الكلية التقنية الادارية/الموصل

قسم تقنيات الاحصاء والمعلوماتية

المستوى الثالث

الاحصاء الحيوي1

Biostatistics1

مدرس المادة

هبة لقمان امين

مدرس

مقدمة في علم الاحصاء

عرف علم الاحصاء قديما" اذ وردت اشارة عن العد من قبل المؤرخ اليوناني هيرودوتس حيث ذكر انه في العام 480 ق.م استعمل احد قادة الجيوش طريقة بدائية بسيطة لمعرفة عدد جيشه . وفي القران الكريم وردت اشارات كثيرة تقرن الاحصاء بعملية العد ، قال تعالى (لقد احصاهم وعدهم عدا") [سورة مريم ، الاية 94] ، (... وان تعدوا نعمة الله لاتحصوها) [سورة مريم ، الاية 34] . وفي السيرة النبوية مثال رائع لعملية العد وذلك عندما قام النبي (ص) بتقدير بالغ الدقة لعدد جيش قريش يوم بدر حينما علم عنان قريش تنحر لجيشها كل يوم تسع من الابل .

ومع مرور الزمن ومع ظهور وتطور الدول ازداد الاهتمام بالبيانات والمعلومات وتصنيفها وتبويبها لتستخدم في تسهيل امور الدولة والمجتمع كحفظ سجلات الجند ، وتحديد السكان الخاضعين للضريبة واحصاءات ايرادات الدولة ونفقاتها ، ويمكن القول ان علم الاحصاء قد نشأ مرتبطا" بالمجالات الاجتماعية والاقتصادية .

فيما بعد اهتم الباحثون بجمع البيانات والارقام من بعض الظواهر التي تتكرر تلقائيا" او تجريبيا" لايجاد تفسيرات علمية لهذه الظواهر ، فبدأ علم الاحصاء يخدم العلوم التجريبية مثل الصحة العامة ، الزراعة ، التجارة ، ادارة الاعمال ، التأمين ، علم الاجتماع ، علم النفس وغيرها من العلوم .

وكنتيجة لدخول الاحصاء وسيلة تحليل في المجالات العلمية المختلفة اصبح من الضروري اللجوء الى اساليب متقدمة وعميقة في التحليل الاحصائي فبدت الحاجة الى الى نظريات في علم الاحصاء الرياضي Mathematical Statistic وكان قد سمي في مراحل سابقة بعلم العد ، وعلم الحساب السياسي ، وعلم الاحصاء الحيوي ، والاحصاء الاجتماعي والاقتصادي .

تعريف الاحصاء واهميته

يتكون مصطلح Statistics والذي بناءه Stat – is – tics المشتق من كلمة State أي الدولة والذي يعني مجموعة الحقائق الخاصة بشؤون الدولة .

وقد وردت تعريفات كثيرة لعلم الاحصاء حيث اصبح يعرف حديثا" بانه مجموعة النظريات والطرق العلمية التي تهدف الى جمع وعرض ووصف وتحليل البيانات واستخدام النتائج في التفسير او التعميم او التنبؤ او التقدير او التحقق .

كم عرف بانه العلم الذي يهتم بتوفير الحقائق الرقمية للظواهر المختلفة ومن ثم ترتيبها وعرضها ثم تحليلها للوصول الى نتائج محددة بدقة بهدف فهم الظاهرة من جهة ووضع المقترحات المختلفة لمتابعة سيرها المستقبلي من جهة اخرى . ولعل ابسط هذه التعريفات هو الذي يرى علم الاحصاء بانه علم جمع ونصنيف وتبويب البيانات وتحليلها وتفسيرها . يكتسب الاحصاء اهميته في مجالات الادارة والاقتصاد والمحاسبة والطب والهندسة والفيزياء وعلوم الاحياء والزراعة والعلوم الاخرى(صلاح الدين حسين الهيتي ، الاساليب الاحصائية في العلوم الادارية ، 2004).

الكلمات الاحصائية الاساسية الخمسة :The five basic words of statistics

هناك كلمات خمسة اساسية في علم الاحصاء وهي : المجتمع ، العينة ، المعلمة ، الاحصائية والمتغير . ان الشخص لايستطيع ان يتعلم الاحصاء مالم يعرف معنى هذه الكلمات الخمسة .(من كتاب "حتى انت تستطيع تعلم الاحصاء"

1- المجتمع : Population (من كتاب خاشع الراوي)

هو عبارة عن جميع القيم او المفردات محل الدراسة .فمثلا" اذا كانت دراستنا تتعلق بأطوال طلبة جامعة الموصل فان المجتمع في هذه الحالة هو اطوال جميع الطلبة في جامعة الموصل ، وقد يكون هذا المجتمع محدودا" مثل عدد الوحدات الانتاجية لمصنع معين ، او يكون غير محدودا" مثل عدد البكتريا في حقل ما.

2- العينة : Sample

هي عبارة عن جزء من المجتمع . وهي مجموعة من المشاهدات اختيرت بطريقة ما من المجتمع . ان دراسة مجتمع ككل قد يكون صعبا" او يحتاج الى وقت وجهد ومال لذا فقد استعيض عن دراسة المجتمع بدراسة العينة ومنها نستنتج خواص المجتمع .

3- المعلمة : Parameter

هي قيمة عددية تصف خاصية من خصائص المجتمع .

4- الاحصائية : Statistic

هي قيمة عددية تصف خصائص العينة .

5- المتغير : Variable (كتاب خاشع الراوي)

هو أي ظاهرة تظهر اختلافات بين مفرداتها ويرمز له بالرمز y (او أي رمز اخر مثل z او x ...) . فعند جمع بيانات حول ظاهرة ما فاننا نرمز للظاهرة بالرمز y وكل مفردة او مشاهدة منها نرمز لها بالرمز y_i فمثلا" عند دراسة اطوال الطلبة في احدى الجامعات فاننا نرمز لصفة الطول بالرمز y وطول أي طالب بالرمز y_i (وتسمى المشاهدة او المفردة (Observation) .

هذا وان قيمة y_i قد تختلف من طالب الى اخر ولهذا نقول بان y متغير Variable .

وهناك نوعين من المتغيرات :

1- متغيرات وصفية او نوعية Qualitative variables

هي تلك الظواهر او الصفات التي لايمكن قياسها مباشرة بالارقام العددية مثل صفة لون العيون (ازرق ، اسود ، بني) والحالة الاجتماعية (غني ، متوسط ، فقير) والجنس (ذكر ، انثى) ...

2- متغيرات كمية Quantitative variables

هي تلك الظواهر او الصفات التي يمكن قياسها مباشرة بأرقام عددية مثل : صفة الطول ، الوزن ، كمية المحصول ...

وتنقسم المتغيرات الكمية الى قسمين :

أ – متغيرات مستمرة او متصلة Continuous variable

فالمتغير المستمر هو المتغير الذي تأخذ المشاهدة او المفردة فيه اية قيمة رقمية في مدى معين . فلو فرضنا بان اطوال طلبة جامعة ما تتراوح بين 130.5 و 170 سم فنقول بان :

$$130.5 \leq y \leq 170$$

أي ان المتغير y ممكن ان ياخذ اية قيمة بين 130.5 ســم و 170 ســم . وكامثلة اخرى على المتغيرات المستمرة هي : درجات الحرارة ، الوزن ، الزمن ، كمية المحصول ...

وبصورة عامة يمكن القول بان كل البيانات التي تقاس تعتبر بيانات لمتغير مستمر.

ب – متغيرات غير مستمرة او منفصلة Discrete variables

المتغير المنفصـل هو المتغير الذي تاخذ المشـاهدة او المفردة فيه قيما" متباعدة او متقطعة غير مسـتمرة . فلو فرضنا ان عدد افراد الاسرة في اربع عوائل هي : 2 ، 3 ، 4 ، 5 فنقول بان y : 2,3,4,5

وكامثلة على المتغيرات غير المستمرة او المنفصلة هي عدد الثمار على النباتات ، عدد الطلبة ... فهي في الغالب تكون اعداد صـحيحة . وبصـورة عامة فان كل البيانات التي نحصـل عليها من العد تعتبر بيانات لمتغير منفصل .

اقسام علم الإحصاء : Branches of statistics

يُقسم الاحصاء الى قسمين :

القسم الاول : الاحصاء الوصفي (وصف البيانات)

تعتبر طريقة جمع البيانات وتبويبها وتلخيصـها من اهم وظائف علم الاحصـاء ، اذ لايمكم الاسـتفادة من البيانات الخام ، ووصف الظواهر المختلفة محل الاهتمام الا اذا تم جمع البيانات وعرضها في شكل جدلي ، او بياني من ناحية ، وحسـاب بعض المؤشـرات الاحصـائية البسـيطة التي تدلنا على طبيعة البيانات من ناحية اخرى .

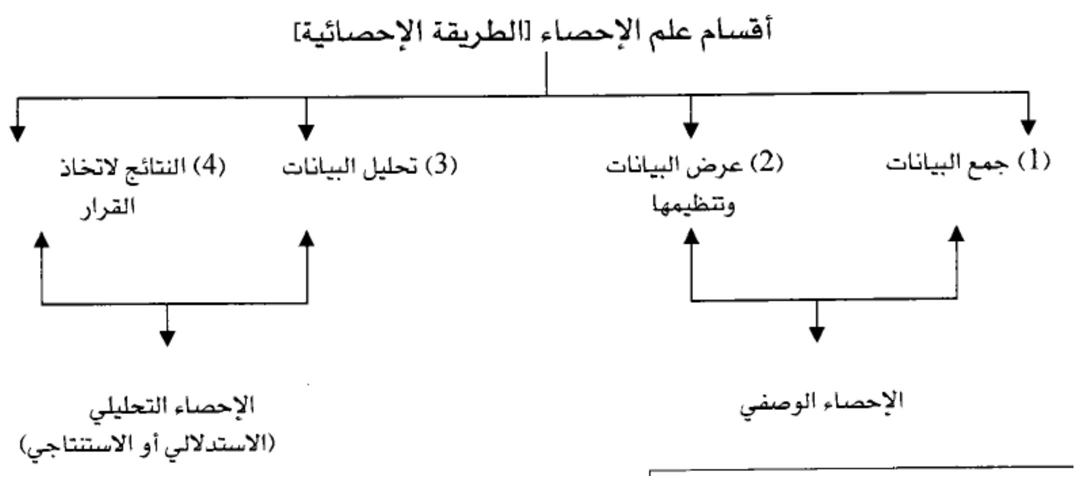
القسم الثاني : الاستدلال الاحصائي

وهو ايضـا" من اهم الوظائف المستخدمة في مجال البحث العلمي ، ويستند الاستدلال الاحصـائي على فكرة اختيار جزء من المجتمع يسـمى عينة بطرقة علمية مناسـبة ، بغرض اسـتخدام بيانات هذه العينة في التوصل الى نتائج ، يمكن تعميمها على مجتمع الدراسة ، ومن ثم يهتم الاستدلال الاحصائي بموضوعين هما :

1- التقدير Estimation : وفيه يتم حسـاب مؤشـرات من بيانات العينة تسـمى احصـاءات Statistics ، تستخدم كتقدير لمؤشرات المجتمع وتسمى معلمات Parameters . ويطلق على المقاييس الاحصائية المحسـوبة من بيانات العينة في هذه الحالة بالتقدير بنقطة Point Estimation ، كما يمكن ايضـا" اسـتخدام المقاييس الاحصـائية المحسـوبة من بيانات العينة في تقدير المدى الذي يمكن ان يقع داخله معلمة المجتمع باحتمال معين ، ويسمى ذلك التقدير بفترة Interval Estimate .

2- اختبار الفروض Tests of Hypotheses : وفيه يتم استخدام بيانات العينة للوصول الى قرار علمي سليم بخصوص الفروض المحددة حول معلمات المجتمع .

ويمكن عرض اقسام علم الاحصاء (الطريقة الاحصائية) حسب المخطط التالي :



(المخطط من كتاب احمد عبد السميع طبية)

الرموز الاحصائية :

مثال : لدينا 5 درجات للطبة في مادة الاحصاء : 77 , 89 , 70 , 85 , 80 : y_i

الظاهرة هي : درجات الطلاب ، المتغير هو y_i

y_1 : القيمة الاولى للمتغير او المشاهدة الاولى .

y_2 : القيمة الثانية للمتغير او المشاهدة الثانية .

وهكذا ...

$y_n = 77$ أي القيمة الاخيرة ($n = 5$) للمتغير او المشاهدة الاخيرة

وعادة يرمز لمجموع قيم المتغير بالرمز :

$$\sum_{i=1}^{n} y_i$$

فالرمز $\sum$ هو حرف اغريقي يسمى (Sigma) أي مجموع الـ ... او summation ، والرقمان 1 و n هما حدا المجموع .

مجموع قيم y ابتداء" من المشاهدة الاولى وحتى الاخيرة أي :

$$\sum_{\mathrm{i}=1}^{\mathrm{n}} y_i = y_1 + y_2 + y_3 + \cdots + y_n$$

وهناك مجموع جزئي مثل

$$\sum_{i=3}^{5} y_i$$

أي مجموع المشاهدة الثالثة والرابعة والخامسة :

$$\sum_{\mathrm{i}=3}^{5} y_i = y_3 + y_4 + y_5$$

ويرمز لمجموع مربعات جميع المشاهدات بالرمز

$$\sum_{\mathrm{i}=1}^{\mathrm{n}} y_1^2$$

وهذا يساوي رياضيا" :

$$\sum_{i=1}^{n} y_i^2 = y_1^2 + y_2^2 + y_3^2 + \cdots + y_n^2$$

ويرمز لمربع مجموع المشاهدات بالرمز $(\sum_{i=1}^{n} \quad _i)^2$ وهو يساوي :

$$(\sum y_i)^2 = (y_1 + y_2 + y_3 + \ldots + y_n)^2$$

كما ويرمز لمجموع حاصل ضرب قيم متغيرين x و y بالرمز $\sum x_i y_i$

$$\sum x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

ولحاصل ضرب مجموعين لقيم متغيرين بالرمز $(\sum x_i)(\sum y_i)$

$$(\sum x_i)(\sum y_i) = (x_1 + x_2 + \cdots + x_n)(y_1 + y_2 + \cdots + y_n)$$

اذا كانت c عدد ثابت فان :

$$\sum c y_i = c y_1 + c y_2 + \cdots + c y_n$$

$$= c\,(y1 + y2 + \cdots + y_n)$$

$= c \sum y_i$

جمع قيم متغيرين او اكثر هو مجموع جمعهم :

$\sum(x_i + y_i) = \sum x_i + \sum y_i$

$= (x_1 + x_2 + \cdots + x_n) + (y_1 + y_2 + \cdots + y_n)$

مثال (1) : (الراوي، (1985))

نفرض بان قيم المتغير y هي كما يلي :

$y_i = 3,9,6,2$

وان قيم المتغير x هي :

$x_i = 4,2,3,7$

اوجد قيمة كل مما يأتي :

$a)\ \sum_{i=1}^{n} y_i \qquad b)\ \sum_{i=2}^{3} y_i \qquad c)\ \sum_{i=1}^{n} y_i^2$

$d)\ \left(\sum y_i\right)^2$ $e)\ \sum x_i y_i$ $f)\ \left(\sum x_i\right)\left(\sum y_i\right)$

الجواب :

a) $n = 4$

$$\sum_{i=1}^{n=4} y_i = 3 + 9 + 6 + 2 = 20$$

b) هنا لدينا مجموع جزئي من القيمة الثانية لـ y الى القيمة الثالثة لـ y :

$$\sum_{i=2}^{3} y_i = 9 + 6 = 15$$

c) $\sum y_i^2=(3)^2+(9)^2\ldots=130$

d) $(\sum y_i)^2 = (3+9+6+2)^2 = 400$

e) $\sum x_i y_i =(4)(3)+(2)(9)+\ldots=62$

f) $(\sum x_i)(\sum y_i) = (16)(20)=320$

مثال (2) :

لديك البيانات التالية :

$x_i = 2,6,3,1 \quad y_i = 3,9,6,2$

اوجد مايلي :

$a) \sum (y_i - x_i)^2$ $b) \sum (x_i - 3)\ (y_i - 5)$ $c) \sum x_i y_i^2$

$d) \sum (y_i - 3)$ $e) \sum y_i - 3$

التوزيعات التكرارية وتمثيلها بيانيا"

اولا : التوزيعات التكرارية :

عند حصولنا على بيانات فاننا نطلق عليها اسم بيانات خام Raw data ، وبعد تلخيص هذه البيانات وتنظيمها في توزيعات تكرارية يطلق عليها بيانات مبوبة .

التوزيعات التكرارية هي عبارة عن جداول تتضمن قيم مرتبة للظاهرة المدروسة مع تكراراتها .

مثال :

البيانات التالية تمثل انواع التمور لخمسة عشر مزرعة :

سكري	خلاص	برحي	خلاص	برحي
برحي	سكري	برحي	صقعي	خلاص
صقعي	برحي	سكري	خلاص	برحي

المطلوب تكوين جدول توزيع تكراري من البيانات السابقة

الحل :

الجدول التكراري يكون على النحو التالي لان البيانات هنا وصفية .

الصفة	التكرار f_i
سكري	3
خلاص	4
برحي	6
صقعي	2
	15

ملاحظة : يجب ان تلاحظ ان مجموع التكرارات دائما" يساوي عدد البيانات .

هناك نوعان من الجداول الاحصائية وهما : (د.خاشع الراوي)

1- الجدول البسيط : وهو الجدول الذي توزع فيه البيانات حسب صفة واحدة . ويتالف من عمودين : الاول يمثل تقسـيمات الصـفة او الظاهرة الى فئات او مجموعات والثاني يبين عدد المفردات التابعة لكل فئة او مجموعة مثل :

فئات الوزن (كغم)	عدد الطلبة
62-60	5
65-63	15
68-66	45
71-69	27
74-72	8
المجموع	100

2- الجدول المركب : وهو الجدول الذي توزع فيه البيانات حسب صفتين او ظاهرتين او اكثر في نفس الوقت.

فمثلا" الجدول لصفتين يتألف من :

الصفوف : تمثل فئات او مجاميع احدى الصفتين .

الاعمدة : تمثل فئات او مجاميع الصفة الاخرى .

اما المربعات التي تقابل الصفوف والاعمدة فتحتوي على عدد المفردات او التكرارات المشتركة في فئات ومجاميع كلا الصفتين مثل :

الوزن (كغم) الطول (سم)	60-51	70-61	80-71	المجموع
121 – 140	20	6	4	30
141 – 160	2	40	10	52
161 – 180	2	6	10	18
المجموع	24	52	24	100

تعاريف

1- الفئات :classes

هي المجاميع التي قسمت اليها قيم المتغير.

2- البيانات المبوبة :grouped data

هي البيانات التي بوبت ونظمت في جدول توزيع تكراري .

3- البيانات الغير مبوبة : ungrouped data

هي البيانات التي جمعت ولم تبوب .

تبويب البيانات :

في مايلي خطوات تكوين توزيع تكراري في حالة البيانات الكمية :

1- ترتيب البيانات : هناك طريقتان للترتيب :

أ – ترتيب البيانات تصاعديا" .

ب – ترتيب البيانات تنازليا" .

2- حساب قيمة المدى :

المدى = اكبر قيمة – اصغر قيمة

3- اختيار عدد مناسب للفئات :

حيث يفضل ان لاتزيد الفئات عن خمس عشرة فئة و لا تقل عن خمسة فئات ، حيث ان اختيار عدد اقل من خمس فئات سيؤدي الى ضياع الكثير من المعلومات ، وكذلك اختيار اكثر من خمس عشرة فئة يقلل من الوضوح في المعلومات .

ولتسجيل الفئات طرق مختلفة لعل ابسطها هو ان تجعل كل فئة لها حد ادنى واعلى حيث تبدأ محدودة وتنتهي بأقل من قيمة محدودة على ان تبدأ الفئة التالية بهذه القيمة الاخيرة .

الجدول التالي ، اذا اخذنا اول فئة تبدأ بـ 3 وتنتهي بأقل من 9 وهكذا ، يمكن ان نكتب الفئات كالتالي :

من 3 الى اقل من 9	او 3 - 8	او 3 -
من 9 الى اقل من 15	او 9 - 14	او 9 -
من 15 - اقل من 21	او 15 - 20	او 15 -
وهكـــــــذا ...		

4- حساب طول الفئة :

$$\text{طول الفئة} = \frac{\text{المدى}}{\text{عدد الفئات}}$$

ويكون من المناسب تقريب قيمة طول الفئة الى اقرب عدد صحيح يلي تلك القيمة . فمثلا" اذا كانت قيمة طول الفئة تساوي العدد 4.5 فاننا نقربها الى العدد 5 .

مثال :
البيانات التالية تمثل اطوال نبات معين في احدى الحقول الزراعية :

27	26	25	24	23	21	21	20
6	5	5	4	3	36	30	28
9	9	9	8	8	7	7	6
13	13	13	13	12	12	10	10
19	18	17	17	16	15	15	14

المطلوب :
1-رتب البيانات تصاعديا" .

2- ضع البيانات في جدول توزيع تكراري .

الحل :

1- يكون الترتيب التصاعدي على النحو التالي :

3	4	5	5	6	6	7	7	8	8
9	9	9	10	10	12	12	13	13	13
13	14	15	15	16	17	17	18	19	20
21	21	23	24	25	26	27	28	30	26

2- ثم نضع البيانات في جدول توزيع تكراري كالتالي :

اولا" : اختيار عدد مناسب للفئات وليكن 6 فئات .

ثانيا" : حساب طول الفئة :

طول الفئة = $\frac{36-3}{6} = \frac{33}{6} = 5.5$

ويقرب طول الفئة ليساوي 6 .

ثم يكون الجدول على النحو التالي :

الفئات	التكرار f_i
3 - 8	10
9 - 14	12
15 - 20	8
21 - 26	6
27 - 32	3
33 - 38	1
المجموع	40

جداول التوزيع التكراري النسبي : Relative frequency distribution

يبين الاهمية النسبية لكل فئة ويحسب التكرار النسبي لكل فئة كما يلي :

التكرار النسبي لكل فئة = $\frac{\text{تكرار تلك الفئة}}{\text{المجموع الكلي للتكرارات}} * 100 = \frac{f_i}{\sum f_i} * 100$

مثال على البيانات الكمية :

بين الاهمية النسبية لكل فئة للجدول التالي :

الفئات	التكرار f_i
3 - 8	10
9 - 14	12
15 - 20	8
21 - 26	6
27 - 32	3
33 - 38	1
المجموع	40

الحل :

لايجاد الاهمية النسبية او التكرار النسبي لكل فئة نستخدم القانون التالي لكل فئة :

$$\text{التكرار النسبي لكل فئة} = \frac{\text{تكرار تلك الفئة}}{\text{المجموع الكلي للتكرارات}} * 100 = \frac{f_i}{\sum f_i} * 100$$

$$\text{الاهمية النسبية للفئة الاولى} = \frac{\text{تكرار الفئة الاولى}}{\text{المجموع الكلي للتكرارات}} * 100 = \frac{f1}{\sum f_i} * 100$$

$$= \frac{10}{40} * 100 = 25\ \%$$

الاهمية النسبية للفئة الثانية = $\frac{\text{تكرار الفئة الثانية}}{\text{المجموع الكلي للتكرارات}} * 100 = \frac{f_2}{\sum f_i} * 100$

$$= \frac{12}{40} * 100 = 30\ \%$$

وهكذا لكل الفئات ، فيصبح الجدول كما يلي :

الفئات	التكرار f_i	التكرار النسبي
3 - 8	10	(10 / 40) * 100 = 25 %
9 - 14	12	(12 / 40) * 100 =30 %
15 - 20	8	(8 / 40) * 100 = 20 %
21 - 26	6	(6 / 40) * 100 =15 %
27 - 32	3	(3 / 40) * 100 =7.5 %
33 - 38	1	(1 / 40) * 100 = 2.5 %
المجموع	40	(40 / 40) * 100 =100%

مثال على البيانات الكمية المنفصلة :

فيما يلي بيانات عينة من 40 مزرعة عن نوع التمر الذي تنتجه تلك المزارع :

سكري خلاص برحي خلاص برحي خلاص صقعي خلاص

برحي سكري برحي صقعي خلاص برحي نبوت سيف برحي

صقعي برحي سكري خلاص برحي برحي صقعي خلاص

برحي خلاص برحي سكري نبوت سيف صقعي نبوت سيف صقعي

خلاص برحي صقعي نبوت سيف سكري برحي صقعي خلاص

المطلوب:

1- اعرض البيانات في جدول توزيع تكراري.

2-كون التوزيع التكراري النسبي.

حيث ان البيانات منفصلة فيمكن تبويبها حسب انواع التمرو ثم ايجاد الاهمية النسبية لكل نوع :

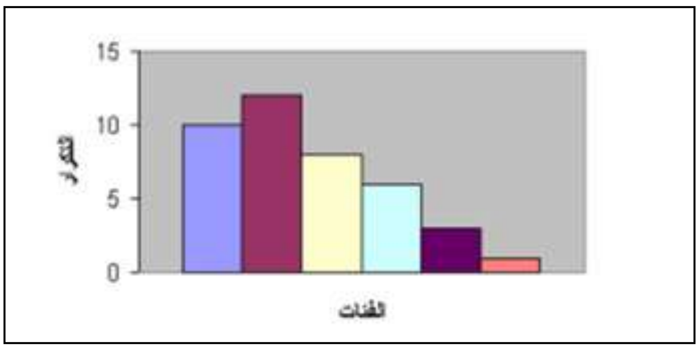
<table>
<tr><th>الصنف</th><th>العلامات</th><th>العدد</th><th>النسبة</th></tr>
<tr><td>برحي</td><td>|||| |||| |||</td><td>13</td><td>(13/40)*100 = 32.5</td></tr>
<tr><td>سكري</td><td>||||</td><td>5</td><td>(5/40)*100 = 12.5</td></tr>
<tr><td>صقعي</td><td>|||| |||</td><td>8</td><td>20</td></tr>
<tr><td>خلاص</td><td>|||| ||||</td><td>10</td><td>25</td></tr>
<tr><td>نبوت سيف</td><td>||||</td><td>4</td><td>10</td></tr>
<tr><td>المجموع</td><td></td><td>40</td><td>100</td></tr>
</table>

ثانيا" : العرض البياني

1- المدرج التكراري

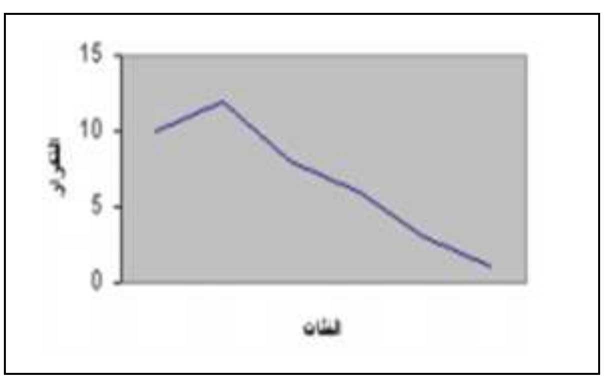
يعتبر المدرج التكراري نوعا" من الاعمدة البيانية ، ولرسم المدرج التكراري نضع حدود الفئات على المحور الافقي والتكرارات على المحور الراسي (العمودي) ، ويرسم فوق كل فئة مستطيل تمثل قاعدته طول الفئة وارتفاعه تكرار الفئة .

والشكل التالي يمثل شكل المدرج التكراري لبيانات جدول توزيع تكراري :



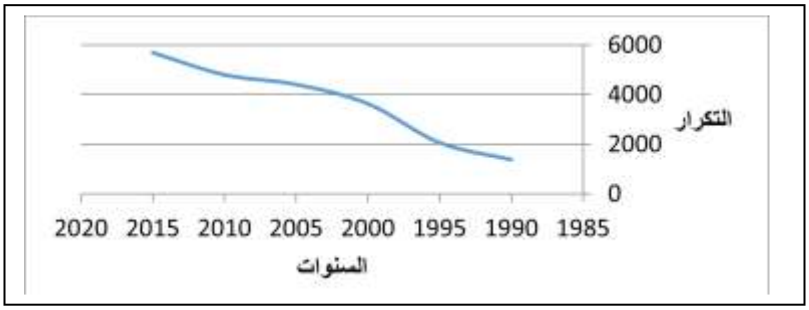
2- المضلع التكراري

يرسم المضلع التكراري بنفس طريقة عمل المدرج التكراري وذلك على محورين متعامدين ، الافقي يمثل الفئات والعمودي يمثل التكرارات ، وبدلا" من رسم مستطيلات في المدرج التكراري توضع نقطة فوق مركز الفئة ا، ارتفاعها يمثل تكرار تلك الفئة . وبعد الانتهاء من تمثيل النقط لجميع الفئات نصل بالمسطرة كل نقطتين متجاورتين فنحصل على المضلع التكراري المفقتوح ويكون شكل المضلع من بيانات جدول توزيع تكراري كالتالي :



3- المنحنى التكراري :

باتباع نفس الخطوات السابقة في رسم المضلع التكراري ولكن يتم تمهيد الخطوط المتكسرة في شكل منحني وكما في الشكل :



مقاييس النزعة المركزية

مقاييس النزعة المركزية :

يقصد بمقاييس النزعة المركزية ميل البيانات للتراكم حول قيمة ما تسمى بالمتوسط ، وهناك عدد من المقاييس لقياس هذا الميل منها :

1- الوسط الحسابي
2- الوسيط .
3- المنوال .

1- الوسط الحسابي : Mean Arithmetic

يُعد الوسط الحسابي اهم مقاييس النزعة المركزية ويُعرف بانه القيمة التي اذا أُعطيت لجميع مفردات الظاهرة كان مجموع قيم المفردات مساويا" لمجموع القيم الاصلية لها .ويسمى ايضا " بالمعدل . ويُرمز له بالرمز $\bar{x}$ او $\bar{y}$.

أ – البيانات الغير المبوبة :

يمكن حساب الوسط الحسابي للبيانات غير مبوبة حسب العلاقة التالية :

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

مثال (1) :

اوجد الوسط الحسابي للبيانات التالية :

x_i : 15 10 10 15 30

الحل :

n = 5

$$\bar{x} = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{15+10+10+15+30}{5} = \frac{80}{5} = 16$$

مثال (2) :

اوجد الوسط الحسابي للبيانات غير مبوبة التالية :

y_i : 5 3 4 3 4 5

$$\bar{y} = \frac{\sum_{i=1}^{6} y_i}{6} = \frac{5+3+4+3+4+5}{6} = \frac{24}{6} = 4$$

ب – البيانات المبوبة :

القانون التالي يمثل ايجاد الوسط الحسابي للبيانات المبوبة :

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

حيث ان :

x_i : تمثل مركز الفئة التي تسلسلها i ، ويحسب كما يلي :

مركز الفئة = (بداية الفئة + نهاية الفئة) / 2 .

f_i : التكرار عند القيمة i .

مثال : جد الوسط الحسابي لجدول التوزيع التكراري التالي :

الفئات	التكرار f_i
60- 64	5
65 -69	15
70 - 74	20
75 - 79	30
80 - 84	15
85 - 89	10
90 - 94	5
المجموع	100

الحل :

الوسط الحسابي للبيانات المبوبة هو :

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

1 - لإيجاد مركز الفئة الذي يمثل العمود الثالث للجدول ، نطبق قانون ايجاد مركز الفئة لكل الفئات وكما يلي :

مركز الفئة الاولى = (60 + 64) / 2 = 62

مركز الفئة الثانية = (65 + 69) / 2 = 67

وبنفس الطريقة نجد بقية مراكز الفئات .

2 – نضرب كل تكرار في مركز الفئة ونكون من ذلك عمود بأسم ($x_i * f_i$) .

بعد ذلك يتكون لنا الجدول التالي :

الفئات	التكرار f_i	مركز الفئة x_i	$x_i * f_i$
60- 64	5	(60+64)/2= 62	62* 5 = 310
65 -69	15	(65 + 69) / 2 = 67	67 * 15 = 1005
70 - 74	20	72	1440
75 - 79	30	77	2310
80 - 84	15	82	1230
85 - 89	10	87	870
90 - 94	5	92	460
المجموع	100		7625

3 – نجد الوسط الحسابي حسب القانون :

$$\bar{x} = \frac{7625}{100} = 76.25$$

مثال (2) :

جد الوسط الحسابي لجدول التوزيع التكراري التالي :

التكرار f_i	الفئات
5	60 - 62
18	63 - 65
42	66 - 68
27	69 - 71
8	72 - 74
100	

الحل :

1 – نجد مراكز الفئة لكل فئة .

مركز الفئة الاولى = $(60 + 62) / 2 = 61$

مركز الفئة الثانية = $(63 + 65) / 2 = 64$

مركز الفئة الثالثة = $(66 + 68) / 2 = 67$

وهكذا لبقية الفئات .

2 – نضرب كل مركز فئة في تكرارها ، ونكوّن عمود ($y_i * f_i$)

يصبح الجدول كما يلي :

الفئات	التكرار f_i	مركز الفئة y_i	$y_i * f_i$
60 - 62	5	(60+62)/2 = 61	305
63 - 65	18	64	1152
66 - 68	42	67	2814
69 - 71	27	70	1890
72 - 74	8	73	584
	100		6745

3 – نجد الوسط الحسابي بتطبيق قانون الوسط الحسابي للبيانات المبوبة :

$$\bar{y} = \frac{6745}{100} = 67.45$$

ميزات الوسط الحسابي :

1 – سهولة حسابه .

2 – مجموع انحرافات القيم عن وسطها الحسابي يساوي صفر:

1 – للبيانات غير مبوبة

$$\sum(y_i - \bar{y}) = 0$$

2 – للبيانات المبوبة :

$$\sum f_i(y_i - \bar{y}) = 0$$

مثال :

y_i	$y - \bar{y}$
5	5 - 6 = -1
7	7 - 6 = 1
8	8 - 6 = 2
4	4 - 6 = -2
$\sum y_i = 24 \rightarrow \bar{y} = \frac{24}{4} = 6$	0

عيوب الوسط الحسابي :

1 – تاثره بالقيم الشاذة .

2 – عدم امكانية حسابه في حالة البيانات الوصفية .

2 – الوسيط Median (Me)

هو القيمة التي تتوسط قيم البيانات بعد ترتيبها تصاعديا" او تنازليا" ، بحيث يكون عدد المفردات التي قبلها مساويا" لعدد المفردات التي بعدها .

أ – الوسيط في البيانات غير مبوبة :

لحساب قيمة الوسيط نرتب البيانات تصاعديا" او تنازليا"

- فاذا كان عدد المشاهدات او المفردات (n) فرديا" فيكون :

$$\text{الوسيط} = Me = y_{\frac{n+1}{2}}$$

- واذا كان عدد المشاهدات (n) زوجيا" فيكون :

$$\text{الوسيط} = Me = \frac{y_{\frac{n}{2}} + y_{\frac{n}{2}+1}}{2}$$

مثال (1) :

اوجد الوسيط للقيم التالية :

90 100 40 70 80 60 50

الحل :

1 – نحدد n

هنا n = 7 ، العدد فردي ، اي اننا نستعمل القانون التالي :

$$\text{الوسيط} = Me = y_{\frac{n+1}{2}}$$

2 – نرتب البيانات تصاعديا" او تنازليا" ، في هذا المثال سنرتب البيانات تصاعديا" :

100 90 80 70 60 50 40

3 – نطبق القانون :

$$\text{الوسيط} = Me = y_{\frac{7+1}{2}} = y_{\frac{8}{2}} = y_4 = 70$$

اي ان القيمة الرابعة هي الوسيط وتساوي 70 .

مثال (2) :

جد الوسيط للبيانات التالية :

40 50 60 70 100 80

الحل :

1 – نحدد n ، هنا n = 6 ، العدد زوجي اذن نستخدم القانون التالي :

$$\text{الوسيط} = Me = \frac{y_{\frac{n}{2}} + y_{\frac{n}{2}+1}}{2}$$

2 – نرتب البيانات اما تصاعديا" او تنازليا ، في هذا المثال سنرتب البيانات تنازليا" :

40 50 60 70 80 100

3 – نطبق القانون :

$$\text{الوسيط} = Me = \frac{y_{\frac{6}{2}} + y_{\frac{6}{2}+1}}{2} = \frac{y_3 + y_4}{2} = \frac{60+70}{2} = \frac{130}{2} = 65$$

ملاحظة : نجد في القانون y_3 و y_4 ، هذا يعني القيمة الثالثة من المشاهدات بعد ترتيبها وهي تساوي 60 والقيمة الرابعة من المشاهدات وهي تساوي 70 .

ب – الوسيط في البيانات المبوبة :

لحساب قيمة الوسيط في البيانات المبوبة نتبع الخطوات التالية :

1 - نحدد فئة الوسيط ، التي تعني ان قيمة الوسيط تقع في هذه الفئة ، وذلك من خلال القانون $y_{\frac{n+1}{2}}$.

2 – بعد تحديد فئة الوسيط ، نحدد الحد الادنى لتلك الفئة ويرمز له بالرمز L .

3 – نحدد التكرار المقابل لفئة الوسيط ويُرمز له بالرمز f_i .

4 – نكوّن عمود تكرار المتجمع الصاعد ومنه نحدد التكرار المتجمع لبداية فئة الوسيط ويرمز له بالرمز F_i.

5 – نحدد طول الفئة h .

6 – نستخدم القانون التالي لايجاد الوسيط :

$$Me = L + \frac{\frac{\sum f_i}{2} - F_i}{f_i} * h$$

مثال :

جد الوسيط للجدول التالي :

الفئات	التكرار
1 – 3	5
4 – 6	7
7 – 9	15
10 - 12	9
13 - 15	7

الحل :

1 – عدد الفئات يساوي 5 ، اذن فئة الوسيط تكون بالشكل التالي :

$$y_{\frac{n+1}{2}} = y_{\frac{5+1}{2}} = y_{\frac{6}{2}} = y_3$$

اي ان الفئة الثالثة هي الفئة التي تقع فيه قيمة الوسيط .

2 – الحد الادنى لفئة الوسيط $L = 7$.

3 – التكرار المتجمع عند بداية فئة الوسيط $F_i = 12$.

4 – تكرار فئة الوسيط $f_i = 15$.

5 – طول الفئة $h = 3$.

6 – نستخدم قانون الوسيط : وبذلك يكون الجدول كالتالي :

تكرار متجمع صاعد F_i	الحد الادنى للفئة فاقل	التكرار f_i	الفئات
0	اقل من 1	5	1 - 3
5	اقل من 4	7	4 – 6
5+7 = 12	اقل من 7	15	7 – 9
5+7+15= 27	اقل من 10	9	10 - 12
5+7+15+9 = 36	اقل من 13	7	13 - 15
5+7+15+9+7 = 43	اقل من 15	$\sum f_i = 43$	

6 – نستخدم قانون الوسيط :

$$Me = 7 + \frac{\frac{43}{2} - 12}{15} * 3 = 7 + \frac{21.5 - 12}{15} * 3 = 7 + \frac{9.5}{15} * 3$$

$$= 7 + 1.9 = 8.9$$

وهذه القيمة تقع ضمن فئة الوسيط التي تم تحديدها ، وهي الفئة الثالثة .

3 – المنوال The Mode (D) :

هو القيمة الاكثر شيوعا" (تكرارا") في البيانات .، ويرمز له بالرمز D .

أ – البيانات غير مبوبة :

اوجد المنوال للبيانات التالية : 4 5 6 7 5 8 5 3

الحل :

بما ان 5 هو الرقم الاكثر تكرارا" ، اذن المنوال D يساوي 5 ، اي ان :

$$D = 5$$

ملاحظة : قد لا يوجد منوال للبيانات ، او قد يكون للبيانات اكثر من منوال .

مثال :

جد المنوال للبيانات التالية :

أ - 51 48 50 49 47

ب - 3 7 5 3 5 6

الحل :

أ – لا يوجد منوال لانه لم تتكرر اي قيمة من القيم .

ب – المنوال هما القيمة 3 والقيمة 5 لانهما تكررتا اكثر من غيرهما ، اي ان :

$$D = 3 , 5$$

ب – البيانات المبوبة :

لايجاد المنوال في البيانات المبوبة نتبع الخطوات التالية :

1 – نحدد فئة المنوال (الفئة المنوالية) وهي التي تمتلك اعلى تكرار .

2 – نستخدم القانون التالي لإيجاد قيمة المنوال :

$$D = L + \frac{d_1}{d_1 + d_2} * h$$

حيث ان :

L : الحد الادنى للفئة المنوالية .

d_1 : الفرق بين تكرار الفئة المنوالية وتكرار الفئة السابقة لها .

d_2 : الفرق بين تكرار الفئة المنوالية وتكرار الفئة اللاحقة لها .

h : طول الفئة .

جد المنوال للجدول التالي :

الفئات	التكرار
3 – 8	10
9 – 14	12
15 - 20	8

الحل :

1 – نجد الفئة المنوالية ، بما ان 12 هو اعلى تكرار ، اذن الفئة الثانية هي التي تمثل الفئة المنوالية.

2 – نستخدم القانون لايجاد المنوال ، حيث ان :

L = 9

$d_1 = 12 - 10 = 2$

$d_2 = 12 - 8 = 4$

$h = 6$

اذن المنوال هو :

$$D = 9 + \frac{2}{2+4} * 6 = 9 + \frac{2}{6} * 6 = 9 + 2 = 11$$

المنوال يساوي 11 وهذه القيمة تقع ضمن الفئة المنوالية الثانية، والتي تم تحديدها في البداية.

مقاييس التشتت

تعريف التشتت :

يمثل التشتت مدى انحراف (تقارب او تباعد) البيانات بعضها عن بعض .

وهناك عدة مقاييس للتشتت اهمها :

اولا" : مقاييس التشتت المطلق :

أي التي تكون وحداتها القياسية نفس وحدات القيم الاصلية، واهمها :

1- المدى R.

2- التباين والانحراف المعياري S^2 , S .

ثانيا" : مقاييس التشتت النسبي :

أي التي تكون خالية من وحدات القياس . منها :

1- معامل الاختلاف C.V .

ثالثا" : الدرجة المعيارية Z .

اولا" : مقاييس التشتت المطلق :

1- المدى R : هو ابسط مقاييس التشتت ، وهو الفرق بين اعلى قيمة واقل قيمة لمجموعة من البيانات . ويُحسب المدى في حالة البيانات غير المبوبة كما يلي :

المدى = اعلى قيمة – اقل قيمة

$$R = y_{max} - y_{min}$$

مثال :

تم زراعة 9 وحدات تجريبية بمحصول القمح ، وتم تسميدها بنوع معين من الاسمدة الفسفورية ، وفيما يلي بيانات كمية الانتاج من القمح بالطن / هكتار :

5 6 5 5 5 5 5 4 5

المطلوب : حساب المدى .

الحل :

اقل قيمة = 4 ، اعلى قيمة = 6

المدى = اعلى قيمة – اقل قيمة

المدى = 6 – 4

= 2 اي ان المدى يساوي 2 طن / هكتار .

مثال :

اوجد المدى للبيانات التالية :

12 6 7 3 15 10 18 5

الحل :

اقل قيمة = 3 ، واعلى قيمة = 18

اذن المدى :

$$R = y_{max} - y_{min}$$

$$= 18 - 3$$

$$= 15$$

2-التباين والانحراف المعياري (القياسي) S^2 , S .

التباين : مجموع مربعات انحرافات القيم عن وسطها الحسابي مقسوما على درجات الحرية ، ويرمز له بالرمز S^2 . قانون التباين هو :

الصيغة الاولى :

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

الصيغة الثانية :

$$s^2 = \frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1}$$

اما الانحراف المعياري فهو الجذر التربيعي للتباين . قانون الانحراف المعياري هو :

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

مثال : اوجد الانحراف المعياري او (الانحراف القياسي) للبيانات غير المبوبة التالية :

y : 15 20 10 15 30

الحل :

1- نكتب البيانات على شكل عمودي لتسهيل الحل .

2- نجد الوسط الحسابي للبيانات $\bar{y}$.

3- في حل استخدام الصيغة الاولى للقانون ، نطرح الوسط الحسابي من كل قيمة من قيم البيانات $(y_i - \bar{y})$، مع ملاحظة ان مجموع هذا العمود يجب ان يساوي الصفر .

4- نربع كل قيمة من قيم العمود السابق $(y_i - \bar{y})^2$، ثم نجد مجوع هذه القيم ، اي ايجاد مجموع هذا العمود .

5- نطبق قانون التباين

6- نطبق قانون الانحراف المعياري .

y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
15	15 – 18 = -3	$(-3)^2 = 9$
10	10 – 18 = -8	$(-8)^2 = 64$
15	15 – 18 = -3	$(-3)^2 = 9$
20	20 – 18 = 2	$(2)^2 = 4$
30	30 – 18 = 12	$(12)^2 = 144$
$\bar{y} = \frac{\sum_{i=1}^{5} y_i}{n} = \frac{90}{5} = 18$	$\sum (y_i - \bar{y}) = 0$	230

نطبق قانون التباين :

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^{5}(y_i - \bar{y})^2}{5-1} = \frac{230}{4} = 57.5$$

اذن الانحراف المعياري هو :

$$s = \sqrt{57.5} = 7.58$$

اما اذا تم استخدام الصيغة الثانية للتباين فيكون الحل كما في الجدول التالي :

y_i	y_i^2
15	$(15)^2 = 225$
10	$(10)^2 = 100$
15	$(15)^2 = 225$
20	$(20)^2 = 400$
30	$(30)^2 = 900$
$\sum_{i=1}^{5} y_i = 90$	$\sum y_i^2 = 1850$

$$s^2 = \frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1} = \frac{1850 - \frac{(90)^2}{5}}{5-1} = \frac{1850 - 1620}{4} =$$

$$s^2 = \frac{230}{4} = 57.5$$

وهذه القيمة هي نفس القيمة التي حصلنا عليها عند تطبيق الصيغة الاولى للتباين .

وعند اخذ الجذر التربيعي نحصل على الانحراف المعياري .

مثال :

احسب الانحراف المعياري للبيانات التالية بتطبيق الصيغتين للتباين :

x : 5 7 8 8

الحل :

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	x_i^2
5	-2	4	25
7	0	0	49
8	1	1	64
8	1	1	64
$\bar{x} = \frac{\sum_{i=1}^{4} x_i}{n} = \frac{28}{4} = 7$	0	6	$\sum x_i^2 = 202$

بتطبيق الصيغة الاولى للتباين ، نحصل على :

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{5}(x_i - \bar{x})^2}{4-1} = \frac{6}{3} = 2$$

وبتطبيق الصيغة الثانية ، نحصل على :

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{202 - \frac{(28)^2}{4}}{3} = \frac{202 - 196}{3} = \frac{6}{3} = 2$$

اذن الانحراف المعياري هو :

$$s = \sqrt{2} = 1.414$$

ثانيا" : مقاييس التشتت النسبي :

من مقاييس التشتت النسبي هو معامل الاختلاف ، حيث يُستخدم معامل الاختلاف في المقارنة بين تشتت مجموعتين او اكثر من البيانات . يرمز لمعامل الاختلاف C.V ، ويُحسب حسب القانون التالي :

$$C.V = \frac{s}{\bar{y}} * 100$$

حيث ان :

s : يمثل الانحراف المعياري للظاهرة .

$\bar{y}$: يمثل الوسط الحسابي للظاهرة .

والظاهرة التي معامل اختلافها اكبر تكون اكثر تشتتا" من الاخرى .

مثال :

تم اختيار مجموعتين من الاغنام النامية في احد المزارع وتم استخدام عليقة لتسمين المجموعة الاولى ، بينما تم استخدام عليقة اخرى لتسمين المجموعة الثانية ، وبعد فترة زمنية تم جمع بيانات عن اوزان المجموعتين بالكيلوغرام ، وتم الحصول على المقاييس التالية :

المقاييس	المجموعة الاولى	المجموعة الثانية
الوسط الحسابي $\bar{y}$	173	198
الانحراف المعياري s	23	25

المطلوب :

مقارنة درجة تشتت المجموعتين .

الحل :

1 – نجد معامل الاختلاف النسبي لكل مجموعة .

معامل الاختلاف النسبي للمجموعة الاولى :

$$C.V1 = \frac{s}{\bar{y}} * 100 = \frac{23}{173} * 100 = 13.3\ \%$$

معامل الاختلاف النسبي للمجموعة الثانية :

$$C.V2 = \frac{s}{\bar{y}} * 100 = \frac{25}{195} * 100 = 12.8\ \%$$

يُلاحظ ان درجة تشتت اوزان المجموعة الثانية اقل من درجة تشتت اوزان المجموعة الاولى .

مثال :

لديك البيانات التالية :

y_i : 5 7 8 8

جد :

1 - الوسط الحسابي

2 – الانحراف المعياري .

3 – معامل الاختلاف .

الحل :

$$\bar{y} = \frac{\sum y_i}{n} = \frac{5+7+8+8}{4} = \frac{28}{4} = 7$$

$$s^2 = \frac{\sum(\, y_i - \bar{y}\,)^2}{n-1} = \frac{6}{3} = 2$$

$$\therefore \quad s = \sqrt{s^2} = \sqrt{2} = 1.414$$

$$c.v = \frac{s}{\bar{y}} * 100 = \frac{1.414}{7} * 100 = 20.2\%$$

ثالثا" : الدرجة المعيارية (الدرجة القياسية) Z :

في كثير من الاحيان نحتاج الى مقارنة مفردتين من مجموعتين مختلفتين ، في هذه الحالة يجب تحويل وحدات كل مفردة الى وحدات قياسية حتى تكون المقارنة ذات معنى وذلك باستخدام الوسط الحسابي والانحراف القياسي لكل مجموعة .

تسمى القيمة Z_i درجة قياسية اذا كانت تساوي :

$$Z_i = \frac{y_i - \bar{y}}{s}$$

والدرجات القياسية خالية من الوحدات المستخدمة في القياس ويمكن ان تكون قيمها موجبة او سالبة . هذا واذا حولنا جميع قيم المجموعة الى درجات قياسية فان الوسط الحسابي لهذه الدرجات القياسية يساوي صفر، وان تباينها يساوي 1 .

مثال :

تم اختيار مجموعتين من الاغنام النامية في احد المزارع وتم استخدام عليقة لتسمين المجموعة الاولى ، بينما تم استخدام عليقة اخرى لتسمين المجموعة الثانية ، وبعد فترة زمنية تم جمع بيانات عن اوزان المجموعتين بالكيلوغرام ، وتم الحصول على المقاييس التالية :

المقاييس	المجموعة الاولى	المجموعة الثانية
الوسط الحسابي $\bar{y}$	173	198
الانحراف المعياري s	23	25

تم اختيار احد الاغنام من المجموعة الاولى بعد تطبيق برنامج التسميد ، ووجد ان وزنه 178 كيلوغرام ، وتم اختيار احد لاغنام من المجموعة الثانية ، ووجد ان وزنه 180 كيلوغرام ، قارن بين هاتين القيمتين من حيث اهمية كل منها في المجموعة التي ينتمي اليها.

الحل :

عند مقارنة الوزن الاول مع الوزن الثاني للاغنام ، نجد ان وزن الخروف الثاني اكبر من وزن الخروف الثاني ، ولكن عند تحويل الوزنين حسب الدرجة المعيارية نجد ان :

$$Z_1 = \frac{y_1 - \bar{y}_1}{s_1} = \frac{178 - 173}{23} = \frac{5}{23} = 0.22$$

$$Z_2 = \frac{y_2 - \bar{y}_2}{s_2} = \frac{180 - 198}{25} = \frac{-18}{23} = -0.75$$

من مقارنة Z_1 و Z_2 حيث ان $Z_2 < Z_1$ ، يتضح بأن وزن الخروف الاول افضل من وزن الخروف الثاني وهو عكس ما توصلت اليه المقارنة السابقة .

مثال :

حصل طالب على درجة 84 في الامتحان النهائي بمادة الوراثة علما" بان الوسط الحسابي في امتحان مادة الوراثة لجميع الطلاب كان 76 وبانحراف قياسي قدره 10 ، اما في امتحان مادة التغذية فقد حصل نفس الطالب على درجة 90 ، حيث كان الوسط الحسابي في امتحان مادة التغذية لجميع الطلاب يساوي 82 والانحراف القياسي يساوي 16 ، ففي أي الموضوعين كانت قابلية هذا الطالب اعلى ؟

الحل :

عند مقارنة درجة الامتحانين نجد ان درجته في التغذية 90 اعلى من درجته في الوراثة 84 ولكن عند تحويل هاتين الدرجتين الى درجات قياسية نجد ان :

$$Z_1 = \frac{y_1 - \bar{y}_1}{s_1} = \frac{84 - 76}{10} = \frac{8}{10} = 0.8$$

$$Z_2 = \frac{y_2 - \bar{y}_2}{s_2} = \frac{90 - 82}{16} = \frac{8}{16} = 0.5$$

ومن هذا يتضح ان قابليته في مادة الوراثة اعلى من قابليته في مادة التغذية .

الارتبــــــاط

Correlation

الارتباط الخطي البسيط :

اذا كان الغرض من التحليل هو تحديد نوع وقوة العلاقة بين متغيرين ، يُستخدم الارتباط ، اما اذا كان الغرض من تحليل اثر احد المتغيرين على الاخر ، يستخدم الانحدار ، وفي هذا الفصل سوف يتم عرض اسلوب تحليل الانحدار الخطي البسيط ، اي في حالة افتراض ان العلاقة بين المتغيرين تأخذ الشكل الخطي ، وسوف نجري حسابه في حالة البيانات الكمية .

الارتبــاط : من محاضرات الاول – احصاء

يقصــــد بالارتباط بانه قيمة العلاقة وطبيعتها بين المتغيرين y, x او اكثر ويقال بان المتغيرين y,x مرتبطين ببعضــــهما اذا كان التغير في احد هذين المتغيرين يؤدي الى احداث تغير في المتغير الاخر واذا كان تغير الظاهرتين في نفس الاتجاه ســواء بالزيادة او النقصــان عندئذ يقال بان الارتباط ما بين المتغيرين موجب ، اما اذا كان تغيرهما بشــــكل معاكس أي ان الزيادة في احدهما يؤدي الى انخفاض في الاخر او النقص في احدهما يؤدي الى الزيادة في الاخر عندئذ يقال ان الارتباط بين هذين المتغيرين هو ارتباط عكســي او ارتباط ســالب كما في الشكل ، مثلا" الارتباط بين طول الشخص ووزنه ، يفترض ان يكون موجب لانه من المتوقع ان يزداد وزن الشــخص بزيادة طوله ، في حين ان الارتباط بين ســعر ســلعة معينة والكمية المطلوبة من هذه الســلعة يفترض ان يكون ســالبا" لانه من المتوقع عند انخفاض ســعر هذه

الســـلعة ان يزداد الطلب عليها والعكس صـــحيح ،وقد لا يكون هناك اي علاقة بين المتغيرين ، مثل العلاقة بين طول الطالب و درجة الطالب في مادة دراسية معينة .

معامل الارتباط :

يرمز لمعامل الارتباط بالرمز r ، هو عبارة عن مقياس كمي نســـبي يقيس قوة الارتباط بين متغيرين ، حيث تتراوح قيمته بين (1+ و 1-) اي ان :

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

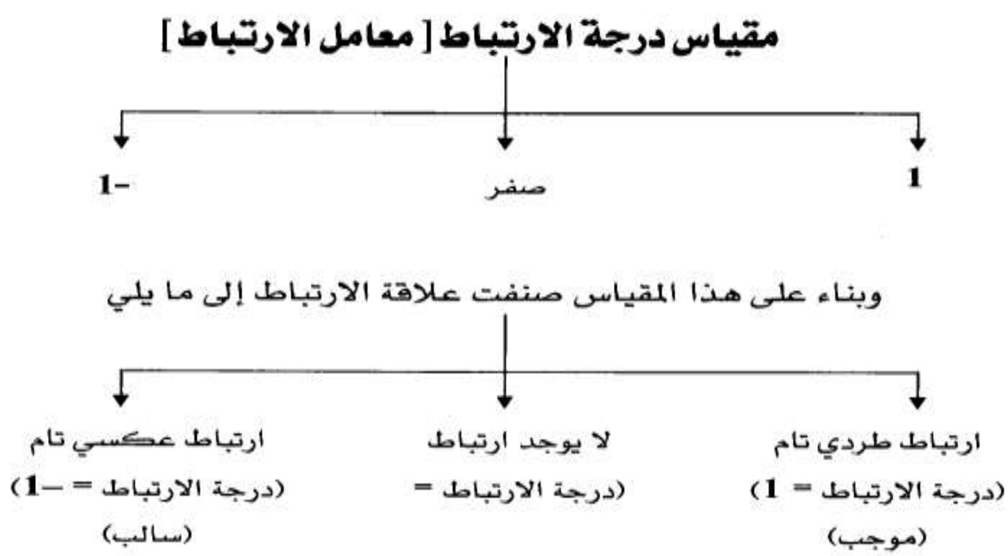
يُقال ان :

1 - الارتباط طردي تام اذا كان معامل الارتباط = 1+ اي ان : $r = +1$

2 – الارتباط عكسي تام اذا كان معامل الارتباط = 1 – اي ان : $r = -1$

3 – المتغيرين مستقلان اذا كان معامل الارتباط = 0 اي ان : $r = 0$

ويمكن توضيح ذلك بالمخطط التالي :



أما قوة درجة الارتباط فإنها تصنف وفق الأصناف التالية

- قوي طردي (بين $\frac{1}{2}$ و 1)
- ضعيف بين ($-\frac{1}{2}$ و $\frac{1}{2}$)
- قوي عكسي بين (-1 ، $-\frac{1}{2}$)

ملاحظة هامة: تزداد قوة الارتباط كلما اقترينا من الأطراف وتقل كلما ابتعدنا عن الأطراف.

لحساب معامل الارتباط r ، تُستخدم صيغة بيرسون التالية :

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

او الصيغة التالية :

$$r = \frac{\sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}\sqrt{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}}$$

مثال :

لديك المتغير x يمثل عرض الورقة لنبات معين ، والمتغير y والذي يمثل طول ورقة النبات المعين وحسب ما يلي :

x	13	19	13	18	14	17	14	17	15	16
y	15	22	13	20	13	20	15	19	15	18

المطلوب :

1 – حساب معامل ارتباط بيرسون بين عرض الورقة x و طول الورقة y للنبات .

2 – هل توجد علاقة ارتباط خطية ؟

3 – مانوع هذه العلاقة (موجبة ام سالبة) ؟ وما مدى قوتها ؟

الحل :

سوف نحسب معامل ارتباط بيرسون حسب الصيغتين المذكورتين اعلاه .

اولا" : الصيغة الاولى

1 – نحسب الوسط الحسابي لكل من x و y :

$$\bar{y} = \frac{\sum_{i=1}^{10} y_i}{n} \ , \bar{x} = \frac{\sum_{i=1}^{10} x_i}{n}$$

2 – نطرح الوسط الحسابي لكل متغير من قيم المتغير ، اي نحسب $(x_i - \bar{x})$ و $(y_i - \bar{y})$.

3 – يجب ان تكون مجاميع النقطة الثانية تساوي صفر .

4 – نحسب $(x_i - \bar{x})^2$ و $(y_i - \bar{y})^2$

5 – نطبق قانون الصيغة الاولى .، او قانون الصيغة الثانية .

- تطبيق القانون حسب الصيغة الاولى :

لتطبيق الصيغة الاولى نحتاج الى البيانات الظاهرة في الجدول التالي :

x	y	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
13	15	-2.6	6.76	-2	4	5.2
19	22	3.4	11.56	5	25	17
13	13	-2.6	6.76	-4	16	10.4
18	20	2.4	5.76	3	9	7.2
14	13	-1.6	2.56	-4	16	6.4
17	20	1.4	1.96	3	9	4.2
14	15	-1.6	2.56	-2	4	3.2
17	19	1.4	1.96	2	4	2.8
15	15	-0.6	0.36	-2	4	1.2
16	18	0.4	0.16	1	1	0.4
$\bar{x}$ =15.6	$\bar{y}$= 17	0	40.4	0	92	58

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

$$r = \frac{58}{\sqrt{40.4}\sqrt{92}} = \frac{58}{(6.35)(9.59)} = \frac{58}{60.89} = 0.95$$

- لتطبيق الصيغة الثانية ، نحتاج الى البيانات المذكورة في الجدول التالي :

	x	y	x^2	y^2	x*y
	13	15	169	225	195
	19	22	361	484	418
	13	13	169	169	169
	18	20	324	400	360
	14	13	196	169	182
	17	20	289	400	340
	14	15	196	225	210
	17	19	289	361	323
	15	15	225	225	225
	16	18	256	324	288
$\sum$	156	170	2474	2982	2710

$$r = \frac{\sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}\sqrt{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}}$$

$$= \frac{2710 - \frac{(156)(170)}{10}}{\sqrt{2474 - \frac{(156)^2}{10}}\sqrt{2982 - \frac{(170)^2}{10}}}$$

$$r = 0.95$$

التحليل :

بما ان قيمة r تساوي 0.95 ، هذا يعني ان الارتباط بين طول الورقة وعرضها يكون :

1- موجب طردي لان القيمة موجبة ، كلما زاد الطول زاد العرض .

2- بما ان القيمة اكبر من 0.5 لذلك فان الارتباط قوي .

معامل سبيرمان لارتباط الرتب :

Spearman rank Correlation Coefficient

لحساب معامل سبيرمان لارتباط الرتب يقوم بترتيب كل من المتغيرين ترتيباً تصاعدياً أو تنازلياً (أما تصاعدياً لكلا المتغيرين أو تنازلياً لكليهما). وفي حالة الترتيب التصاعدي تأخذ أقل قيمة من قيم المتغير الرتبة رقم 1، والقيمة الأعلى منها مباشرة الرتبة رقم 2 وهكذا (بالنسبة لكل من المتغيرين). أما في حالة الترتيب التنازلي تأخذ أكبر قيمة من قيم المتغير الرتبة رقم 1، والقيمة الأقل منها مباشرة الرتبة رقم 2 وهكذا (بالنسبة لكل من المتغيرين). وعند تساوي قيمتين (أو أكثر) من قيم المتغير نعطي كل قيمة رتبة مختلفة (كما لو كانت القيم غير متساوية) ثم نحسب متوسط هذه الرتب، ويعطى هذا المتوسط لكل من هذه القيم المتساوية.

وبعد ترتيب المتغيرين نحسب الفروق بين رتب كل من المتغيرين (ونرمز للفروق بالرمز d) ثم نقوم بتربيع هذه الفروق ونحصل على مجموعها أي نحصل على $\sum d^2$ ثم نعوض في معامل سبيرمان لارتباط الرتب والذي يأخذ الشكل التالي :

$$r_s = 1 - \frac{6(\sum d^2)}{n(n^2-1)}$$

حيث ان :

$\sum d^2$ هو مجموع مربعات الفروق بين رتب المتغيرين.

n هي عدد أزواج القيم.

مما سبق نستطيع إجمال بعضاً من الملاحظات فيما يلي :

1 – مجموع الفروق بين الرتب يساوي صفر.

2 – أن قيمة معامل ارتباط الرتب تتحصر بين – 1، + 1 .

3 – كذلك نلاحظ أن مجموع الرتب لكل من المتغيرين تساوي $\frac{n(n+1)}{2}$.

مثال

البيانات التالية تمثل تقديرات سبعة طلاب في مادتي الاحصاء والرياضيات :

مادة الاحصاء	جيد	مقبول	ممتاز	جيد	جيد جداً	مقبول	جيد
مادة الرياضيات	جيد جداً	مقبول	جيد جداً	جيد	جيد	جيد	ممتاز

والمطلوب حساب معامل سبيرمان لارتباط الرتب بين هاتين المادتين ؟

الحل :

تنظم الحل في الجدول التالي مع ملاحظة ما يلي :

1 – بالنسبة لمادة الاحصاء، فإن التقدير الأعلى سيحصل على الرتبة رقم 1 والأقل منه مباشرة سيحصل على الرتبة رقم 2 وهكذا.. أي أن الترتيب تنازلي. ونكرر العمل نفسه مع مادة الرياضيات وكما يلي :

الاحصاء x : مقبول مقبول جيد جيد جيد جيد جدا ممتاز

1 2 3 4 5 6 7

2+1 5+4+3

2 – عند حصول المادتين أو أكثر على التقدير نفسه نعطي لكل تقدير مبدئياً رتبة كما لو كانوا مختلفين ثم نحسب متوسط هذه الرتب، وهذا المتوسط هو الذي يعطى لكل تقدير.

3 – ثم نحسب الفروق بين رتب المادتين ونرمز لها بالرمز d ثم نربع هذه الفروق فنحصل على d^2 ونعوض في القانون عن $\sum d^2$ مع ملاحظة أن n = 7.

d^2 مربعات الفرق	d الفرق بين الرتب	رتب Y	رتب X	مادة الرياضيات Y	مادة الاحصاء X
2.25	1.5	2.5	4	جيد جداً	جيد
0.25	- 0.05	7	6.5	مقبول	مقبول
2.25	- 1.5	2.5	1	جيد جداً	ممتاز
1.00	- 1.0	5	4	جيد	جيد
9.00	- 3.0	5	2	جيد	جيد جداً
2.25	1.5	5	6.5	جيد	مقبول
9.00	3.0	1	4	ممتاز	جيد
26.0	Zero				المجموع

$$r_s = 1 - \frac{6(Sd^2)}{n(n^2-1)} = 1 - \frac{6(\sum d^2)}{n(n^2-1)} = 1 - \frac{6(26)}{7(49-1)}$$

$$= 1 - \frac{156}{336} = 1 - 0.46$$

$$r_s = 0.54$$

وهذا يعني أن الارتباط بين إجابات المتقديرات المادتين ارتباط طردي قوي .

ارتباط الصفات

توجد بيانات وصفية أحيانا لها صفة مميزه لا نستطيع ترتيبها كالحالة الاجتماعية ، ولقياس العلاقة لهذه البيانات نشأت الحاجة إلى إيجاد مقاييس تقيس الارتباط بين هذه الصفا ت وسنذكر منها (معامل الاقتران ومعامل التوافق).

أولا : معامل الاقتران Coefficient of Association : C.A

ويستخدم لقياس قوة الارتباط عندما يكون لدينا متغيرين نوعيين ولكل متغير صفتين فقط إي يكون لدينا جدول 2*2 ، كدراسة العلاقة بين الإصابة بمرض معين (مصاب ، غير مصاب ، والتدخين (مدخن ، غير مدخن) لمجموعة من المدخنين ، ولا يستخدم في المتغيرات التي تمتلك أكثر من صفتين وتكون قيمته محصورة بين الصفر والواحد الصحيح تقسم المفردات حسب المتغيرات والصفات كالتالي :

المجموع	القيمة (X2)	القيمة (X1)	المتغير (X) المتغير (Y)
A+B	B	A	القيمة (Y1)
C+D	D	C	القيمة (Y2)
A+B+C+D=N	B+D	A+C	المجموع

ويقال أن العلاقة قوية كلما اقتربت قيمة معامل الاقتران من الواحد وبشكل عام يقال أن هناك علاقة عندما تكون قيمة C.A أكبر من 0.5 .

مثال :

لمعرفة اثر ادوية معينة في الشفاء من احد الامراض ، أُجرِيَت تجربة على عينة عشوائية من الافراد قوامها 350 فرد وكانت نتائج التجربة كالاتي :

مجموع	X2	X1	المتغير (X) العلاج المتغير (Y)- النتيجة
230	30 (B)	(A) 200	شفوا من المرض
120	100 (D)	(C) 20	لم يشفوا من المرض
350	130	220	مجموع

$$C.A = \frac{AD - BC}{AD + BC}$$

$$C.A = \frac{(200)(100) - (30)(20)}{(200)(100) + (30)(20)} = \frac{20000 - 600}{20000 + 600} = \frac{19400}{20600} = \frac{194}{206} = 0.941$$

وتدل هذه على علاقة ارتباط قوية بين العلاج والشـــفاء من المرض ، كما أن إشـــارة معامل الاقتران ليس لها أي مدلول فهي تختلف حســب ترتيب تكرار الصــفات في الجدول لذلك تؤخذ قيمته المطلقة للدلالة على وجود العلاقة من عدمها ، وتكون قيمة هذا المعامل دائما أقل من الواحد الصحيح.

ثانيا : معامل التوافق Coefficient of Contingency : C.C

في حالة انقسام كل من الظاهرتين X , Y الى اكثر من صفتين ، عندئذ لا يمكن استخدام معامل الاقتران في حســـاب درجة العلاقة بينهما وانما يُســـتخدم في ذلك مقياس اخر يدعى بمعامل التوافق والي يُحسب وفق الصيغة التالية :

$$C.C = \sqrt{\frac{K - 1}{K}}$$

حيث ان :

K : تمثل المجموع الكلي لناتج قسمة مربع التكرار في كل خلية على حاصل ضرب مجموعي التكرارين العمودي والافقي لتلك الخلية التي فيها التكرار .

مثال :

عند دراســـــة العلاقة بين الرائحة ولون الزهرة لعينة مكونة من 30 زهرة ، كانت لدينا النتائج التالية :

المتغير (X) اللون / المتغير (Y) الرائحة	X1 اصفر	X2 ابيض	X3 احمر	مجموع	
بدون رائحة Y1	6	7	6	19	T1.
لها رائحة Y2	4	2	5	11	T2.
مجموع	10	9	11	30	

T.1 T.2 T.3

المطلوب :

احسب معامل التوافق C.C بين اللون و رائحة الزهور .

الحل :

$$K_i = \sum \frac{K_{ij}^2}{(Ti.)(T.j)}$$

$$K_1 = \left[\frac{6^2}{(10)(19)} + \frac{7^2}{(9)(19)} + \frac{6^2}{(11)(19)}\right] = \frac{1}{19}\left[\frac{6^2}{10} + \frac{7^2}{9} + \frac{6^2}{11}\right] = 0.675$$

$$K_2 = \left[\frac{4^2}{(10)(11)} + \frac{2^2}{(9)(11)} + \frac{5^2}{(11)(11)}\right] = \frac{1}{11}\left[\frac{4^2}{10} + \frac{2^2}{9} + \frac{5^2}{11}\right] = 0.378$$

K= K1 + K2 = 0.675 + 0.378

K = 1.05

$$C.C = \sqrt{\frac{K-1}{K}} = \sqrt{\frac{1.05-1}{1.05}} = 0.22$$

بما ان قيمة معامل التوافق اقل من 0.5 ، اذن قوة الارتباط بين اللون والرائحة ضعيفة .

مثال :

المتغير (X) اللون / المتغير (Y) الرائحة	X1 اصطدام	X2 انقلاب	X3 انحراف	X4 احتراق	مجموع	
Y1صحو	5	7	4	2	18	T1.
Y2 ممطر	15	20	26	1	62	T2.
Y1 ضباب	12	25	30	1	68	T3.
Y1 مترب	10	18	32	2	62	T4.
مجموع	42	70	92	6	210	
	T.1	T.2	T.3	T.4		

$$K_i = \sum \frac{K_{ij}^2}{(Ti.)(T.j)}$$

$$K_1 = \left[\frac{5^2}{(18)(42)} + \frac{7^2}{(18)(70)} + \frac{4^2}{(18)(92)} + \frac{2^2}{(18)(6)}\right]$$

$$= \frac{1}{18}\left[\frac{6^2}{42} + \frac{7^2}{70} + \frac{6^2}{92} + \frac{2^2}{6}\right] = 0.1187$$

$$K_2 = \left[\frac{15^2}{(62)(42)} + \frac{20^2}{(62)(70)} + \frac{26^2}{(62)(92)} + \frac{1^2}{(62)(6)}\right]$$

$$= \frac{1}{62}\left[\frac{6^2}{42} + \frac{7^2}{70} + \frac{6^2}{92} + \frac{2^2}{6}\right] = 0.2298$$

$$K_3 = \left[\frac{12^2}{(68)(42)} + \frac{25^2}{(68)(70)} + \frac{30^2}{(68)(92)} + \frac{1^2}{(68)(6)}\right]$$

$$= \frac{1}{68}\left[\frac{12^2}{42} + \frac{25^2}{70} + \frac{30^2}{92} + \frac{1^2}{6}\right] = 0.3280$$

$$K_4 = \left[\frac{10^2}{(62)(42)} + \frac{18^2}{(62)(70)} + \frac{32^2}{(62)(92)} + \frac{2^2}{(62)(6)}\right]$$

$$= \frac{1}{62}\left[\frac{10^2}{42} + \frac{18^2}{70} + \frac{32^2}{92} + \frac{2^2}{6}\right] = 0.3033$$

K= K1 + K2+ K3 + K4 = 1.0498

$$C.C = \sqrt{\frac{K-1}{K}} = \sqrt{\frac{1.0498-1}{1.0498}} = 0.2178$$

***الانحدار الخطي البسيط Simple Linear Regression:**

إن الغرض من استخدام اسلوب تحليل الانحدار الخطي البسيط هو دراسة وتحليل اثر متغير كمي على متغير كمي اخر ، ومن الامثلة على ذلك:

-دراسة اثر كمية السماد على انتاجية الدونم.

-دراسة اثر الانتاج على التكلفة.

-دراسة اثر كمية البروتين التي يتناولها الابقار على الزيادة في الوزن.

-اثر الدخل على الانفاق الاستهلاكي.

وهكذا هناك امثلة في كثير من النواحي الاقتصادية ،والزراعية،والتجارية،وغيرها من المجالات.

***نموذج الانحدار الخطي البسيط**

في تحليل الانحدار الخطي البسيط ، نجد ان الباحث يهتم بدراسة اثر احد المتغيرين ويسمى بالمتغير المستقل او المتنبأ منه ، على المتغير الثاني ويسمى بالمتغير التابع او المتنبأ به.

ومن ثم يمكن عرض نموذج الانحدار الخطي في شكل معادلة خطية من الدرجة الاولى تعكس المتغير التابع كدالة في المتغير المستقلكما يلي:-

$$\boldsymbol{y_i = \beta_0 + \beta_1 x_i + e_i}$$

y:هو المتغير التابع(الذي يتأثر).

x:هو المتغير المستقل (الذي يؤثر).

β_0 :هو الجزء المقطوع من المحور الرأسي y وهو يعكس قيمة المتغير التابع في حالة انعدام قيمة المتغير المستقل x اي في حالة x=0.

β_1 :ميل الخط المستقيم $(\beta_0 + \beta_1 x)$ ويعكس مقدار التغير في y اذا تغيرت x بوحدة واحدة.

e:هو الخطأ العشـــوائي والذييعبر عن الفرق بين القيمة الفعلية y والقيمة المقدرة $(\hat{y} = \beta_0 + \beta_1 x)$
اي ان $e = y - (\beta_0 + \beta_1 x)$.

***تقدير نموذج الانحدار الخطي البسيط:**

يمكن تقدير معاملات الانحدار (β_0, β_1) **في الانموذج باستخدام طريقة المربعات الصغرى ، وهذا التقدير هو الذي يجعل مجموع مربعات الاخطاء العشـــوائية** $\sum e^2 = \sum(y - (\beta_0 + \beta_1 x))^2$ **اقل ما يمكن ويحسب هذا التقدير بالمعادلة التالية:-**

$$\hat{\beta}_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

حيث ان :$\bar{x}$ **هو الوسـط الحسـابي لقيم x و** $\bar{y}$ **هو الوسـط الحسـابي لقيم y وتكون القيمة المقدرة للمتغير التابع هو** $(\hat{y} = \beta_0 + \beta_1 x)$ **ويطلق على هذا التقدير تقدير معادلة الانحدار y على x.**

مثال :فيما يلي بيانات عن كمية البروتين اليومي بالغرام التي يحتاجها العجل الرضـــيع ومقدار الزيادة في وزن العجل بالكيلو غرام وذلك لعينة من العجول الرضيعة حجمها 10 .

كمية البروتين	10	11	14	15	20	25	46	50	59	70
الزيادة في الوزن	10	10	12	12	13	13	19	15	16	20

المطلوب:

1–ارسم شكل الانتشار وماهو توقعاتك لشكل العلاقة.

2–قدر معادلة انحدار الوزن على كمية البروتين.

3–فسر معادلة الانحدار.

4–ماهو مقدار الزيادة في الوزن عند اعطاء العجل (50) غرام من البروتين؟ وماهو مقدار الخطأ العشوائي؟

الاحصاءات الحيويـــــة

الاحصاءات الحيوية : هي مجموعة الاحداث التي تصيب الانسان منذ ولادته وحتى وفاته ، وهي تشمل المواليد والوفيات والزواج والطلاق والهجرة واحصاءات الامراض واسبابها .

وتخدم الاحصاءات الحيوية عدة اغراض اهمها :

1- التخطيط في المجالات التعليمية والصحية والاقتصادية والاجتماعية .
2- تنظيم وتحسين الخدمات العامة والخاصة .
3- قياس المستوى العلمي والثقافي للمجتمع .
4- البحث العلمي بجميع فروعه مثل البيئة ، الطب ، الاجتماع والتعليم ... الخ .
5- المقارنات المحلية والدولية .

تنقسم الاحصاءات الحيوية الى قسمين :

1- احصاءات الوفيات .
2- احصاءات الخصوبة .

1- احصاءات الوفيات

تشمل احصاءات الوفيات :

أ –معدل وفيات الاطفال الرضع (الاقل من سنة واكثر من شهر) .

$$\text{معدل الوفيات الاطفال الرضع} = \frac{\text{عدد وفيات الاطفال الرضع}}{\text{عدد الاطفال المولودين احياء في نفس العام}} * 1000$$

ب – معدل الوفاة الخام (العام) .

$$\text{معدل الوفيات الخام (العام)} = \frac{\text{عدد الوفيات المسجلة خلال العام}}{\text{عدد السكان في منتصف العام}} * 1000$$

جـ - معدل وفيات الاطفال حديثي الولادة (من الولادة وحتى 28 يوم او شهر) .

$$\text{معدل وفاة حديثي الولادة} = \frac{\text{عدد وفيات حديثي الولادة}}{\text{عدد المواليد الاحياء}} * 1000$$

د- معدل المواليد الموتى .

$$\text{معدل المواليد الموتى} = \frac{\text{عدد المواليد الموتى}}{\text{عدد المواليد الاحياء}} * 1000$$

هـ - معدل وفيات الامومة .

$$\text{معدل وفيات الامومة} = \frac{\text{عدد وفيات النساء بسبب الحمل الولادة}}{\text{عدد المواليد الاحياء}} * 1000$$

مثال :

اذا كان عدد الوفيات في بلد ما سنة 1980 يساوي 30000 نسمة ، فاذا علمت ان عدد السكان في منتصف السنة يساوي 20000000 نسمة ، جد معدل الوفاة الخام .

الحل :

$$\text{معدل الوفيات الخام (العام)} = \frac{\text{عدد الوفيات المسجلة خلال العام}}{\text{عدد السكان في منتصف العام}} * 1000$$

$$\text{معدل الوفيات الخام(العام)} = \frac{3000}{20000000} * 1000$$

$$= 1.5/\text{الف} \sim 2/\text{الف}$$

مثال :

اذا كان عدد المواليد الاحياء 225000 طفل وعدد المواليد الموتى 7500 طفل وعدد الوفيات الاقل من سنة يساوي 4000 طفل ، منهم 250 طفل حديثي الولادة ، جد ما يلي :

1- معدل المواليد الموتى .

2- معدل وفيات الاطفال الرضع .

3- معدل وفيات الاطفال حديثي الولادة .

الحل:

1-معدل المواليد الموتى

$$\text{معدل المواليد الموتى} = \frac{\text{عدد المواليد الموتى}}{\text{عدد المواليد الاحياء}} * \mathbf{1000}$$

$$\text{معدل المواليد الموتى} = \frac{\mathbf{7500}}{\mathbf{225000}} * \mathbf{1000}$$

$$= \mathbf{33.3}/\text{الف} \sim \mathbf{33}/\text{الف}$$

2- معدل وفيات الاطفال الرضع .

$$\text{معدل الوفيات الاطفال الرضع} = \frac{\text{عدد وفيات الاطفال الرضع}}{\text{عدد الاطفال المولودين احياء في نفس العام}} * \mathbf{1000}$$

لان عدد الاطفال الاقل من سنة يساوي 4000 ، منهم 250 طفل حديثي الولادة اذن :
الاطفال الرضع = 400 – 250 = 3750

$$\text{معدل الوفيات الاطفال الرضع} = \frac{\mathbf{3750}}{\mathbf{225000}} * \mathbf{1000}$$

$$= 16.7/\text{الف} \sim 17/\text{الف}$$

3-- معدل وفيات الاطفال حديثي الولادة .

$$\text{معدل وفاة حديثي الولادة} = \frac{\text{عدد وفيات حديثي الولادة}}{\text{عدد المواليد الاحياء}} * 1000$$

$$\text{معدل وفاة حديثي الولادة} = \frac{250}{225000} * 1000$$

$$= 1.1/\text{الف} \sim 1/\text{الف}$$

مثال :

اذا كان عدد وفيات النساء في بلد ما بسبب الحمل او الولادة 14000 امرأة ، وعدد المواليد الاحياء 225000 طفل . احسب معدل وفيات الامومة .

الحل :

$$\text{معدل وفيات الامومة} = \frac{\text{عدد وفيات النساء بسبب الحمل الولادة}}{\text{عدد المواليد الاحياء}} * 1000$$

$$\text{معدل وفيات الامومة} = \frac{14000}{225000} * 1000$$

$$= 62.2/\text{الف}$$

2-احصاءات الخصوبة :

تشمل احصاءات الخصوبة :-

$$1 - \text{معدل الولادة الخام} = \frac{\text{عدد المواليد الاحياء}}{\text{عدد السكان في منتصف العام}} * 1000$$

$$2 - \text{معدل الخصوبة العام} = \frac{\text{عدد المواليد الاحياء}}{\text{عدد النساء في سن الحمل في منتصف العام}} * 1000$$

$$3 - \text{معدل الخصوبة للنساء المتزوجات} = \frac{\text{عدد المواليد الاحياء في العام}}{\text{عدد النساء المتزوجات في منتصف العام}} * 1000$$

$$4 - \text{معدل الخصوبة حسب فئات العمر} = \frac{\text{عدد المواليد الاحياء في فئة عمر محددة}}{\text{عدد النساء في تلك الفئة في منتصف العام}} * 1000$$

مثال:اذا كان عدد المواليد الاحياء في سنة ما 4000 طفل وكان عدد النساء في سن الحمل في منتصف العام 40000 امرأة ، وكان عدد السكان في منتصف العام 150000 نسمة جد مايلي:-

1-معدل الخصوبة العام.

2-معدل الولادة العام.

الحل:

1- $\text{معدل الخصوبة العام} = \frac{\text{عدد المواليد الاحياء}}{\text{عدد النساء في سن الحمل في منتصف العام}} * 1000$

$$= \frac{4000}{40000} * 1000 = 100/\text{الف}$$

2-$\text{معدل الولادة الخام} = \frac{\text{عدد المواليد الاحياء}}{\text{عدد السكان في منتصف العام}} * 1000$

$$= \frac{4000}{150000} * 1000$$

$$= 26.6/\text{الف}$$

$$\sim 27/\text{الف}$$

*احصائيات الامراض:

الاحصائيات الصحية:هي الاحصاءات التي تبين الحالة الصحية العامة في البلد وهذا يعني معرفة عدد المرضى المصابين بمرض معين ونسبتهم من مجموع السكان او من الفئات المختلفة التي تؤلف المجوع ، كما تظهر الاحصاءات الصحية الامكانات المتوفرة في البلد لمكافحة الامراض والتغلب عليها من حيث معرفة عدد الاطباء واطباء الاسنان او الصيادلة والممرضات ونسبتهم من مجموع السكان.

تشمل احصاءات الامراض:-

$$\mathbf{1} - \text{معدل الاصابة بمرض معين} = \frac{\text{عدد الاصابات الجديدة من مرض معين خلال السنة}}{\text{عدد السكان في منتصف السنة}} * \mathbf{1000}$$

مثال (1): وصل عدد الاطفال الذين اصيبوا بمرض معين في بلد ما سنة 1995 الى (224) طفل ، اذا علمت ان عدد سكان هذا البلد وصل الى حدود (6500000) عند منتصف سنة 1995 ، احسب معدل الاصابة بالمرض بين الاطفال لنفس السنة؟

الحل:

$$\text{معدل الاصابة بمرض معين} = \frac{\text{عدد الاصابات الجديدة من مرض معين خلال السنة}}{\text{عدد السكان في منتصف السنة}} * \mathbf{1000}$$

$$\text{معدل الاصابة} = \frac{\mathbf{224}}{\mathbf{6500000}} * \mathbf{1000}$$

$$= \mathbf{0.03}/\text{الف}$$

2-معدل انتشار المرض:

$$\text{معدل انتشار المرض} = \frac{\text{عدد الاصابات الموجودين قديما وحديثا في منطقة معين}}{\text{عدد السكان في تلك المنطقة}} * \mathbf{1000}$$

مثال (2):اوضحت دراسة في احدى المدن ان عدد المصابين بالكوليرا عند اجراء تعداد هو (4234) حالة ، احسب معدل الانتشار اذا علمت ان عدد سكان هذه المدينة وصل (23418) نسمة.

الحل:

$$\text{معدل الانتشار} = \frac{\mathbf{4234}}{\mathbf{323418}} * \mathbf{1000}$$

$$= 13/\text{الف}$$

$$3 - \text{نسبة الهلاك} = \frac{\text{عدد حالات الوفاة بسبب المرض}}{\text{عدد حالات الاصابة بهذا المرض}} * 1000$$

مثال (3): من المثال (1) ، اذا علمت ان (224) طفل توفي منهم (97) طفل ، احسب نسبة حالات الهلاك؟

الحل:

$$\text{نسبة الهلاك} = \frac{97}{224} * 1000$$

$$= 433/\text{الف}$$

4-معدل وفيات المرضى داخل المستشفى:

يقيس هذا المعدل عدد الوفيات الحاصلة في المستشفى خلال السنة لكل (1000) من المرضى الخارجين من المستشفى خلال تلك السنة ويساوي:

$$\text{معدل وفيات المرضى داخل المستشفى} = \frac{\text{عدد الوفيات الحاصلة في المستشفى خلال السنة}}{\text{عدد المرضى الخارجين من المستشفى خلال تلك السنة}} * 1000$$

مثال:بلغ عدد حوادث الوفاة في مستشفى معين خلال السنة (342) حادثة ،فاذا كان عدد المرضى الذين غادروا تلك المستشفى خلال تلك السنة (12500)حالة، فما هو معدل وفيات المرضى داخل المستشفى؟

الحل:

$$\text{معدل وفيات المرضى داخل المستشفى} = \frac{342}{12500} * 1000 = 27.36 \sim 27 \text{ بالالف}$$

اي ان هناك 27 حادثة وفاة تقريبا تحدث خلال تلك السنة لكل 1000 من المرضى الخارجين من المستشفى.

5-معدل عدم الاكتمال:

يعبر هذا المقياس عن عدد الولادات الحية بوزن اقل من كيلو غرامين ونصف خلال تلك السنة لكل 1000 من العدد الكلي للولادات الحية خلال تلك السنة يساوي:

$$\text{معدل عدم الاكتمال} = \frac{\text{عدد الولادات الحية بوزن اقل من}\left(\mathbf{2.5}\ \text{كغم}\right)\text{خلال السنة}}{\text{العدد الكلي للولادات الحية خلال السنة}} * \mathbf{1000}$$

مثال:بلغ عدد الولادات الحية التي يقل وزن المولود فيها عن (2.5 كغم) خلال السنة (10925) ، فإذا كان عدد الولادات الحية خلال تلك السنة (1375902) ، فما هو معدل عدم الاكتمال؟

الحل:

$$\text{معدل عدم الاكتمال} = \frac{\mathbf{10925}}{\mathbf{1375902}} * \mathbf{1000} = \mathbf{7.94} \sim \mathbf{8}\ \text{بالالف}$$

اي هناك 8 حالة يقل وزنهم عن (2.5 كغم) من 1000 حالة من عدد الولادات الحية.

جـــــداول الحياة

Tables Life

***طرق تكوين جداول الحياة:**

هناك طريقتين اساسيتين لتكوين جداول الحياة:-

1-الطريقة المباشرة:

وحسب هذه الطريقة يتم رصد التناقص التدريجي لجيل فعلي من السكان ولدوا في يوم واحد ، وهذه الجداول تسمى (جداول الحياة الواقعية) وهذه الطريقة لاتستخدم في الوقت الحاضر لوجود عدة مشاكل في هذه الطريقة منها هجرة السكان.

2-الطريقة غير المباشرة:

وهي الطريقة التي تستخدم لجيل نظري من السكان حيث تحسب احتمالات الوفاة لفئات العمر المختلفة اعتمادا على عدد الوفيات في تلك الاعمار وتصنيف السكان حسب العمر التي يتم الحصول عليها من التعداد ، وهذه الجداول تسمى (جداول الحياة النظرية)

***انواع جداول الحياة:**

جداول الحياة نوعان تبعا لطول فئات العمر:-

1-الجداول المفصلة: وهي الجداول التي تنظم حسب فئات العمر الاحادية (سنوات العمر) ومثل هذه الجداول قد تدعى بجداول الحياة الكاملة.

2-الجداول المختصرة:وهي الجداول التي تنظم بفئات اوسع وقد تكون بفئات خمسية (غالبا) اوعشرية (في احيان اقل) ومثل هذه الجداول هي الاكثر شيوعا.

***فوائد جداول الحياة:**

1-يمكن معرفة عدد الوفيات الممثلة بأي عمر كان.

2-يمكن معرفة عدد الباقين على قيد الحياة لذلك فهي تستخدم بشكل واسع في مجال التأمين على الحياة.

3-يمكن استخدامها لتقدير عدد السكان للسنوات بين التعدادات موزعين حسب فئات العمر.

4-تستخدم مقاييس جداول الحياة فيما يتعلق بأحتمال الحياة او الوفاة ومعدل طول الحياة للمقارنة بين قطرين او اكثر.

5-استخدام جداول الحياة كوسيلة لتحقيق دقة الاحصاءات وخاصة في ما يتعلق بتوزيعات السكان حسب الاعمار او توزيعات الوفيات حسب الاعمار.

تسمى جداول الحياة Tables Life في بعض الاحيان بجداول الوفيات وهي جداول تلخص بصورة رقمية مبسطة تاريخ حياة مجموعة من الاشخاص يولدون في الوقت نفسه ويتعرضون للانقراض بالتدريج بفعل الوفاة يطلق عليهم اسم فوج او جيل ، حيث يتم تتبع هذه المجموعة خلال مراحل العمر المختلفة حتى تنتهي بوفاة كل أفرادها . ويمكن ان يبنى جدول الحياة لكافة الاعمار بحسب آحاد العمر فيسمى بذلك جدول حياة كامل Table Life Complete(او بحسب فئات العمر الخمسية) فيسمى عندئذ جدول حياة مختصر

وعليه يمكن تقسيم جداول الحياة الى قسمين :

1- جدول حياة كامل : يكون طول الفئة في الجدول سنة واحدة .

2- جدول حياة مختصر : يكون طول الفئة على الغالب 5 سنوات ، وهذه تكون اكثر دقة واكثر سهولة .

من ذلك يتبين ان جداول الحياة تُظهر التالي :

1- عدد الذين يتوفون كل عام .

2 - عدد الاشخاص الباقين على قيد الحياة في كل سنة من سنوات العمر .

3- تظهر احتمال الوفاة واحتمال الحياة لكل شخص في فئات العمر المختلفة .

4- تبين عدد السنوات التي تعيشها كل مجموعة حتى نهاية العمر .

5 – تبين متوسط طول العمر المتوقع لكل شخص في فئات العمر المختلفة .

العدد الذي يوخذ في البداية (سواء كان 1000 او 10000 او 100000 او أي عدد اخر) يدعى الاصل او الاساس Radix .

يتم عمل جداول الحياة من العمر صفر وايضا يمكن عملها من اي عمر كان ، لكن الغالب عملها من الولادة الى الوفاة .

ويمكن عمل هذه الجداول لمجموع السكان ككل او يمكن عملها لكل جنس (ذكور واناث) ويمكن عمل جداول منفصلة لاهل المدن ، اهل القرى ، المدن الكبيرة ، او المناطق الادارية او الجغرافية المختلفة او المجموعات المهنية كمجموعة العمال في صناعة معينة .

<u>*مقاييس جداول الحياة:</u>

تتألف جداول الحياة (الوفيات) من اعمدة مختلفة تحتوي على المقاييس التي تتألف منها جداول الحياة وهذه المقاييس هي:-

1-فئات العمر (i) :تتألف من فئات العمر من الولادة الى الوفاة ، ويكون طول الفئة سنة واحدة في بعض الجداول ، لكن على الغالب يكون طول الفئة (5) سنوات وهذه تكون اكثر دقة واكثر سهولة.

2-عدد الاحياء (S_i) : هو يساوي عدد الاشخاص الباقين على قيد الحياة في اية فئة من فئات العمر (S) فمثلاً S_0 هؤلاء سيتوفى بعضهم عند بلوغهم السنة الاولى من العمر ، والاشخاص الباقون على قيد الحياة من هذه السنة هم S_1 ، والاشخاص الذين يصلون الى السنة الخامسة عشر هم S_{15} ، اي ان :

$$S_0 = 100000$$

$$S_5 = 95725$$

3-عدد الوفيات (d_i): هو عدد الوفيات خلال الفترة من العمر *(i)* حتى العمر التالي *(i+1)* فالمقدار (d_7) مثلا هو عدد الوفيات التي تحصل في العمر (7 سنوات) حتى العمر التالي (8 سنوات) مثلا اي انه لو طرح عدد الوفيات في الفئة S من عدد الباقين على قيد الحياة من تلك الفئة لنتج عدد الباقين على قيد الحياة في الفئة التالية اي:-

$$S_{i+1} = S_i - d_i$$

$$d_i = S_i - S_{i+1}$$

ومجموع الوفيات في كافة الاعمار يساوي العدد الاصلي ، اي ان:-

$$100000 = \sum d_{n-1} = S_0 = d_0 + d_1 + d_2 + \cdots + d_{n-1}$$

حيث ان (n) يمثل اعلى مراحل العمر حيث يتوفى كل شخص ،مثلا :

$$d_0 = S_s - S_0$$

$$100000 - 95725 = 4275$$

4-احتمال الحياة (P_i) : وهو احتمال لكل شخص في العمر (i) ان يعيش الى العمر التالي (i+1) وهو يساوي:-

$$P_i = \frac{S_{i+1}}{S_i} = \frac{\text{عدد الاشخاص في العمر التالي}(S+i)}{\text{عدد الاشخاص في العمر السابق}(i)}$$

حيث ان:

$$\boldsymbol{S_{i+1} = S_i - d_i}$$

فإن :

$$\boldsymbol{P_i = \frac{S_i - d_i}{S_i}}$$

5-احتمال الوفاة (q_i) : هو احتمال الوفاة لكل شخص في الوصول من العمر (i) الى العمر (i+1) ويساوي:

$$q_i = \frac{d_i}{S_i}$$

واحتمال الوفاة يكون مكملا لاحتمال الحياة اي ان:-

$$P_i + q_i = 1$$

6-متوسط عدد الاحياء $(\overline{S})$: وهو معدل عدد الاحياء في كل فئة من فئات العمر ، ويساوي :

$$\boldsymbol{\overline{S} = \frac{S_i + S_{i+1}}{2}}$$

7-مجموع السنوات لكل فئة (L_i) : وهو يمثل مجموع السنوات التي يعيشها الاحياء في كل فئة من فئات العمر للوصول الى العمر التالي وهو يساوي:

$$\boldsymbol{L_i = \overline{S}_i} * \text{طول الفئة}$$

حيث ان: $\overline{S}_i$ يمثل متوسط عدد الاحياء.

8-مجموع السنوات التي تعيشها كل مجموعة حتى نهاية العمر (T_i) تستخرج من تجميع العمود السابق تنازلياً (او تعطى القيمة الاولى للفئة الاولى ثم تستخرج القيمة الثانية عن طريق :

$$T_i - (\bar{S}_i * \text{طول الفئة})$$

9-متوسط طول الحياة (e_i) (توقع الحياة): وهو متوسط طول الحياة المقبلة التي يتوقع ان يعيشها كافة الاشخاص في العمر (i) وهو يساوي:

$$e_i = \frac{T_i}{S_i} = \frac{L_i + L_{i+1} + L_{i+2} + \cdots + L_{i-1}}{S_i}$$

$$= \frac{\text{مجموع عدد السنوات في كافة فئات العمر ابتداءً من العمر}\ (i)\ \text{فما فوق}}{\text{عدد الاحياء في العمر}\ (i)}$$

مثال :

كون جدول الحياة من العمر صفر الى العمر 80 ، اذا كان لديك البيانات التالية :

1 – الاساس = 100000

2- طول الفئة = 5

3- عدد الوفيات في الفئة الاولى = 11420 .

4- مجموع السنوات لكل الفئات العمرية = 5994262 .

جدول الحياة

فئات العمر i	عدد الاحياء S_i	عدد الوفيات d_i	عدد الاحياء بعد طرح عدد الوفيات	احتمال الحياة $P_i = \frac{d_i - S_i}{S_i}$	احتمال الوفاة q_i	متوسط الاحياء $\bar{S} = \frac{\boldsymbol{S_i+(S_i-d_i)}}{\boldsymbol{2}}$	عدد سنوات $L_{i=}\bar{S}_i$ طول الفئة *	مجموع السنوات T_i	توقع الحياة $\boldsymbol{e_i = \frac{T_i}{S_i}}$
0	100000 (الاساس)	11420	88580	$\frac{100000 - 11420}{100000}$ $= 0.8858$	$\frac{d_i}{S_i}$ $= 0.1142$	94290	$= 5$ $* 94290$ $= 471450$	5994262 تعطى	$= \frac{5994262}{100000}$ $= 59.9$
5	88580	530	88050	0.99402	0.00598	88315	441575	5994262-471450=5522812	62.3
10	88050	470	87580	0.99466	0.00534	87815	439075	5522812-441575=5081237	57.7
15	87580	645	86935	0.99263	0.00737	87258	436288	4642162	53.0
.									
.									
.									
75	36460	13091	23369	0.64	0.359	29915	149573	207996	5.7
80	23399	23369	30	0	1	11685	58423	58423	2.5