف B^{+} ،

ج) احتمال أن الشخص يحمل الصنف O ، د) احتمال أن الشخص يحمل الصنف O^{-} ،

هـ) احتمال أن الشخص الذي يحمل الصنف A^{-} يمكن أن يأخذ الدم من شخص ما،

و) احتمال أن الشخص الذي يحمل الصنف AB^+ يمكن أن يأخذ الدم من شخص ما،

2) بفرض احتمال إصابة شخص في مجتمع معين بمرض M هو 0,25، أعطي المجتمع لقاح T، فكانت نسبة استجابة غير المصاب 0.8، ما احتمال أن شخص لقح وأعطي نتيجة إيجابية أنه مصاب بالمرض.

3) البيانات التالية تمثل مستويات العدوى"X" في إحدى المناطق حسب الجنس"Y":

Y \ X	L_1	L_2	L_3	L_4	المجموع
ذكر	334	280	91	386	1091
أنثى	105	101	12	109	327
المجموع	439	381	103	495	1418

احسب الاحتمالات التالية:

أ) أن درجة العدوى لشخص ما في المجتمع هي L_1 ، ب) أن شخص من المصابين بالعدوى هو ذكر،

ج) أن شخص من المصابين بالعدوى ليس أنثى، د) أن درجة العدوى لشخص L_2 وأنه أنثى،

ه) إما أن درجة العدوى لشخص L_3 أو أنه ذكر، و) أن أحد الأشخاص ليس مصاب بعدوى L_4 ،

4) كم عدد الحالات التي يمكن بها ترتيب ثلاثة إعلانات يراد وضعها في لوحة مستشفى ما,

5) أراد باحث اختبار ثلاثة عقاقير ولديه حيوانات تكفي لاختبار ثلاثة عقاقير فقط، كم عدد مجاميع العقاقير التي يمكن اختبارها.

2-2: بعض التوزيعات النظرية: "Some Theoretical Distributions"،

يعتمد الإحصاء في تفسير وتحليل البيانات المعطى على قاعدة معينة، كأن يسند البيانات إلى توزيع نظري معروف، فإذا كان اختبار ذلك التوزيع مناسب كانت النتائج دقيقة...،

من تلك التوزيعات ما يسمى بالمتقطع (discrete) ومنها ما يسمى بالمتصل "المستمر" (continuous)، حيث سنتطرق لأشهر نوعين من التوزيعات المتقطعة وهما "توزيع ثنائي الحدين وتوزيع بواسون Binomial & Poisson distribution"،

كما سنتطرق لواحد من أهم التوزيعات المتصلة وهو "التوزيع الطبيعي Normal distribution"،

أولاً: التوزيعات النظرية المتقطعة،

- توزيع ثنائي الحدين "Binomial distribution":

وهو مشتق مما يعرف بعملية برنولي "Bernoulli" التي تجعل من أي مشاهدة الحالتين "نجاح أو فشل، ذكر أو أنثى، حي أو ميت،..."، وعليه فإنه لأي حادثة من n محاولة وباعتبار "P" هو احتمال النجاح، وأن "q=p-1" هو احتمال الفشل، وليكن X كمتغير عشوائي "دالة حقيقية معرفة على شكل حوادث تمثل مجموعة جزئية من فضاء العينة"،

كما أنه من المعروف أن عدد الطرق للحصول على x من النجاحات في n من المحاولات هو $\binom{n}{x}$،

لذا فإن احتمال ظهور X في حالات النجاح من المحاولات n، هو ما يسمى بتوزيع ثنائي الحدين والذي يعطى بالقاعدة: $\boldsymbol{f(x)=\binom{n}{x}p^{x}q^{n-x},\ x=0,1,2,\ldots,n}$ ،

كما أ ن الوسط الحسابي لهذا التوزيع هو $\boldsymbol{np}$ ، وتباينه هو $\boldsymbol{npq}$ ،

خصائص التوزيع:

- تجرى هذه التجربة لعدد محدد من المحاولات n كل محاوله مستقلة عن الأخرى،
- كل محاولة تسفر عن نتيجتين أحدهما فشل "p" والأخرى نجاح "q=p-1"،

مثال4: تشير الدراسات إلى أن 40% من الأطفال الذين يصابون بمرض معين إما يموتون أو يصابوا بعاهة، فإذا أصيب 3 أطفال بهذا المرض، أحسب الاحتمالات:

أ) أن الجميع يشفون من المرض، **ب)** أن يشفى شخصين من المرض،

جـ) أن يشفى واحد فقط من المرض، **د)** أن لا يشفى أحد من المرض،

الحل: بداية لدينا $\boldsymbol{p} = 1 - 0.40 = 0.60,\ \boldsymbol{q} = 0.40$ ،

وعليه سيكون:

أ) $\boldsymbol{f}(3) = \binom{3}{3}(0.60)^3(0.40)^0 = 0.216$ ،

ب) $\boldsymbol{f}(2) = \binom{3}{2}(0.60)^2(0.40)^1 = 0.432$ ،

جـ) $\boldsymbol{f}(1) = \binom{3}{1}(0.60)^1(0.40)^2 = 0.192$ ،

د) $\boldsymbol{f}(0) = \binom{3}{0}(0.60)^0(0.40)^3 = 0.064$ ،

مثال5: ليكن 24% من مجتمع بصنف الدم $\boldsymbol{B}$ ، سحب 20 شخص عشوائيا، أوجد الاحتمالات:

أ) وجود 3 أشخاص يحملون صنف الدم $\boldsymbol{B}$ ،

ب) وجود 3 أشخاص أو أكثر يحملون صنف الدم $\boldsymbol{B}$ ،

جـ) وجود أقل من 3 أشخاص يحملون صنف الدم $\boldsymbol{B}$ ،

الحل: أ) $f(3) = \binom{20}{3}(0.24)^3(0.76)^{17} = 0.148$ ،

ب) لدينا $P(x \geq 3) = 1 - P(x < 3) = 1 - P(x \leq 2) = 1 - \{f(0) + f(1) + f(2)\}$ ،

ومنه سيكون:

$$P(x \geq 3) = 1 - \left\{\binom{20}{0}(0.24)^0(0.76)^{20} + \binom{20}{1}(0.24)^1(0.76)^{19} + \binom{20}{2}(0.24)^2(0.76)^{18}\right\} = 0.89$$ ،

جـ) لدينا: $P(x \leq 2) = f(0) + f(1) + f(2)$ ،

ومنه سيكون:

$$P(x \leq 2) = \left\{\binom{20}{0}(0.24)^0(0.76)^{20} + \binom{20}{1}(0.24)^1(0.76)^{19} + \binom{20}{2}(0.24)^2(0.76)^{18}\right\} = 0.1085$$ ،

- توزيع بواسون "Poisson distribution":

حيث أنه لأي متغير عشوائي X يتبع توزيع ثنائي الحدين بمعلمتين p, n يمكن توزيعه بمعلمة λ عندما تكون n كبيرة "بحيث تسعى إلى مالا نهاية"، وذلك وفقا لبواسون حسب القاعدة:

$$f(x)=\frac{e^{-\lambda}\lambda^{x}}{x!},\quad x=0,\ 1,\ 2,...$$ ،

حيث أن e=2.718 "الأساس للوغاريتم الطبيعي"، كما أن الوسط متساويين ولهما القيمة λ والتي تمثل معدل عدد المشاهدات،

شروط حدوث تجربة بواسون:

- تعدُ النجاحات في وحدة الزمن أو المساحة أو الحجم...،
- أن يحدث كل عنصر من عناصر التجربة عشوائياً،
- أن يحدث كل عنصر من عناصر التجربة بصورة مستقلة عن العناصر الأخرى،

أيضا من

مثال6: ليكن احتمال أن لدى شخص حساسية من حقن لقاح ما هو 0.001، فإذا سحب عشوائيا عينة من 2000 شخص، ما احتمال أن بين هؤلاء الأشخاص 3 لديهم نفس الظاهرة،

الحل: لدينا $p=0.001,\ n=2000,\ \lambda=np=2$ ،

أي أن: $f(3)=\frac{e^{-2}2^{3}}{3!}=0.18$ ،

مثال7: أظهرت الدراسات أن عدد المراجعات الطارئة لسنوات بأحد المستشفيات كانت 3 مراجعات يومية وتتبع توزيع بواسون،

أحسب الاحتمالات:

ا) حدوث مراجعتين في يوم، ب) عدم حدوث أي مراجعة ليوم ما، ج) حدوث 3 أو 4 مراجعات ليوم ما،

الحل: لدينا $\lambda=3$، أي أن:

أ) $f(2)=\frac{e^{-3}3^{2}}{2!}=0.225$ ، ب) $f(0)=\frac{e^{-3}3^{0}}{0!}=0.05$ ،

ج) $f(3)+f(4)=\frac{e^{-3}3^{3}}{3!}+\frac{e^{-3}3^{4}}{4!}=0.225+0.169=0.39$ ،

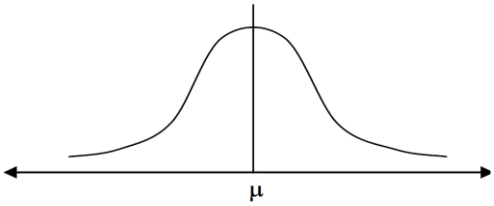
- التوزيع الطبيعي "Normal distribution":

ويسمى أيضا بتوزيع كاوس "Gauss"، كما أن هذا التوزيع ذا أهمية في كثير من الجوانب حيث أن معظم الظواهر تتوزع توزيعا طبيعيا، والذي يعطى وفقا للقاعدة $-\infty < x < \infty \;,\; f(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ ، حيث أن μ تمثل الوسط الحسابي للتوزيع، $\pi = \frac{22}{7}$ النسبة الثابتة، ومعلمتا التوزيع هما μ، σ،

ونتيجة لوجود قيم مختلفة لتلك المعلمات لذا تتعدد التوزيعات الطبيعية، كما أن القيم التي تتمثل بالمنحنى الطبيعي تتحول إلى قيم معيارية "Standard score" لنحصل على توزيع طبيعي معياري وسطه الحسابي 0 وانحرافه 1، أي أن:

$-\infty < z < \infty \;,\; f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2} z^2}$ ، بدرجة معيارية $z = \frac{x-\mu}{\sigma}$ تمثل بعد القيمة عن المتوسط بوحدات الانحراف المعياري، وهناك جداول خاصة بهذا التوزيع، وفيما يلي نتعرض للخواص الرياضية للتوزيع،

- معادلة التوزيع الطبيعي " $f(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ " ترسم منحنى "يعرف بالمنحنى الطبيعي" وهو على شكل الجرس حيث " $f(x) \to 0 \quad \text{as} \quad x \to \pm\infty$ "، كما بالشكل المقابل:

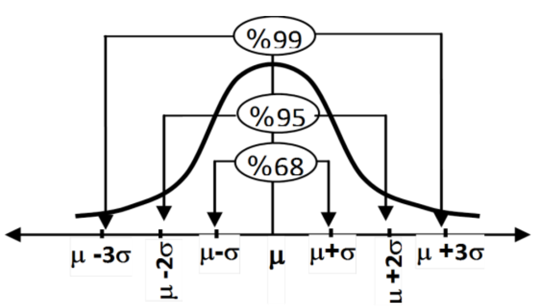


ويتبين تناظر المنحنى حول ما يسمى بمحور التماثل $x = \mu$ ،

كما أن للمنحنى قمة واحدة تحدث عند $x = \mu$ ،

حيث تمثل تلك المعادلة دالة الكثافة، وتتحدد تلك المعادة مع بيانها المتمثل بالمنحنى الطبيعي حالما تعرف قيم معينه لكل من μ, σ ،

- إضافة إلى أن كل من الوسط الحسابي والوسيط والمنوال في التوزيع الطبيعي متساوية، كما أن المساحة تحت المنحى الطبيعي تساوي 1، أي أن المساحة الواقعة إلى يمين محور التماثل وكذلك الواقعة إلى يساره كل منها يساوي 0.5،



حيث يعطي الشكل المقابل النسبة التي تمثله القيم المرافقة.

- ومن خلال عملية المعايرة"Standardization"يمكن لنا إيجاد مساحات تحت المنحنى غير المعياري، حيث يستعان بالجداول التي وضعت لهذا التوزيع ذلك لصعوبة العمليات الرياضية في إيجاد تلك المساحات، وقد أوردنا الجدول المطلوب في نهاية هذا المؤلف كي يعود إليه الدارس لإيجاد المساحات "قيمة الاحتمال"، فلأي متغير طبيعي X، سيكون:

$$\begin{aligned} P(x_1 \le X \le x_2) &= P(\frac{x_1-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{x_2-\mu}{\sigma}) \\ &= P(z_1 \le Z \le z_2) \\ &= P(Z \le z_2) - P(Z \le z_1) \\ &= \text{"the area on left of } z_2 - \text{the area on left of } z_1\text{"} \end{aligned}$$

حيث: $Z = \frac{X-\mu}{\sigma},\ z_2 = \frac{x_2-\mu}{\sigma},\ z_1 = \frac{x_1-\mu}{\sigma}$،

كما نود التنبه إلى ما يلي:

$$\begin{aligned} \text{"(area on left of } a) - \text{(area on left of } b)\text{"} &= P(b \le X \le a) \\ &= P(b < X \le a) \\ &= P(b \le X < a) \\ &= P(b < Z < a) = \text{"(area on left of } a) - \text{(area on left of } b)\text{"} \end{aligned}$$

ذلك لأن: $P(X=a) = P(X=b) = 0$،

ولتوضيح ذلك سنعطي الأمثلة التالية،

أمثلة8:

1- إذا كان شعور القائمين بالعلاج الطبيعي أن درجات اختبار الكفاءة تتوزع طبيعيا بمتوسط10 وانحراف معياري 2.5، فلو أعطي الاختبار لشخص اختير عشوائيا، ما احتمال حصوله في الأقل على 15 درجة،

الحل:

$$\begin{aligned} P(x \ge 15) &= P\left(\frac{x-\mu}{\sigma} \ge \frac{15-10}{2.5}\right) \\ &= P(z \ge 2) = 1 - P(z < 2) \\ &= 1 - 0.9772 = 0.0228 \end{aligned}$$ ،

2- بفرض أن أوزان مجموعة أشخاص تتوزع طبيعيا بمتوسط 140 باون وانحراف معياري 25 باون، ما احتمال وزن شخص اختير عشوائيا أن وزنه بين 100- 170 باون،

الحل:

$$P\ (100 \le x \le 170) = P\left(\frac{100-140}{25} \le \frac{x-\mu}{\sigma} \le \frac{170-140}{25}\right)$$

$$= P\ (-1.6 \le z \le 1.2)$$

$$= P\ (z \le 1.2) - P\ (z \le -1.6)$$

$$= P\ (z \le 1.2) - [1 - P\ (z \le 1.6)]$$

$$= 0.8849 - 1 + 0.9452 = 0.8301$$

3- إذا كان سعة قحف الجمجمة لمجتمع تتوزع طبيعيا بمتوسط 1400سم^3 وانحراف معياري 125سم^3، ما احتمال أن شخص اختير عشوائيا لديه قحفة جمجمة بسعة:

أ) أكبر من 1450سم^3، ب) أقل من 1350سم^3، جـ) مابين 1300سم^3- 1500سم^3،

الحل:

أ)

$$P(x > 1450) = P\left(\frac{x-\mu}{\sigma} > \frac{1450-1400}{125}\right)$$

$$= P(z > 0.4) = 1 - P(z < 0.4)$$

$$= 1 - 0.6554 = 0.3446$$

ب)

$$P(x < 1350) = P\left(\frac{x-\mu}{\sigma} < \frac{1350-1400}{125}\right)$$

$$= P(z < -0.4) = 1 - P(z < 0.4)$$

$$= 1 - 0.6554 = 0.3446$$

جـ)

$$P\ (1300 < x < 1500) = P\left(\frac{1300-1400}{125} < \frac{x-\mu}{\sigma} < \frac{1500-1400}{125}\right)$$

$$= P\ (-0.8 \le z \le 0.8)$$

$$= P\ (z < 0.8) - P\ (z < -0.8)$$

$$= P\ (z < 0.8) - [1 - P\ (z < 0.8)]$$

$$= 0.7881 - 1 + 0.7881 = 0.5762$$

نشاط/ 1) إذا كان قسم الطوارئ بمركز صحي معين يستوعب 6 أشخاص،

أ) إذا كان احتمال استقبال حالة واحدة هو 0.2، أحسب الاحتمالات:

- عدم استقبال أي حالة،
- استقبال حالة واحدة فقط،
- استقبال أكثر من حالة،

ب) لو علم أن الطبيب بالمركز يقوم بعملية واحدة فقط، وعند وجود حالتين يستدعى طبيب آخر، أحسب الاحتمالات:

- استدعاء الطبيب الإضافي مرة على الأقل في الأسبوع،
- استدعاء الطبيب الإضافي 120 مرة على الأقل في السنة "360 يوم"،

2) من خلال التجارب وجد أن من 100 شخص عند إعطائهم مغذي ما يتوزع توزيع بواسون بوسط حسابي 2، أحسب الاحتمالات:

أ) أن تكون النتيجة سلبية عند إعطاء المغذي،

ب) أن يكون عدد الاستجابات السلبية 3 على الأكثر،

3) بفرض أن متوسط أيام الرقود بأحد المستشفيات لإصابة ما هو 60 يوم، بانحراف معياري 15، فإذا كان عدد أيام الرقود تتبع التوزيع الطبيعي، أوجد الاحتمالات:

a) أن عدد أيام رقود شخص اختير عشوائيا تزيد عن 50 يوم،

b) أن عدد أيام رقود شخص اختير عشوائيا تقل عن 30 يوم،

c) أن عدد أيام رقود شخص اختير عشوائيا ما بين 30- 60 يوم،

الفصل الرابع
الاختبارات الإحصائية

إن من أهم أهداف الدراسة الإحصائية بعد جمع المعطيات واستخدام المقاييس هو الاستدلال واتخاذ القرار، غير أن ذلك لا يتأتى إلا بالتحقق من دقة النتائج ومدى تجانسها مع التوقعات المرسومة، وحيث أن مثل ذلك يتم تناوله من خلال مواضيع متعددة كـ "نظرية المعاينة، نظرية التقدير، اختبار الفرضيات، تحليل التباين، الارتباط والمعاملات،..." غير أننا سنتعرض في هذا الفصل لما هو لازم للدارس، كما أن من لديه الرغبة في التوسع والاستطراد سيجد بغيته في عدة مصادر منها ما سنشير له في قائمة المراجع الواردة بهذا...،

1-3: التقدير واختبار الفرضيات: "Estimation and hypothesis test's"،

يعد التقدير أساسيا للاستنتاج الإحصائي معتمدا على بيانات العينة، إذ أن التقدير يكون لمعلمات "مثل: الوسط، التباين، النسب،..." مشابهة في المجتمع الذي سحبت منه العينة، وحيث أن تقدير المعلمة له صفات التقدير الجيد، فعلى سبيل المثال لدينا الوسط الحسابي $\bar{x}$ للعينة يعد تقدير نقطي جيد إلى μ"الوسط الحسابي للمجتمع" غير أنه بسبب تقلبات المعاينة قد لا يمكن تقع أن $\bar{x} = \mu$، وعليه من المنطق إذا تم تقدير μ بفترة تحوي قيمتها المحتملة، ومثل ذلك ما سنتطرق له تباعا...،

بداية نحتاج للمفاهيم التالية،

- حدود الثقة"Confidence Limit": وهما قيمتان تحصران المعلمة باحتمال معين p، يسمى مستوى الثقة "المعنوية"، وهي لأي من المعلمات، مثل $P(L < \mu < U) = \alpha$، حيث α تمثل الثقة، α - 1 تمثل احتمال الخطأ،

فباعتبار نسبة النقص التي يصاب بها المواليد، لتكن q تمثل نسبة الذين يحملون الصفة، p نسبة من لا يحملون الصفة،

المجتمع "Population"/	العينة "Sample"/
$P = \dfrac{A}{N}$	$\hat{P} = \dfrac{a}{n}$

وبمعلومية كل من $n,\ p$، سيكون $\sigma = \sqrt{\dfrac{pq}{n}},\ q = 1 - p$،

أي أن حدود الثقة لنسب المجتمع"Proportion" تتمثل فيما يلي: $\hat{p} - z_{1-\frac{\alpha}{2}}\sigma < P < \hat{p} + z_{1-\frac{\alpha}{2}}\sigma$،

حيث نبحث عن z المقابلة للقيمة α من الجدول،

مثال1:

باعتبار نسبة الولادة "ولد" هي 0.48، أوجد حدود الثقة باحتمال خطأ 5% لعدد 500 ولادة،

الحل:

لدينا $p = 0.48,\ q = 0.52$ ، لذا سيكون $\sigma = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.48)(0.52)}{500}} = 0.0222$ ،

وحيث أن قيمة z المقابلة إلى 5% هي 1.96، لذا فإن:

$$L = 0.48 - 1.96(0.0222) = 0.43 \quad \text{and} \quad U = 0.48 + 1.96(0.0222) = 0.52$$ ،

مما يعني أننا على 95% من الثقة بأن نسبة ولادة "ولد" ما بين 0.43 -0.52،

مثال2:

لوحظ أن 15 ولادة مشوهة من ضمن 1800 ولادة، ما حدود الثقة إذا ولد طفل أن يكون مشوه باحتمال خطأ 31.8%،

الحل:

لدينا $\hat{P} = \frac{15}{1800}$ ، $q = 1 - p = \frac{1785}{1800}$ ، أي أن:

$$\sigma = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(15/1800)(1785/1800)}{1800}} = 0.00214$$ ،

ولأن قيمة z المقابلة إلى 31.8% هي 1، لذا فإن:

$$L = \frac{15}{1800} - 1(0.00214) = 0.00619 \quad \text{and} \quad U = \frac{15}{1800} + 1(0.00214) = 0.0194$$ ،

مما يعني أن نسبة الولادات المشوهة تكون ما بين 0.00619 -0.0194،

مثال3:

إذا كان احتمال الإصابة بوباء معين باليمن $\frac{1}{1000}$ ، فإذا كان في مدينة الحديدة 57 إصابة لكل 50000 شخص، هل أن الحديدة كعينة تمثل اليمن،

الحل:

لدينا: $\sigma = \sqrt{\frac{1}{1000}\left(1 - \frac{1}{1000}\right)} = 0.0001413$ ، $p = \frac{1}{1000},\ \hat{p} = \frac{57}{50000}$ ، ومنه $z = \left|\frac{p - \hat{p}}{\sigma}\right| = \left|\frac{1/1000 - 57/50000}{0.0001413}\right| \approx 1$ ،

أي أن قيمة α المقابلة إلى z=1 هي 31.8%، وهذا يعني أن الحديدة تمثل المجتمع اليمني بثقة 62.2%،

- الفرضية "Hypothesis": وهي ما يعبر عن معلمات المجتمع، ولمعرفة مدى انسجامها مع البيانات المتوفرة يستخدم ما يسمى باختبار الفرضيات،

حيث لدينا ما يعرف بفرضية العدمH_o"nullhypothesis"تشير لعدم وجود فرق وهي إما تقبل أو ترفض، عند رفضها يعني ذلك أن البيانات منسجمة مع فرضية بديلة H_1 "alternative hypothesis"،

سنتطرق فيما يلي لاختبار فرضيات مثل "الوسط الحسابي لمجتمع، الفرق بين وسطي المجتمع، الفرق بين نسيبتي مجتمعين" وذلك في ضوء حالات إما أن يكون فيها التباين معلوم أو غير معلوم، متساوي أو غير متساوي...، كما أن تطبيقات أخرى في هذا الشأن مثل ما يعرف بمقارنة التباين واختبار الفروق بينها "Analysis of Variance" إضافة إلى ما يعرف بالمقارنات المتعددة وما إلى ذلك مما لم نتعرض له في هذا المؤلف، غير أن من له رغبة في التوسع سيجد بغيته في عدة مصادر أخرى منها ما من مرجع أسميناه "الإحصاء والاحتمالات للجميع"...،

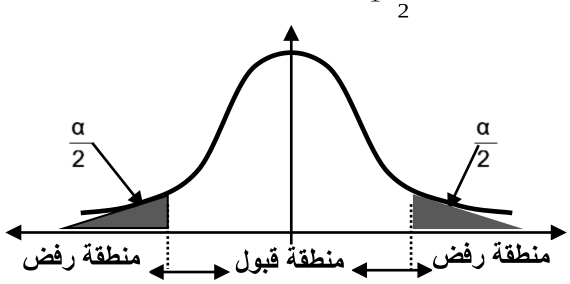
1- اختبار فرضية "الوسط الحسابي μ"للمجتمع:

<u>الحالة الأولى/</u> "المعاينة من مجتمع ذي توزيع طبيعي والتباين معلوم"،

الفرضيات: $H_o: \mu = \mu_o,\ \ H_1: \mu \neq \mu_o$، حيث μ_o معلومة، واستنادا لذلك نستخدم القاعدة $z = \dfrac{\bar{x} - \mu_o}{\sigma / \sqrt{n}}$،

حيث نرفض H_o إذا كان قيمة z المحتسبة أكبر من المجدولة "$z_{cal} \geq z_{tab}\ \ or\ \ z \geq z_{1-\frac{\alpha}{2}}$"،

والشكل المقابل لمزيد من الإيضاح:



ذلك أن المساحة تحت المنحنى الطبيعي تساوي 1، وحيث أن مستوى المعنوية α تمثل نسبة احتمال "مقدار الخطأ" وفقا لفرضيات معتبرة، فعلى سبيل المثال الفرضية $H_o: \mu = \mu_o$ عند المستوى α تعني عدم وجود فروق بين الأوساط، وبحساب قيمة المتغير العشوائي z الخاضع للتوزيع الطبيعي سيكون $0 \leq z \leq 1$،

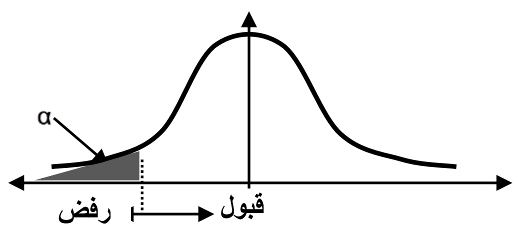
وعندما يكون $z_{cal} \geq z_{tab}$ ذلك يعني أن قيمة z المحتسبة واقعة في منطقة الرفض، وبالعكس إذا كان $z_{cal} < z_{tab}$ ذلك يعني أن قيمة z المحتسبة واقعة في منطقة القبول...،

ملاحظة: الطريقة المبينة في الأعلى تعرف بقاعدة الذيلين، علما أن قواعد أخرى منها ما يعرف بوحيدة الذيل يتم فيها مقارنة قيمة z المحتسبة بالقيمة المجدولة $z_{1-\alpha}$ ،

والفرضيات المبنية على ذلك تتمثل فيما يلي، $H_o : \mu \geq \mu_o,\ \ H_1 : \mu < \mu_o$ ،

حيث نرفض H_o عندما يكون $z_{cal} \leq -z_{tab}\ \ or\ \ z \leq -z_{1-\alpha}$ ،

كما أن الشكل المقابل يفسر المقارنة المتبعة،



والأمثلة التالية توضح ما يلزم...،

مثال4:

في عينة حجمها 10 أشخاص من مجتمع ذي توزيع طبيعي، وجد أن متوسط الإنزيمات للعينة 22 وبانحراف معياري 6.7، اختبر متوسط المجتمع μ=25 عند مستوى 5%.

الحل:

$$z = \frac{(\bar{x}-\mu_o)\sqrt{n}}{\sigma} = \frac{(22-25)\sqrt{10}}{6.7} = -1.415$$ ،

وحيث أن: $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$ ، أي أن $z_{cal} < z_{tab}$،

لذا نقبل الفرضية H_o ، بمعنى عدم وجود فروق،

مثال5:

أستخدم المثال السابق لتحديد ما إذا كان μ<25، عند مستوى معنوية 5%،

الحل:

لدينا $H_o : \mu \geq 25,\ \ H_1 : \mu < 25$ ،

ولأن $$z = \frac{(\bar{x}-\mu_o)\sqrt{n}}{\sigma} = \frac{(22-25)\sqrt{10}}{6.7} = -1.415$$ ،

كما أن $z_{1-\alpha} = z_{0.95} = 1.64$ ، أي أن $z_{cal} > -z_{tab}$،

لذا نقبل الفرضية H_o ، بمعنى عدم وجود فروق،

<u>الحالة الثانية/</u> "المعاينة من مجتمع ذي توزيع طبيعي والتباين مجهول"،

الفرضيات: $H_o: \mu = \mu_o,\ \ H_1: \mu \neq \mu_o$ ، حيث μ_o معلومة، واستنادا لذلك نستخدم القاعدة $z = \dfrac{\bar{x} - \mu_o}{s/\sqrt{n}}$ ،

حيث نرفض H_o إذا كان قيمة z المحتسبة أكبر من المجدولة " $t_{cal} \geq t_{tab}\ \ or\ \ t \geq t_{1-\frac{\alpha}{2}}$ "،

مثال6:

إذا علم أن محددات مصل السكر لعينة حجمها 15 شخص هو 96 وحدة بانحراف معياري 35 وحدة لكل مليمتر، هل أن متوسط المجتمع يختلف عن القيمة 120 عند مستوى معنوية 0.5،

الحل:

لدينا $H_o: \mu = 120,\ \ H_1: \mu \neq 120$ ، كما سيكون $z = \dfrac{\bar{x} - \mu_o}{s/\sqrt{n}} = \dfrac{96-120}{35/\sqrt{15}} = \dfrac{-24}{9.04} = -2.65$ ،

أيضا $t_{1-\frac{\alpha}{2}} = t_{0.975,\ 14} = 2.1448$ ، أي أن $t_{cal} > t_{tab}$ ، لذا نرفض الفرضية H_o ،

مثال7:

استخدم المثال السابق لاختبار الفرضيات: $H_o: \mu \geq 120,\ \ H_1: \mu < 120$ ، عند α=0.05،

الحل:

لدينا $z = \dfrac{\bar{x} - \mu_o}{s/\sqrt{n}} = \dfrac{96-120}{35/\sqrt{15}} = \dfrac{-24}{9.04} = -2.65$ ، ولأن $t_{1-\alpha} = t_{0.95,\ 14} = 1.7613$ ،

لذا سيكون $t_{cal} \leq -t_{tab}$ ، لذا نرفض الفرضية H_o ،

2- اختبار فرضية "الفرق بين وسطي مجتمع":

<u>الحالة الأولى/</u> "التباين معلوم"،

الفرضيات: " $H_o: \mu_1 = \mu_2,\ \ H_1: \mu_1 \neq \mu_2$ "، واستنادا لذلك نستخدم القاعدة: $z = \dfrac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ ،

وعندما تكون الفرضية H_o صحيحة فإن z تتبع التوزيع الطبيعي المعياري،

لذا سيكون: $z=\dfrac{\overline{x_1}-\overline{x_2}}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$ ، وعليه نرفض H_o عندما يكون $z_{cal}\ge z_{tab}\ \ or\ \ z\ge z_{1-\frac{\alpha}{2}}$ ،

مثال8:

في احد المستشفيات تم أخذ عينة من 12 شخص بهم تشوهات، فوجد أن متوسط مقدار سائل الحمض البولي هو 4.5 ملغم/100مليلتر، وفي مستشفى آخر تم أخذ عينة من 15 شخص اعتيادي بنفس العمر النسبي فوجد أن المتوسط 3.4 ملغم/100مليلتر، هل من فرق بين الأوساط عند المستوى α=0.05،

الحل:

لدينا " $H_o:\mu_1=\mu_2,\ \ H_1:\mu_1\ne\mu_2$ "، وحيث أن $z=\dfrac{\overline{x_1}-\overline{x_2}}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}=\dfrac{4.3-3.4}{\sqrt{\dfrac{1}{12}+\dfrac{1}{15}}}=2.82$ ،

أيضا $z_{1-\frac{\alpha}{2}}=z_{0.975}=1.96$ ، أي أن $z_{cal}\ge z_{tab}$ ، لذا نرفض H_o، بمعنى أن الوسطين غير متساويين،

<u>الحالة الثانية/</u> "التباين مجهول"،

الفرضيات: $H_o:\mu_1=\mu_2,\ \ H_1:\mu_1\ne\mu_2$، واستنادا لذلك نستخدم القاعدة $t=\dfrac{\overline{x_1}-\overline{x_2}-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}$ ،

كما نرفض H_o إذا كان " $t_{cal}\ge t_{tab}\ \ or\ \ t\ge t_{d.f,\ 1-\frac{\alpha}{2}}$ "، حيث $d.f=\dfrac{\left(S_1^2/n_1+S_2^2/n_2\right)^2}{\dfrac{\left(S_1^2/n_1\right)^2}{n_1}+\dfrac{\left(S_2^2/n_2\right)^2}{n_2}}$ ،

مثال9:

تم قياس مقدار (CHCO)لعشرين مفردة وعشر مفردات مصابة بمرض معين فكانت النتيجة كما مبينة في الجدول المقابل:

المفردة	n	$\bar{x}$	S
الطبيعية	20	47.2	10.1
المصابة	10	62.6	33.8

هل من فرق بين الأوساط عند α=0.05،

الحل:

لدينا $t=\dfrac{\overline{x_1}-\overline{x_2}}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}=\dfrac{62.6-47.2}{\sqrt{\dfrac{(33.8)^2}{10}+\dfrac{(10.1)^2}{20}}}=\dfrac{15.4}{10.92}=1.41$ ،

ولأن $d.f=\frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1}+\frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2}}=\frac{\left(\frac{33.8}{10}+\frac{10.1}{20}\right)^2}{\frac{\left(\frac{33.8}{10}\right)^2}{10}+\frac{\left(\frac{10.1}{20}\right)^2}{20}}=11$ ،

لذا فإن $t_{d.f,\ 1-\frac{\alpha}{2}}=t_{11,\ 0.975}=2.20$ ، وحيث $t_{cal}<t_{tab}$ ، لذا نقبل الفرضية H_o ، أي أن الوسطين متساويين،

2- اختبار فرضية "الفرق بين نسبتي مجتمعين":

ويستخدم هذا الاختبار في التجانس بين عينتين،

الفرضيات: " $H_o: p_1=p_2,\ \ H_1: p_1\neq p_2$ "، واستنادا لذلك نستخدم القاعدة: $z=\frac{|p_1-p_2|}{Sd}$ ، حيث:

" $\hat{p}=\frac{n_1p_1+n_2p_2}{n_1+n_2},\ \hat{q}=1-\hat{p},\ S^2d=S_1^2+S_2^2,\ S^2d=\frac{\hat{p}(1-\hat{p})}{n_1}+\frac{\hat{p}(1-\hat{p})}{n_2}=\hat{p}\hat{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$ "،

كما نرفض H_o عندما يكون $z_{cal}\geq z_{tab}\ \ or\ \ z\geq z_{1-\frac{\alpha}{2}}$ ،

مثال10:

مدينتين نسبة الولادة غير الطبيعية في الأولى 25 من كل 800 ولادة، وفي الثانية 100 ولادة من كل 2500 ولادة، هل من فرق بين النسبتين عند مستوى معنوية 5%،

الحل:

$p_1=\frac{25}{800}=0.03125,\ \ p_2=\frac{100}{2500}=0.041$ ، لذا سيكون: $\hat{p}=\frac{n_1p_1+n_2p_2}{n_1+n_2}=\frac{25+100}{800+2500}=0.0378$ ،

أي أن $z=\frac{|p_1-p_2|}{Sd}=\frac{|0.03125-0.041|}{\sqrt{(0.0378)(0.9622)\left(\frac{1}{800}+\frac{1}{2500}\right)}}=1.13$ ،

كما أن $z_{1-\frac{\alpha}{2}}=z_{0.975}=1.96$ ، ولأن $z_{cal}<z_{tab}$ ، لذا نقبل H_o ، بمعنى عدم وجود فرق معنوي بين النسبتين،

مثال11:

جرب نوعين من الدواء على 100 أرنب، إذا كان تأثير الدواء "+" وعدم التأثير "-"،

معطى وفقا للجدول المقابل،

ناقش الفروق في التأثير عند α=0.05،

النوع الأول "A"	-	-	+	+
النوع الثاني "B"	-	+	-	+
عدد الأرانب	35	5	15	45

الحل:

لدينا $p_A = \frac{15+45}{100} = 0.6,\ \ p_B = \frac{5+45}{100} = 0.5$ ، ومنه $\hat{p} = \frac{n_1 p_A + n_2 p_B}{n_1 + n_2} = \frac{50+60}{200} = 0.55$ ،

كما أن $z = \frac{|0.6-0.5|}{\sqrt{(0.55)(0.45)\left(\frac{1}{100}+\frac{1}{100}\right)}} = 1.4$ ، وحيث $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$ ، ومنه $z_{cal} < z_{tab}$ ،

لذا نقبل H_o ، بمعنى عدم وجود فرق معنوي بين تأثير الدوائين،

نشاط/ 1) عينة حجمها 25 شخص، سحبت عشوائيا من مجتمع تعداده 100 مليون، حيث وجد أن 60% غير مهتمين بنظافة الأسنان، ما حدود الثقة باحتمال خطأ 5%.

2) لتقدير نسبة الإصابة بمرض السكر لدى من تكبر أعمارهم عن 60 سنة، سحبت عينة بحجم 170 عشوائيا من مجتمع ما، حيث وجد أن 34 شخص مصاب في تلك العينة، ما حدود الثقة باحتمال خطأ 5%.

3) في أحد مراكز العلاج الطبيعي، جرى تقدير متوسط القوة لعضلة الساعد لدى مجموعة من الأشخاص، فإذا كان درجات القوة لهذه العضلة تتبع التوزيع الطبيعي بتباين 144، وبعد أخذ عينة حجمها 25 ظهر أن المتوسط هو 84.3، هل من الممكن استنتاج أن وسط قوة العضلة أقل من 95 عند مستوى 5%.

4) تم اجراء تحديد عصارة البنكرياس لعينة من 15 شخص، بمتوسط 96 وحدة/100مليلتر، وانحراف معياري 35 وحدة/100مليلتر، إذا كان تباين المجتمع مجهول أختبر أن μ=15 عند مستوى 5%.

5) عينة حجمها 10 من الإناث بعمر 12 سنة، وأخرى 10 من الذكور بعمر 12 سنة، متوسط الأطوال 8، 5 على التوالي، بفرض التوزيع الطبيعي للأطوال بانحراف معياري 11، 10 على التوالي، هل أن البيانات تشير بأن معدل طول الإناث أكبر من معدل طول الذكور عند مستوى 5%.

6) إذا كان 24 حيوانا يعانون من نقص فيتامين D، تم تقسيمهم لمجموعتين بالتساوي أعطيت الأولى علاج للنقص وتركت الأخرى بدون علاج، وبعد انتهاء التجربة وجد أن نسبة الكالسيوم ملغم/100مليمتر، كما بالجدول المقابل:

مجموعة1	$\overline{x_1} = 11.1$	$S_1 = 0.5$
مجموعة2	$\overline{x_2} = 7.8$	$S_2 = 0.75$

هل أن البيانات تشير بوجود فرق معنوي عند مستوى 5%.

اختبار مربع كاي χ^2 "Chi-square test":

في بعض التجارب نجد أن البيانات على شكل تكرارات ضمن اختبارات مثل "جودة التطابق (goodness of fit)، الاستقلال (Independent)، التجانس (Homogeneity)،..." ويمكن اعتبار كل ذلك كاختبارات جودة التطابق، لأننا نحتاج لاختبار تطابق التكرارات المشاهدة مع التكرارات النظرية المتوقعة، وعليه نجد أن اختبار χ^2 كاختبار مطابقة هو ما يناسب مثل تلك المقاصد...،

1- اختبار جودة التطابق "Tests of goodness of fit":

ويستخدم هذا الاختبار لاتخاذ القرار فيما إذا كان التوزيع المشاهد للتكرارات غير منسجمة مع توزيع نظري مفترض، وذلك وفقا للقاعدة: $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$، حيث O_i التكرار المشاهد، E_i التكرار النظري، وذلك لصفة وحيدة معينة، وعندما تكون قيمة χ^2 المحتسبة أكبر من المجدولة نرفض الفرضية H_o،

مثال1:

في المجتمع الأوربي وجد أن أصناف الدم حسب الجدول:

A	B	AB	O
43%	12%	5%	40%

،

أخذت عينة حجمها 100، تبين أن تقسيمهم حسب الفصيلة هو كما في الجدول:

A	B	AB	O
35	20	10	35

،

هل أن توزيع الدم متطابق بين المجتمع والعينة عند مستوى معنوية 5%،

الحل:

سنكون الجدول المساعد التالي،

صنف الدم	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
A	35	43	8	64	1.488
B	20	12	8	64	5.333
AB	10	5	5	25	5
O	35	40	5	25	0.625
المجموع	100	100	#	#	12.446

،

وحيث أن قيمة χ^2 المجدولة عند درجات حرية "n-1=3" ومستوى معنوية α=0.05، هي 7.81،

أي أن $\chi^2_{cal} > \chi^2_{tab}$ لذا سنرفض الفرضية H_o،

بمعنى أنه لا تطابق بين قراءات العينة والمجتمع،

مثال2:

تم في أحد الأقطار دراسة نسبة المرضى بعدد من المستشفيات، حيث أخذت عينة حجمها 250 مستشفى، وكانت النتائج كما بالجدول:

110-100	-90	-80	-70	-60	-50	-40	-0	النسبة
4	22	55	62	51	22	18	16	ع.المشافي

،

هل أن تلك البيانات تشير بأن العينة مسحوبة من مجتمع بتوزيع طبيعي،

الحل:

سنكون الجدول المقابل،

$(x_i-\bar{x})^2 f_i$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})$	$x_i f_i$	x_i	f_i	النسب
399936	2496.0	-49.96	320	20	16	-0
11214.0	623.00	-24.96	810	45	18	-40
4923.63	223.80	-14.96	1210	55	22	-50
1254.68	24.602	-4.96	3315	65	51	-60
1574.89	25.402	5.04	4650	75	62	-70
12441.0	226.20	15.04	4675	85	55	-80
13794.0	627.00	25.04	2090	95	22	-90
4911.20	1227.8	35.04	420	105	4	110-100
90049.5	#	#	17490	#	#	المجموع

لاحظ أن:

$$\bar{x}=\frac{\sum x_i f_i}{\sum f_i}=\frac{17490}{250}=69.96$$

،

لذا سيكون $S=\sqrt{\frac{\sum(x_i-\bar{x})^2 f_i}{\sum f_i -1}}=\sqrt{\frac{90049.597}{249}}=19.02$ ، ومنه سنكون الجدول التالي:

$\chi^2=\frac{(O_i-E_i)^2}{E_i}$	E_i	$P(Z)$	Z	O_i	النسب
-	14.55 [0	0	-3.68	16 [0 16	-0 40
0.145	14.55	0.0582	-1.57	-	-0
0.788	22.18	0.0887	-1.05	18	-40
7.173	38.65	15.46	-0.52	22	-50
0.038	49.62	0.1985	0.00	51	-60
2.629	50.48	0.2019	0.58	62	-70
7.197	38.38	0.1535	1.06	55	-80
0.001	21.88	0.0875	1.58	22	-90
3.533 4.350	9.92 4.35	0.0397 0.0174	2.41 -	4 0	-100 110
25.854	#	1	#	250	المجموع

حيث O_i تمثل التكرارات المشاهدة"عدد المستشفيات"، وأن x_i المستخدمة في $Z=\frac{x_i-\bar{x}}{S}$ تمثل الحد الأدنى للفئة،

كما أن $P(Z)$ تمثل الاحتمال المتوقع، ولتوضيح حساب ذلك: $P(Z<-3.68)=1-P(Z<3.68)=1-1=0$ ،

$P(-3.68<Z<-1.57)=P(Z<-1.57)-P(Z<-3.68)=\{1-P(Z<1.57)\}-\{1-P(Z<3.68)=1-0.9418-1+1=0.0582$ ، وهكذا،

أيضا E_i تحسب في كل حالة وفقا للقاعدة $E_i=nP(Z)$ ، وللتوضيح $E_1=nP(Z)=250\times 0.0582=14.55$ ، وهكذا،

ولأن χ^2 المجدولة هي $\chi^2_{n-k-1,\ \alpha} = \chi^2_{9-2-1,\ 0.05} = \chi^2_{6\ 0.05} = 12.592$ ، حيث n تمثل عدد الفئات، k تمثل عدد المعالم المقدرة "μ, σ^2"، ومنه $\chi^2_{cal} > \chi^2_{tab}$ لذا نرفض H_o، بمعنى أن بيانات العينة ليس لها توزيع طبيعي،

ملاحظة:

التكرار المتوقع "E_i" يستلزم أن يكون أكبر من الواحد، لذا يتم دمج الفئة التي تكرارها أقل من الواحد مع الفئة السابقة أو اللاحقة بها،

مثال3:

مجموعة من 100 طبيب، قام كل منهم باختيار عينة عشوائية حجمها 20 مريض، حيث أعطي لكل مريض نوع جديد من الدواء مع تقديم التساؤل حول تفضيل المريض لذلك الدواء، فإذا كان الجدول التالي يمثل البيانات المحصلة:

م	10	9	8	7	6	5	4	3	2	1	0	x_i
10	0	9	10	10	17	15	10	10	8	6	5	ع.الأطباء

بين فيما إذا كانت البيانات تتبع توزيع ثنائي الحدين، حيث x_i تمثل عدد الذين يفضلون الدواء من 25 مريض،

الحل:

سنكون الجدول المقابل،

$\chi^2 = \frac{(O_i - E_i)^2}{E_i}$	E_i	$f(x)$	$O_i x_i$	O_i	x_i
2496.0	2.74 [0.38	0.0038	0	11 [5	0
24.901	2.36	0.0236	6	6	1
0.119	7.08	0.0708	16	8	2
0.944	13.58	0.0136	30	10	3
4.026	18.67	0.1867	40	10	4
1.079	19.60	0.1960	75	15	5
0.027	16.33	0.1633	102	17	6
0.107	11.09	0.1109	70	10	7
2.281	6.23	0.0623	80	10	8
12.407	2.95	0.0295	81	9	9
1.73	1.73	0.0173	0	0	≤ 10
47.624	100	1	500	#	م

ذلك أنه لحساب التكرار المتوقع $f(x)$، نحتاج بداية لحساب كل من p، q،

ذلك أن: $P = \frac{\sum O_i x_i}{n \sum O_i} = \frac{500}{25(100)} = 0.2, \quad q = 1 - 0.2 = 0.8$ ،

ومنه سيكون: $f(x) = \binom{25}{x} (0.2)^x (0.8)^{25-x}$ ،

ولأن χ^2 المجدولة هي $\chi^2_{n-k-1,\ 1-\alpha} = \chi^2_{10-1-1,\ 0.95} = \chi^2_{8\ 0.95} = 15.509$ ،

حيث n تمثل عدد الفئات، k تمثل عدد المعالم المقدرة "P"، ومنه $\chi^2_{cal} > \chi^2_{tab}$ لذا نرفض H_o،

بمعنى أن البيانات المحصلة لا تتبع توزيع ثنائي الحدين،

مثال4:

بأحد المستشفيات تم دراسة عدد حالات الطوارئ اليومية، فإذا كانت البيانات المحصلة خلال 90 يوم كما في الجدول التالي:

ع.الحالات باليوم	0	1	2	3	4	5	6	7	8	9	10	م
ع.الأيام	5	14	15	23	16	9	3	3	1	1	0	90

،

بين فيما إذا كانت البيانات تتبع توزيع بواسون بمعدل $\lambda=3$ لكل يوم،

الحل:

سنكون الجدول المقابل،

x_i	O_i		$\mathrm{f}(x)$		E_i	$\chi^2 = \frac{(O_i - E_i)^2}{E_i}$
0	5		0.050	4.5		0.055
1	14		0.149	13.41		0.026
2	15		0.224	20.16		1.321
3	23		0.224	20.16		0.400
4	16		0.168	15.12		0.051
5	9		0.101	9.09		0.001
6	3		0.05	4.5		0.5
7	3		0.022	1.98		0.525
8	1	2	0.008	0.72	1.8	0.783
9	1		0.003	0.27		
≤ 10	0		0.001	0.09		
م	90		1	90		3.662

ويحسب التكرار المتوقع بالقاعدة: $\mathrm{f}(x) = \frac{e^{-3}3^x}{x!}$ ، وذلك في كل الحالات...،

ومن الجدول $\chi^2_{n-k-1,\,1-\alpha} = \chi^2_{8\ 0.95} = 15.507$، أي أن $\chi^2_{cal} < \chi^2_{tab}$ لذا نقبل H_o، بمعنى أن البيانات تتبع توزيع بواسون،

2- اختبارات الاستقلال "Tests of independence":

ويستخدم هذا الاختبار فيما إذا كان هناك جدول توزيع مزدوج، بمعنى أن المطلوب اختبار الاستقلال لمعيارين مختلفين، فلأي عينة حجمها n، يمكن تبويب البيانات المتعلقة بصنفين من خصائص العينة وفقا للجدول المقابل:

معيار1"L" / **معيار2"k"**	1	2	…	j	…	م
1	n_{11}	n_{21}	…	n_{j1}	…	n_{L1}
2	n_{12}	n_{22}	…	n_{j2}	…	n_{L2}
⋮	⋮	⋮		⋮		
i	n_{1i}	n_{2i}	…	n_{ij}	…	n_{Lj}
⋮	⋮	⋮		⋮		
م	n_{1k}	n_{2k}		n_{ik}		n

حيث n_{ij} تمثل التكرارات المشاهدة O_{ij}، كما أن التكرارات النظرية المقابلة تحسب بالقاعدة: $E_{ij} = \frac{n_{ik} \times n_{Lj}}{n}$ ،

أيضا سيكون: $\chi^2=\sum_{ij}^{Lk}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$، وعندما يحصل أن $\chi^2_{cal}>\chi^2_{tab}$، "وذلك لدرجات حدية (L-1)(k-1)"

"ومستوى معنوية α" سنرفض فرضية العدم H_0 "المعيارين مستقلين"،

مثال5:

أراد باحث دراسة العلاقة بين صنف الدم ومدى صعوبة حالة معينة في المجتمع، أخذ عينة عشوائية حجمها 1500، وكانت النتائج كما يلي:

"k" / "L"	A	B	AB	O	م
0	543	221	90	476	1320
1	44	22	8	31	105
2	28	9	7	31	75
م	615	242	105	538	1500

،

ما العلاقة بين **L**"صنف الدم"، **k**"درجة صعوبة الحالة 0، 1، 2" منعدمة، متوسطة، عالية، على الترتيب، وذلك عند مستوى معنوية α=5%،

الحل:

لحساب التكرار النظري المتوقع:

$$E_{ij}=\frac{n_{ik}\times n_{Lj}}{n},\ \text{So}\ E_{11}=\frac{615\times1320}{1500}=541.2,\ E_{12}=212.96,\ E_{13}=92.4,\ E_{14}=473.44,$$ ،

$$E_{21}=43.05,\ E_{22}=16.94,\ E_{23}=7.35,\ E_{24}=37.66,\ E_{31}=30.75,\ E_{32}=12.10,\ E_{33}=5.25,\ E_{34}=26.9$$

ومنه سيكون: $\chi^2=\sum_{ij}^{Lk}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}=\frac{(543-541.2)^2}{541.2}+\frac{(221-212.96)^2}{212.96}+\ldots+\frac{(31-26.9)^2}{26.9}=5.12$،

كما أن: $\chi^2_{(L-1)(k-1),\,1-\alpha}=\chi^2_{6\ 0.95}=12.592$، أي أن $\chi^2_{cal}<\chi^2_{tab}$ لذا نقبل H_0، بمعنى عدم وجود علاقة بين المعيارين،

مثال6:

مجموعة من 350 مشارك في دراسة صحية، كان سؤالهم عن إتباع رجيم، وكانت النتائج كما في الجدول المقابل:

الإجابة / الجنس	ذكر	أنثى	م
Y	14	25	39
N	159	152	311
م	173	177	350

،

هل من إشارة بوجود علاقة بين إتباع رجيم والجنس عند مستوى α=5%،

الحل:

لحساب التكرار النظري المتوقع: $E_{11}=\frac{615\times1320}{1500}=19.27,\ E_{12}=19.72,\ E_{13}=153.72,\ E_{14}=157.27$،

ومنه سيكون: $\chi^2=\frac{(14-19.27)^2}{19.27}+\frac{(25-19.27)^2}{19.27}+\frac{(159-153.72)^2}{153.72}+\frac{(152-157.27)^2}{157.27}=3.2$،

أيضا: $\chi^2_{(L-1)(k-1),\,1-\alpha} = \chi^2_{1\;0.95} = 3.84$، ومنه $\chi^2_{cal} < \chi^2_{tab}$ لذا نقبل H_o،

بمعنى عدم وجود علاقة بين إتباع رجيم والجنس،

2- اختبارا التجانس "Test of Homogeneity":

وفي هذا الاختبار تسحب عينات مستقلة من المجمعات المراد دراستها، ثم تصنف وفق معيار معين إلى مستويات، بمعنى أن المطلوب اختبار تجانس المجتمع "لاحظ الفرق بين هذا الاختبار والاختبارات السابقة"، كما تتضح طريقة اختبار التجانس من خلال المثال التالي،

مثال5:

تم تقسيم مجموعة من 400 ماعز إلى مجموعتين متساويتين، أعطيت الأولى **"F"** لقاح V_1، والثانية **"S"** لقاح V_2، وكانت النتائج حسب الجدول:

الاستجابه / الصنف	F	P	G	E	∑
F	20	160	16	4	200
S	16	174	6	4	200
∑	36	334	22	8	400

حيث "F، P، G، E" تشير إلى الاستجابات "ضعيفة، متوسطة، جيدة، مرتفعة" على الترتيب،

هل تأثير اللقاحات متشابه عند مستوى $\alpha=5\%$،

الحل:

لحساب التكرار النظري المتوقع:

لدينا: $E_{ij} = \dfrac{n_{ik} \times n_{Lj}}{n}$، ومنه سيكون:

$$E_{11} = \frac{36 \times 200}{400} = 18,\ E_{12} = 167,\ E_{13} = 12,\ E_{14} = 3,\ E_{21} = 18,\ E_{22} = 167,\ E_{23} = 12,\ E_{24} = 3$$ ،

كما أن: $$\chi^2 = \sum_{ij}^{Lk} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(20-18)^2}{18} + \frac{(160-167)^2}{167} + \ldots + \frac{(2-3)^2}{3} = 4.36$$ ،

أيضا لدينا: $\chi^2_{(2-1)(4-1),\,1-\alpha} = \chi^2_{3\;0.95} = 7.815$، أي أن $\chi^2_{cal} < \chi^2_{tab}$ لذا نقبل H_o،

بمعنى تشابه الاستجابة بين اللقاحات،

نشاط/ 1) البيانات التالية تمثل أطوال 300 شخص بالسنتيمتر عند عمر ثمان سنوات،

الطول	5	10	14	21	30	40	45	43	42	30	11	5	4
التكرار المشاهد	116>	-116	-118	-120	-122	-124	-126	-128	-130	-132	-134	-136	-138

اختبر الفرضية H_o بأن البيانات مسحوبة من مجتمع بتوزيع طبيعي عند مستوى معنوية 5%.

2) إذا كان السجل بأحد المستشفيات مكون من 10 حقول، سحبت عينة من 100 سجل وظهرت الأخطاء فيها كما في الجدول المقابل:

ع. الأخطاء في10 حقول	0	1	2	3	4	5<
التكرار المشاهد	8	25	32	24	10	1

،

هل أن البيانات تتبع توزيع ثنائي الحدين عند مستوى معنوية 5%.

3) البيانات التالية تمثل الحوادث لعشر سنوات سابقة، 300 شخص بالسنتيمتر عند عمر ثمان سنوات،

ع.الحوادث شهريا	0	1	2	3	4	5	6	7	8	9	10
ع. الشهور المتكرر فيها الحالة	2	10	15	30	28	15	10	6	2	1	1

،

هل أن البيانات تتبع توزيع بواسون عند مستوى معنوية 5%.

4) وضع مصنع إنتاجا يشك في علاقته بتلف الشعب الهوائية، سحبت عينة من 500 شخص، وظهرت النتائج كما في الجدول المقابل:

المجموع	منخفض	محدود	مرتفع	المستوى
235	17	33	185	Y
265	72	73	120	N
500	89	106	305	∑

،

هل من علاقة بين مستوى التعرض "Y, N" وتلف الشعب "منخفض، محدود، مرتفع" عند مستوى 5%.

5) في دراسة لتلوث الهواء، اختيرت عينة من 200 منزل بمنطقتين "I, II"، كان السؤال لشخص ما في كل المنازل ما إذا كان أحد ساكني المنزل متأثرا من تلوث الهواء، كانت البيانات المحصلة كما في الجدول المقابل:

المجموع	نعم	لا	المنطقة
200	157	43	I
200	119	81	II
400	276	124	∑

،

هل يمكن للباحث القول أن المنطقتين تختلف بالنسبة لمتغير الدراسة عند مستوى معنوية 5%.

6) في دراسة لوفيات الرضع بأحد المستشفيات، حسب مستوى ثقافة الأم الصحية، كانت النتائج المحصلة كما في الجدول المقابل:

المجموع	مثقفة	غيرمثقفة	الوفاة
193	133	60	موجدة
787	327	460	غيرموجو
980	460	520	∑

،

هل من علاقة بين الوفاة وثقافة الأم عند مستوى معنوية 5%.

7) في دراسة لوفيات الرضع بأحد المستشفيات، حسب دخل الأسرة، كانت النتائج المحصلة كما في الجدول المقابل:

المجموع	دخل ضعيف	دخل مرتفع	الوفاة
193	104	89	موجدة
787	400	387	غيرموجودة
980	504	976	∑

هل من علاقة بين الوفاة والدخل عند مستوى معنوية 5%.