خطة الدراسة لمادة الاحصاء الاجتماعي للمرحلة الثالثة

ت	الاسبوع	المادة
١	الاول	مفهوم الاحصاء
٢	الثاني	انواع الاحصاء
٣	الثالث	فوائد الاحصاء
٤	الرابع	تبويب البيانات
٥	الخامس	الجداول الاحصائية البسيطة
٦	السادس	الجداول الاحصائية المزدوجة
٧	السابع	عرض البيانات
٨	الثامن	الاعمدة البيانية
٩	التاسع	الدائرة البانية
١٠	العاشر	الدائرة البيانية
١١	احدى عشر	المستطيل البياني
١٢	اثنتا عشر	مقاييس النزعة المركزية
١٣	ثلاثة عشر	الوسط الحسابي للبيانات غير المبوبة
١٤	اربعة عشر	الوسط الحسابي للبيانات المبوبة
١٥	خمسة عشر	الوسط الحسابي للبيانات المبوبة
١٦	ستة عشر	امتحان الفصل الاول
عطلة نصف السنة		
١٧	سبعة عشر	اعادة للوسط الحسابي
١٨	ثمانية عشر	الوسيط
١٩	تسعة عشر	الوسيط للبيانات غير المبوبة

٢٠	عشرون	الوسيط للبيانات غير المبوبة
٢١	احدى وعشرون	الوسيط للبيانات المبوبة
٢٢	اثنان وعشرون	الوسيط للبيانات المبوبة
٢٣	ثلاثة وعشرون	المنوال
٢٤	اربعة وعشرون	المنوال للبيانات غير المبوبة
٢٥	خمسة وعشرون	المنوال للبيانات غير المبوبة
٢٦	ستة وعشرون	مقدمة حول مقاييس النزعة المركزية
٢٧	سبعة وعشرون	المدى للبيانات غير المبوبة
٢٨	ثمانية وعشرون	المدى للبيانات المبوبة
٢٩	تسعة وعشرون	التباين
٣٠	ثلاثون	الانحراف المعياري
٣١	واحد وثلاثون	الارتباط
٣٢	اثنان وثلاثون	معامل بيرسون للارتباط

وزارة التعليم العالي والبحث العلمي
جامعة القادسية
كلية الاداب
قسم علم الاجتماع

محاضرات
مادة الاحصاء الاجتماعي
المرحلة الثالثة

اعداد
الدكتور فلاح جابر جاسم الغرابي

مفهوم الإحصاء (Statistics)

ان مفهوم الإحصاء مشتق من الكلمة اللاتينية (status) التي تعني الدولة السياسية . حيث كان الإحصاء مصطلح يستخدم لوصف ظروف ومؤسسات الدولة ثم تحول إلى معنى جديد يعني تطبيق الطرق والمناهج الإحصائية على مؤسسات ونظم الدولة والتي تتعلق بوصف وتحليل البيانات العددية والحقائق المتعلقة بشؤون الدولة ،ثم تطور إلى تطبيق الطرق والمناهج الإحصائية في تحليل البيانات والتوصل إلى الاستنتاجات المتعلقة بشؤون الإنسان والدولة والمجتمع .

وعموما يمكن القول ان الإحصاء هو ذلك العلم الذي يدرس الظواهر الطبيعية والاجتماعية والفكرية دراسة علمية تحاول توضيح ووصف وتحليل هذه الظواهر بالأساليب الكمية والتحليلية التي يعتمدها العالم الإحصائي .

أو هو ذلك العلم الذي يبحث في جمع البيانات وتنظيمها وعرضها وتحليلها واستقراء النتائج واتخاذ القرارات بناء عليها .

ويقصد بجمع البيانات عملية الحصول على القياسات أو الأرقام أو قيم المشاهدات للتجارب التي يجريها الباحث مثل معرفة أعمار الطلبة أو ارتفاع الأشجار أو قياس درجات الحرارة .

أما تنظيم وعرض البيانات فهو عملية وضع البيانات في جداول منسقة وعرضها بطرق مناسبة كالأشكال البيانية. أما تحليل البيانات فهو عبارة عن إيجاد قيم المقاييس . وان استقراء النتائج واتخاذ القرارات يمثل الاستنتاجات التي يتوصل اليها الباحث بناء على نتائج التحليل الإحصائي .

أنواع الإحصاء

يقسم الإحصاء إلى نوعين أساسيين هما

١ – الإحصاء الوصفي : وهو يهتم بوصف وتوضيح وعرض البيانات أو المتغيرات دون القيام بتحليلها واستنتاج الحقائق عنها ، فهو يكتفي بعرضها عرضا جدوليا أو بيانيا مثل تصنيف الطلبة حسب الجنس إلى ذكور وإناث ،أو توزيع الطلبة حسب المحافظات أو توزيع الأشجار الموجودة في بستان حسب نوعها .

٢ – الإحصاء التحليلي : وهذا لا يكتفي بالعرض الوصفي للحالات والظواهر المدروسة بل يذهب إلى تحليل البيانات باستعمال مناهج وطرق الإحصاء التحليلي والتوصل إلى نتائج واتخاذ القرارات بشأنها .

بعض المفاهيم الإحصائية

المتغير (variable) : هو حالة أو وضع ما يراد دراسته مثل العمر ، الطول ، الوزن ، عدد أفراد العائلة ، درجات الحرارة ، عدد الأشجار ، كمية السماد ....الخ . وتصنف المتغيرات إلى أنواع وهي :

١ – متغيرات مستمرة ومتغيرات غير مستمرة (متقطعة) : ويراد بالمتغيرات المستمرة تلك التي تؤخذ قيم تشمل أجزاء من الواحد (الكسر) مثل ارتفاع الأشجار بالمتر والوزن بالكغم حيث يمكن أن يكون الطول أو الوزن مثلا (٨٥ ، ١٢٧,٥) .

أما المتغير غير المستمر (المتقطع) فهو المتغير الذي يعبر عنه بالأعداد الصحيحة فقط مثل عدد أفراد العائلة أو عدد الأشجار في بستان .

٢ – متغيرات مستقلة ومتغيرات تابعة : غالبا ما تتضمن الدراسة أو التجربة أو البحث العلمي متغيرات متعددة (متغيرين أو أكثر) مثلا دراسة تأثير مسافات الزراعة على كمية الحاصل فإنها تتضمن متغيرين وان مسافات الزراعة تمثل متغير مستقل ، أما كمية الحاصل فتمثل متغير تابع ،وان المتغير الذي يؤثر يسمى مستقل أما المتغير الذي يتأثر فيسمى تابع . وعند دراسة تأثير عمق الحراثة وكمية البذور ومسافات الزراعة وكمية السماد على كمية الحاصل فان متغيرات (عمق الحراثة ،كمية البذور ، مسافات الزراعة ، كمية السماد) هي متغيرات مستقلة ، أما متغير كمية الحاصل فهو متغير تابع .

٣ – متغيرات وصفية ومتغيرات كمية : ويقصد بالمتغيرات الوصفية تلك المتغيرات التي تقاس أو يعبر عنها بالوصف فقط ولا يمكن قياسها بصورة عددية مثل مهنة الأب ، جنس الطالب ، نوع الأشجار المزروعة . أما المتغيرات الكمية (العددية) فهي تلك المتغيرات التي تقاس أو يعبر عنها بصورة عددية أو كمية مثل العمر ، الطول ، الوزن ، كمية الإنتاج ، المساحة المزروعة .

المجتمع : هو جميع وحدات الظاهرة المدروسة ، وان مجتمع طلبة جامعة بابل يشمل جميع طلبة الكليات في الجامعة .

العينة : هي جزء من المجتمع وهذه يتم اختيارها تبعا لاعتبارات معينة .

فوائد الإحصاء للباحث

١ – تحويل الظواهر والعلاقات والمتغيرات إلى أرقام يمكن الاستفادة منها في التحليل العلمي .

٢ – تساعد الطرق الإحصائية الباحث في استخلاص النتائج العامة من النتائج الجزئية .

٣ – تساعد الطرق الإحصائية الباحث على التنبؤ بالنتائج التي يحتمل أن يحصل عليها في مستقبلا أو في ظروف خاصة .

٤ – تساعد الطرق الإحصائية الباحث في تحديد اثر عامل دون غيره من العوامل مما لا يمكن تحقيقه عمليا .

تنظيم (تبويب) وعرض البيانات

بعد أن يجمع الباحث البيانات التي يريدها قد يكون من الصعب عليه أن يستوعب هذه البيانات على ما هي عليه دون أن يضعها في صورة مبسطة تسهل دراستها من خلال تبويبها وتقسيمها إلى مجاميع متماثلة أو متشابهة مستخدما لذلك جداول التوزيع التكراري .وجداول التوزيع التكراري هذه قد تكون مفردة (بسيطة) وهي تتضمن عرض متغير واحد فقط ، أو مزدوجة وتتضمن عرض متغيرين .

وفيما يلي مثال لكل نوع من الجداول :

أولا : عرض البيانات الوصفية في جدول التوزيع التكراري

مثال : البيانات التالية تمثل الجنس والمحافظة لعينة من الطلبة

الجنس	ذكر	أنثى	أنثى	ذكر	ذكر	أنثى	ذكر	ذكر	أنثى	أنثى
المحافظة	بابل	بغداد	كربلاء	النجف	بغداد	بابل	واسط	بابل	النجف	كربلاء

المطلوب :

أولا : توزيع الطلبة المبحوثين في جدول توزيع تكراري مفرد (بسيط) تبعا للجنس ، وتبعا للمحافظة

جدول يبين توزيع الطلبة المبحوثين حسب الجنس

الجنس	العدد
ذكر	5
أنثى	٥
المجموع	10

جدول يبين توزيع الطلبة المبحوثين حسب المحافظة

المحافظة	العدد
بابل	3
بغداد	2
كريلاء	2
النجف	2
واسط	1
المجموع	10

ثانيا : : توزيع الطلبة المبحوثين في جدول توزيع تكراري مزدوج تبعا للجنس والمحافظة

جدول توزيع الطلبة المبحوثين تبعا للجنس والمحافظة

المحافظة \ الجنس	ذكر	أنثى	المجموع
بابل	2	1	3
بغداد	1	1	2
كربلاء	ــــ	2	2
النجف	1	1	2
واسط	1	ــــ	1
المجموع	5	5	10

ومن الممكن عرض نفس الجدول بالصورة التالية

جدول توزيع الطلبة المبحوثين تبعا للجنس والمحافظة

الجنس \ المحافظة	بابل	بغداد	كربلاء	النجف	واسط	المجموع
ذكر	2	1	ــ	1	1	5
أنثى	1	1	2	1	ــ	5
المجموع	3	2	2	2	1	10

مثال :البيانات التالية تمثل نوع التربة ونوع المحصول المزروع لعينة من المزارع ، المطلوب عرض البيانات في جدول توزيع تكراري

نوع لتربة	رملية	مزيجيه	رملية	طينية	طينية	مزيجيه	رملية	طينية	مزيجيه	رملية
المحصول	باذنجان	طماطم	بطيخ	باميا	باذنجان	خيار	طماطم	خيار	بطيخ	باميا

نوع لتربة	طينية	مزيجيه	مزيجيه	رملية	طينية	رملية	رملية	مزيجيه	رملية	طينية
المحصول	خيار	باذنجان	طماطم	بطيخ	بطيخ	خيار	طماطم	باذنجان	بطيخ	باميا

جدول توزيع البيانات حسب نوع

التربة والمحصول

نوع التربة / المحصول	طينية	مزيجيه	رملية	المجموع
باذنجان	1	2	1	4
طماطم	ـــ	2	2	4
بطيخ	1	1	3	5
باميا	2	ـــ	1	3
خيار	2	1	1	4
المجموع	6	6	8	20

العرض البيانى للبيانات الغير مبوبة :

والمقصود بالبيانات الغير مبوبة تلك البيانات المفردة أى لا يوجد بها فئات وهناك عدة طرق لعرض البيانات الغير مبوبة .

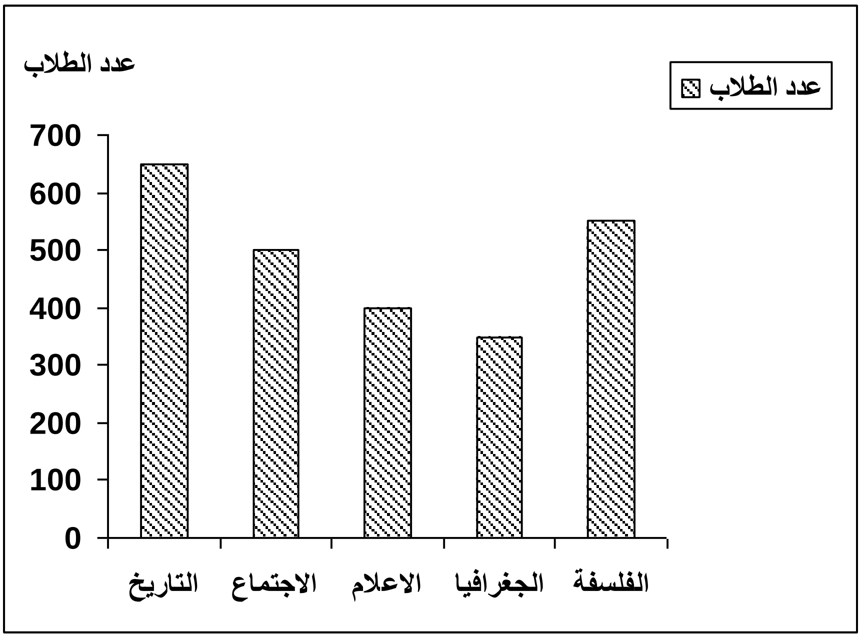
(١) طريقة الأعمدة البيانية البسيطة :

وفى هذه الطريقة يمثل محور السينات قيم المتغير أما محور الصادات يمثل القيمة المقابلة لقيمة المتغير ويتم رسم عمود حول المتغير وارتفاعه يمثل قيمة المتغير .

مثال :

الجدول التالى يوضح أعداد الطلاب ببعض أقسام كلية الآداب جامعة المنصورة والمطلوب عرض هذه البيانات باستخدام طريقة الأعمدة البيانية البسيطة ؟

القسم	التاريخ	الاجتماع	الإعلام	الجغرافيا	الفلسفة
عدد الطلاب	650	500	400	350	550



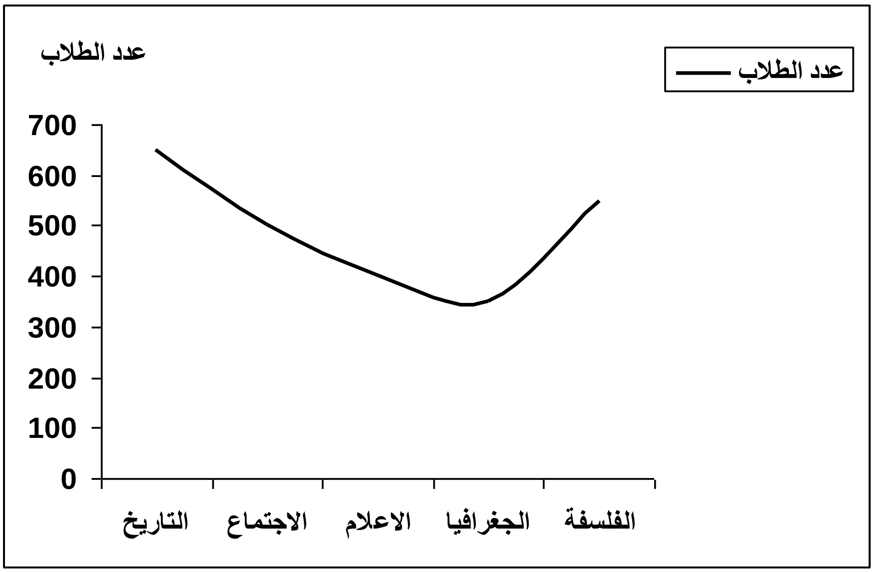
(٢) طريقة المنحنى البيانى البسيط :

وفى هذه الطريقة يمثل محور السينات المتغير أما محور الصادات يمثل قيمة المتغير ويتم توقيع نقاط بين كل قيمة من قيم المتغير على محور السينات والقيمة المقابلة على محور الصادات ثم يتم توصيل تلك النقاط بخط منحنى باليد .

مثال :

الجدول التـالى يوضـح أعداد الطـلاب بـبعض أقسـام كليـة الآداب جامعـة المنصـورة والمطلـوب عـرض هـذه البيانــات باسـتخدام طريقــة المنحنــى البياني البسيطة؟

القسم	التاريخ	الاجتماع	الإعلام	الجغرافيا	الفلسفة
عدد الطلاب	650	500	400	350	550



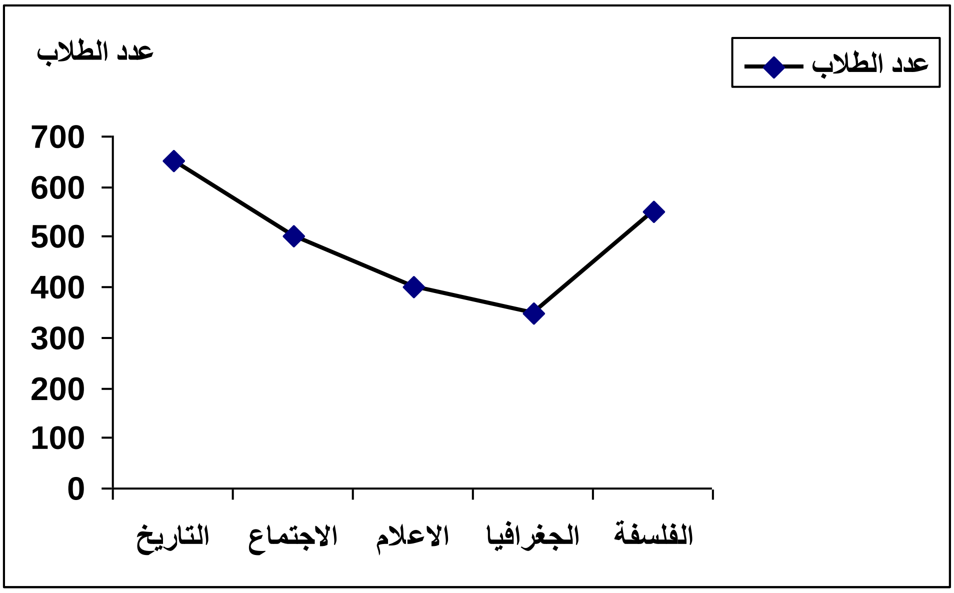
(٣) طريقة الخط البيانى المنكسر :

وفى هذه الطريقة يمثل محور السينات المتغير أما محور الصادات يمثل قيمة المتغير ويتم توقيع نقاط بين كل قيمـة من قيم المتغير علـى محور السينات والقيمة المقابلة على محور الصادات ثم يتم توصيل تلك النقاط بخط منكسر باستخدام المسطرة .

مثال :

الجدول التـالى يوضـح أعداد الطـلاب بـبعض أقسـام كليـة الآداب جامعـة المنصورة والمطلوب عرض هذه البيانـات باستخدام طريقـة الخط البيـاني المنكسر؟

القسم	التاريخ	الاجتماع	الإعلام	الجغرافيا	الفلسفة
عدد الطلاب	650	٥٠٠	400	350	550



(٤) **طريقة الدائرة البيانية** :

وفى هذه الطريقة يتم رسم دائرة ثم نحسب زاوية قطاع كل قيمة على حدة ونقوم برسم تلك الزاوية داخل الدائرة حتى تنتهى الدائرة.

ونحسب زاوية قطاع الجزء من العلاقة :

$$\text{زاوية قطاع الجزء} = \frac{\text{التكرار الفعلى للجزء}}{\text{مجموع التكرارات}} \times 360$$

مثال :

الجدول التالى يوضح أعداد الطلاب ببعض أقسام كلية الآداب جامعة المنصورة والمطلوب عرض هذه البيانات باستخدام طريقة الدائرة البيانية ؟

القسم	التاريخ	الاجتماع	الإعلام	الجغرافيا	الفلسفة
عدد الطلاب	650	500	400	350	550

الحل :

نحسب مجموع التكرارات = 550+350+400+500+650

مجموع التكرارات = 2450

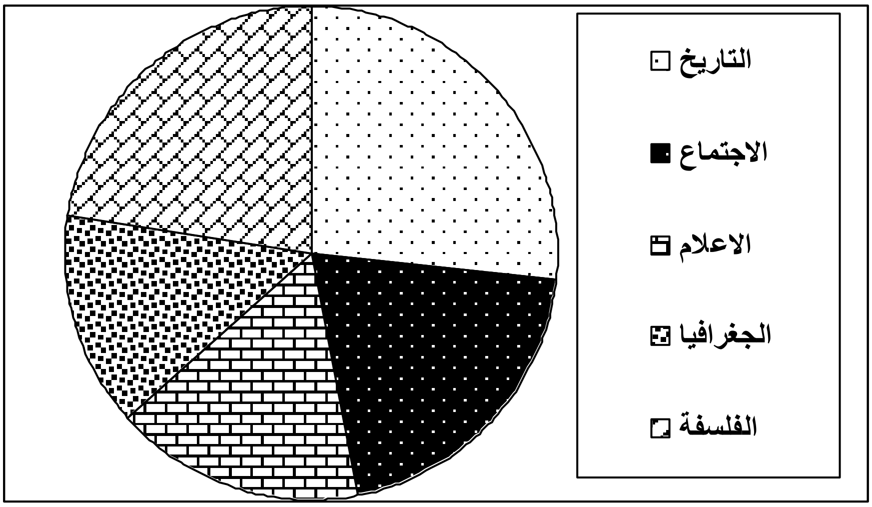
زاوية قطاع التاريخ = $\frac{650}{2450} \times 360 = 95.5°$

زاوية قطاع الاجتماع = $\frac{500}{2450} \times 360 = 73.5°$

زاوية قطاع الإعلام = $\frac{400}{2450} \times 360 = 58.7°$

زاوية قطاع الجغرافيا = $\frac{350}{2450} \times 360 = 51.4°$

زاوية قطاع الفلسفة = $\frac{550}{2450} \times 360 = 80.8°$



(٥) طريقة الأعمدة البيانية المتلاصقة :

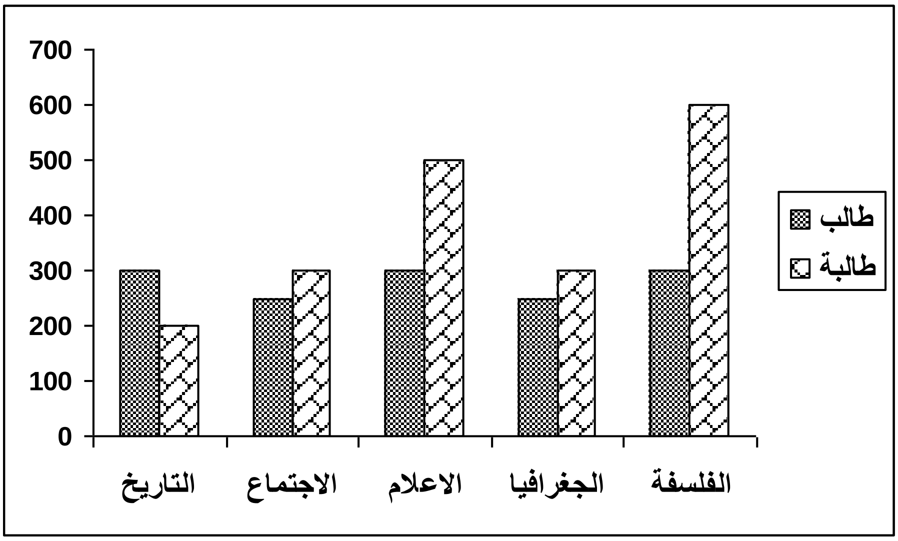
تسمى هذه الطريقة أيضا بطريقة الأعمدة البيانية المتجاورة وهى تشبه طريقة العمدة البيانية البسيطة ولكن يتم رسم عدد من الأعمدة متلاصقة يمثل كل منهم احد قيم المتغير .

مثال :

الجدول التالى يوضح أعداد الطلاب ببعض أقسام كلية الآداب جامعة المنصورة والمطلوب عرض هذه البيانات باستخدام طريقة الأعمدة البيانية المتلاصقة ؟

القسم	التاريخ	الاجتماع	الإعلام	الجغرافيا	الفلسفة
طالب	300	250	300	250	300
طالبة	200	300	500	300	600

الحل :



(٦) طريقة الأعمدة البيانية المجزأة :

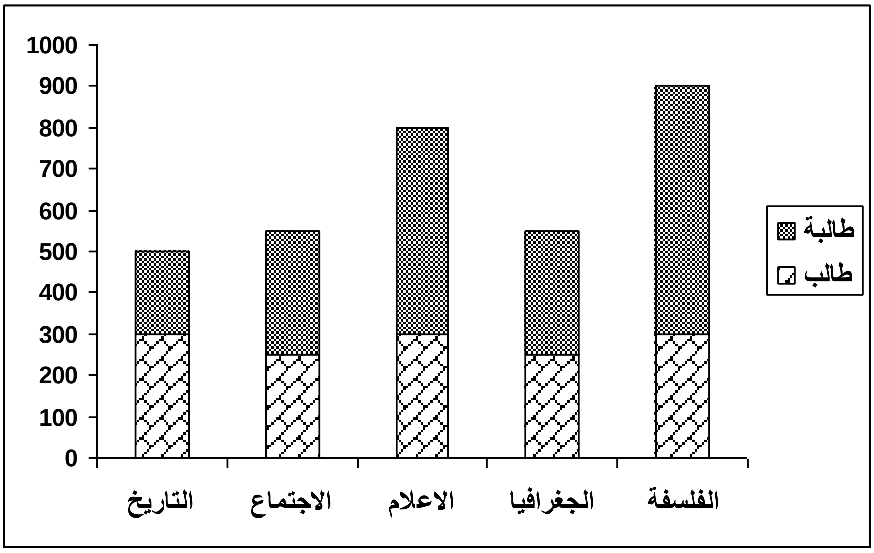
هذه الطريقة تشبه طريقة الأعمدة البيانية البسيطة ولكن يتم رسم عمود يمثل القيمة الأولى للمتغير ثم يليه أو يرتفعه عمود بباقى قيمة المتغير وتكون بادية العمود الثانى هى نهاية العمود الأول .

مثال :

الجدول التالى يوضح أعداد الطلاب ببعض أقسام كلية الآداب جامعة المنصورة والمطلوب عرض هذه البيانات باستخدام طريقة الأعمدة البيانية المجزأة ؟

القسم	التاريخ	الاجتماع	الإعلام	الجغرافيا	الفلسفة
طالب	300	250	300	250	300
طالبة	200	300	500	300	600

الحل :



مقاييس النزعة المركزية
Central Tendency measurement

مقدمة

في كثير من النواحي التطبيقية يكون الباحث في حاجة إلى حساب بعض المؤشرات التي يمكن الاعتماد عليها في وصف الظاهرة من حيث القيمة التي تتوسط القيم أو تنزع إليها القيم ، ومن حيث التعرف على مدى تجانس القيم التي يأخذها المتغير، وأيضا ما إذا كان هناك قيم شاذة أم لا . والاعتماد على العرض البياني وحدة لا يكفى ، ولذا يتناول هذا الفصل، والذي يليه عرض بعض المقاييس الإحصائية التي يمكن من خلالها التعرف على خصائص الظاهرة محل البحث، وكذلك إمكانية مقارنة ظاهرتين أو أكثر ، ومن أهم هذه المقاييس ، مقاييس النزعة المركزية والتشتت .

مقاييس النزعة المركزية

تسمى مقاييس النزعة المركزية بمقاييس الموضع أو المتوسطات ، وهى القيم التي تتركز القيم حولها ، ومن هذه المقاييس ، الوسط الحسابي ، والمنوال ، والوسيط ، والوسط التوافقي وفيما يلي عرض لأهم هذه المقاييس

١- الوسط الحسابي Arithmetic Mean

من أهم مقاييس النزعة المركزية ، وأكثرها استخداما في النواحي التطبيقية ، ويمكن حسابه للبيانات المبوبة وغير المبوبة ، كما يلي :

أولا: الوسط الحسابي للبيانات غير المبوبة

يعرف الوسط الحسابي بشكل عام على أنه مجموع القيم مقسوما على عددها . فإذا كان لدينا **n** من القيم ، ويرمز لها بالرمز : $x_1, x_2, \dots, x_n$.

فإن الوسط الحسابي لهذه القيم ، ونرمز له بالرمز $\bar{x}$ يحسب بالمعادلة التالية :

$$\text{الوسط الحسابي} = \frac{\text{مجموع القيم}}{\text{عدد القيم}}$$

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

حيث يدل الرمز Σ على المجموع .

مثال

فيما يلي درجات 8 طلاب في مادة الإحصاء .

40 36 40 35 37 42 32 34

والمطلوب إيجاد الوسط الحسابي لدرجة الطالب في الامتحان .

الحل

لإيجاد الوسط الحسابي للدرجات تطبق المعادلة رقم (٣-١) كما يلي:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$
$$= \frac{34+32+42+37+35+40+36+40}{8} = \frac{296}{8} = 37$$

أي أن الوسط الحسابي لدرجة الطالب في اختبار مادة الإحصاء يساوي 37 درجة

ثانيا: الوسط الحسابي للبيانات المبوبة

من المعلوم أن القيم الأصلية ، لا يمكن معرفتها من جدول التوزيع التكراري ، حيث أن هذه القيم موضوعة في شكل فئات ، ولذا يتم التعبير عن كل قيمة من القيم التي تقع داخل حدود الفئة بمركز هذه الفئة ،

ومن ثم يؤخذ في الاعتبار أن مركز الفئة هو القيمة التقديرية لكل مفردة تقع في هذه الفئة.

فإذا كانت k هي عدد الفئات ، وكانت $x_1, x_2, \ldots, x_k$ هي مراكز هذه الفئات، $f_1, f_2, \ldots, f_k$ هي التكرارات ، فإن الوسط الحسابي يحسب بالمعادلة التالية:

$$\boxed{\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_k f_k}{f_1 + f_2 + \ldots + f_k} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}}$$

مثال

الجدول التالي يعرض توزيع 40 تلميذ حسب أوزانهم .

فئات الوزن	32– 34	34– 36	36– 38	38– 40	40– 42	42– 44
عدد التلاميذ	4	7	13	10	5	1

والمطلوب إيجاد الوسط الحسابي.

الحل

لحساب الوسط الحسابي باستخدام المعادلة رقم (٣-٢) يتم إتباع الخطوات التالية :

إيجاد مجموع التكرارات $\sum f$ **.** **٢- حساب مراكز الفئات** x

.

ضرب مركز الفئة في التكرار المناظر له $(x\,f)$**، وحساب المجموع** $\sum xf$

حساب الوسط الحسابي بتطبيق المعادلة رقم (٣-٢) .

فئات الوزن (C)	التكرارات f	مراكز الفئات x	$x\ f$
32-34	4	(32+34)÷2=33	4×33=132
35-37	7	٣٦	7×36=252
38-40	13	٣٩	13×39=507
41-43	10	٤٢	10×42=420
44-46	5	٤٥	5×45=225
47-49	1	٤٨	1×48=48
المجموع	40		1584

إذا الوسط الحسابي لوزن التلميذ هو :

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i f_i}{\sum_{i=1}^{6} f_i} = \frac{1584}{40} = 39.6\ k.g$$

أي أن متوسط وزن التلميذ يساوي 39.6 k.g

المحاضرة الخامسة عشر

٢- الوسيط Median

هو أحد مقاييس النزعة المركزية، والذي يأخذ في الاعتبار رتب القيم ، ويعرف الوسيط بأنه القيمة التي يقل عنها نصف عدد القيم $(n/2)$ ، ويزيد عنها النصف الآخر $(n/2)$، أي أن 50% من القيم أقل منه، 50% من القيم أعلى منه. وفيما يلي كيفية حساب الوسيط في حالة البيانات غير مبوبة ، والبيانات المبوبة.

أولا: الوسيط للبيانات غير المبوبة

لبيان كيف يمكن حساب الوسيط للبيانات غير المبوبة ، نتبع الخطوات التالية:

ترتب القيم تصاعديا .

تحديد رتبة الوسيط، وهي : رتبة الوسيط = $\left(\frac{n+1}{2}\right)$

إذا كان عدد القيم (n) فردي فإن الوسيط هو:

$$\boxed{\text{الوسيط = القيمة رقم} \left(\frac{n+1}{2}\right)}$$

إذا كان عدد القيم (n) زوجي، فإن الوسيط يقع بين القيمة رقم $(n/2)$، والقيمة رقم $((n/2)+1)$، ومن ثم يحسب الوسيط بتطبيق المعادلة التالي:

$$\text{الوسيط} = \frac{\text{القيمة رقم}\left(\frac{n}{2}\right) + \text{القيمة رقم}\left(\frac{n}{2}+1\right)}{2}$$

مثال

تم تقسيم قطعة أرض زراعية إلى 17 وحدة تجريبية متشابهة ، وتم زراعتها بمحصول القمح ، وتم استخدام نوعين من التسميد هما : النوع (a) وجرب على 7 وحدات تجريبية ، والنوع (b) وجرب على 10 وحدات تجريبية ، وبعد انتهاء الموسم الزراعي ، تم تسجيل إنتاجية الوحدة بالطن / هكتار ، وكانت على النحو التالي :

النوع (a)				1.5	2.3	3	2	3.25	2.75	1.2
النوع (b)	3	2.5	4	1.5	2.5	2	3.75	3.5	1.8	4.5

والمطلوب حساب وسيط الإنتاج لكل نوع من السماد المستخدم، ثم قارن بينها.

الحل

أولا : حساب وسيط الإنتاج للنوع الأول (a)

ترتيب القيم تصاعديا :

قيمة الوسيط

الإنتاج	1.2	1.5	2	2.3	2.75	3	3.25
الرتبة	1	2	2	4	5	6	7

رتبة الوسيط

عدد القيم فردى $(n = 7)$

إذا رتبة الوسيط هي: $((n+1)/2 = (7+1)/2 = 4)$.

ويكون الوسيط هو القيمة رقم **4** ، أي أن وسيط الإنتاج للنوع **a** هو:

$Me_a = 2.3$ طن / هكتار

ثانيا : حساب وسيط الإنتاج للنوع الثاني (b) :

ترتيب القيم تصاعديا .

قيمة الوسيط = $\frac{2.5 + 3}{2}$

2.75

الإنتاج	1.5	1.8	2	2.5	2.5	3	3.5	3.75	4	4.5
الرتبة	1	2	3	4	5	6	7	8	9	10

5.5

رتبة الوسيط

عدد القيم زوجي $(n = 10)$ إذا

رتبة الوسيط هي : $((n+1)/2 = (10+1)/2 = 5.5)$.

الوسيط = الوسط الحسابي للقيمتين الواقعتين في المنتصف (رقم **5** ، **6**)

.

$$Me_b = \frac{2.5+3}{2} = 2.75$$ طن / هكتار

ويمقارنة النوعين من السماد ، نجد أن وسيط إنتاجية النوع (a)

أقل من وسيط إنتاجية النوع (b) ، أي أن : $Me_b > Me_a$.

ثانيا: الوسيط للبيانات المبوبة

لحساب الوسيط من بيانات مبوبة في جدول توزيع تكراري ، يتم إتباع الخطوات التالية .

تكوين الجدول التكراري المتجمع الصاعد .

تحديد رتبة الوسيط : $\left(\frac{n}{2}\right)=\left(\frac{\sum f}{2}\right)$

ويحسب الوسيط ، بتطبيق المعادلة .

$$Me=L_k+\frac{h_k}{f_k}\left[\frac{\sum f_i}{2}-f_{k-1}^{/}\right]$$

حيث أن :

L_k **: هي الحد الأدنى للفئة الوسيطية المقابلة لأكبر تكرار.**

h_k : طول الفئة الوسيطية

f_k : تكرار الفئة الوسيطية.

f'_{k-1} : التكرار المتجمع الصاعد السابق للفئة الوسيطية

3-المنوال Mode

يعرف المنوال بأنه القيمة الأكثر شيوعا أو تكرارا ، ويكثر استخدامه في حالة البيانات الوصفية ، لمعرفة النمط (المستوى) الشائع، ويمكن حسابه للبيانات المبوبة وغير المبوبة كما يلي:

أولا: حساب المنوال في حالة البيانات غير المبوبة

$$\boxed{\text{المنوال } (\mathbf{Mod}) = \text{القيمة (المستوى) الأكثر تكرارا}} \qquad (٣-١٢)$$

ثانيا: حساب المنوال في حالة البيانات المبوبة.

$$Mod = L_k + \frac{(f_k - f_{k-1})}{(f_k - f_{k-1}) + (f_k - f_{k+1})} * h_k$$

حيث أن :

Lk : الحد الأدنى لفئة المنوال المقابلة لأكبر تكرار.

f_k : تكرار الفئة المنوالية

f_{k-1} : التكرار السابق للفئة المنوالية

f_{k+1} : التكرار اللاحق للفئة المنوالية

h_k : طول فئة المنوال

مقاييس التشتت

Measures of Dispersion

تناولنا في محاضرات سابقة المتوسطات أو مقاييس النزعة المركزية والتي تعبر عن المستوى العام للظاهرة التربوية او السياسية أو الاجتماعية محل البحث. ولكن ترى هل هذا يعتبر كافياً لوصف البيانات وتحليلها كمياً لنصل بالتالي إلى فهم أكثر وضوحاً للظاهرة محل الدراسة ؟ للإجابة على هذا السؤال نسوق المثال التالي :

مثال ١ :

نفترض أن لدينا شريحتين من شرائح المجتمع تعيشان في منطقتين مختلفتين وكانت دخولهم الأسبوعية (بالدولار الأمريكي) هي كما يلي :

المجموعة الأولى A :

70	75	71	75	74	76	73	78

المجموعة الثانية B :

99	56	80	100	29	70	65	93

ويحساب الوسط الحسابي للشريحتين المذكورتين في كل من المجموعتين

– حسب ما أوضحناه في الفصل السابق – فإن الوسط الحسابي لدخول المجموعة الأولى :

$$\overline{X}_A = \frac{70+75+71+75+74+76+73+78}{8}$$

$$\overline{X}_A = \frac{592}{8} = 74 \text{ دولاراً}$$

أي أن الوسط الحسابي لدخل المجموعة A هو 74 دولاراً.

وكذلك بالنسبة للمجموعة B فإن الوسط الحسابي لدخولها هو :

$$\overline{X}_B = \frac{99+56+80+100+29+70+65+93}{8}$$

$$\overline{X}_B = \frac{592}{8} = 74 \text{ دولاراً}$$

والوسط الحسابي لدخل المجموعة B هو أيضاً 74 دولاراً.

ومعنى هذا أن المستوى العام لدخل الأفراد في المجموعتين واحد ويساوي

74 دولاراً. ولكن بإمعان النظر في دخول المجموعة الأولى نجد أنها متجانسة إلى حدٍ كبير أي أنها قريبة جداً من بعضها أو من الوسط الحسابي والذي يساوي 74 دولاراً. وهنا نقول أن تشتت الدخول قليل أو صغير. بينما دخول المجموعة الثانية غير متجانسة فهي بعيدة عن بعضها، أو عن الوسط الحسابي بشكل كبير. وهنا نقول أن تشتت دخول المجموعة الثانية كبير فهي أقل تجانساً (أو أكثر تشتتاً) من المجموعة الأولى.

(الفرق بين أكبر دخل وأصغر دخل في المجموعة الأولى: ٧٨– ٧٠ =٨ دولارات فقط، بينما الفرق بين أكبر دخل وأصغر دخل في المجموعة الثانية: ١٠٠– ٢٩–٧١ دولاراً)

نستنتج من هذا المثال بأن المتوسط ليس كافياً لتوصيف البيانات أو تحليلها كمياً. فها هو المتوسط واحد في المجموعتين ورغم

ذلك فإن البيانات تختلف تماماً في مدى تشتتها (أو تجانسها). وهنا تبرز الحاجة إلى مقاييس كمية أو إحصائية ليقاس مدى تشتت البيانات ولتعطي للباحث بالتالي صورة أكثر وضوحاً وصدقاً للظاهرة السياسية محل الدراسة. وفيما يلي نتناول بعض مقاييس التشتت التي تخدم الغرض من هذا الكتاب وهي :

المدى، والتباين والانحراف المعياري، ومعامل اختلاف

المدى The Range

يعرّف المدى بأنه " الفرق بين أكبر قيمة وأصغر قيمة ".

فمدى المجموعة A في المثال رقم (١) هو : $78 - 70 = 8$

ومدى المجموعة B في نفس المثال هو : $100 - 29 = 71$

وواضح تماماً أن مدى المجموعة الثانية أكبر بكثير جداً من مدى المجموعة الأولى مما يعني أن المجموعة الثانية أكثر تشتتاً من المجموعة الأولى.

ومن تعريف المدى يتضح أنه مقياس بسيط جداً وأن كل المطلوب معرفته هو أكبر قيمة وأصغر قيمة فقط. ومن هذا التعريف نخرج بالملاحظات التالية :

١ – أن المدى يعتمد في حسابه على قيمتين فقط (أكبر قيمة وأصغر قيمة) وأنه يهمل بالتالي باقي القيم.

٢ – أنه لا يقيس تشتت البيانات عن متوسطها، فهو لا يشير – من قريب أو من بعيد – إلى متوسط البيانات (أو مركزها).
٣ – أنه حساس جداً لأي قيمة شاذة أو متطرفة والمثال التالي يوضح هذه النقطة بشيء من التفصيل.

مثال (٢) :
إذا كانت أعمار أعضاء السلطة التشريعية في بلد ما هي :
72 78 20 74 73 75 70
فإن المدى في هذه الحالة هو : 78 – 20 = 58
وهنا نلاحظ وجود قيمة شاذة بالنسبة لباقي القيم وهي 20 وهي أصغر قيمة. وإذا أهملت هذه القيمة (أو لم تكن موجودة أصلاً) لكان المدى :
78 – 70 = 8
وهذا يعني أن وجود قيمة شاذة (20) رفعت قيمة المدى من 8 سنوات إلى 58 سنة. وهذا يوضح مدى حساسية هذا المقياس للقيم الشاذة (أو المتطرفة).
ولكل هذه الأسباب – أو الملاحظات – فإن كثيراً من الإحصائين والباحثين لا يعتمدون كثيراً على المدى كمقياس للتشتت. ويستخدم فقط إذا كان المطلوب فكرة سريعة أو عامة (وليست دقيقة) عن مدى تشتت البيانات.

التباين The Variance :

يعتبر التباين أحد مقاييس التشتت المهمة لأنه من ناحية يأخذ جميع القيم في الاعتبار عند حسابه، ومن ناحية أخرى لأنه يقيس التشتت عن الوسط الحسابي للقيم، هذا بالإضافة إلى أنه تسهل معالجته رياضياً، وأنه يدخل في تكوين عدد من المقاييس والاختبارات الإحصائية المهمة.

والفكرة الأساسية للتباين هي حساب إنحرافات جميع القيم عن وسطها الحسابي

(أي حساب الفرق بين كل قيمة والوسط الحسابي)، وسوف نجد أن بعض القيم أكبر من الوسط فتكون الفروق (أو الإنحرافات) بالموجب، والبعض الأخر أصغر من الوسط فتكون الفروق (أو الإنحرافات) بالسالب. ودائماً يكون مجموع هذه الإنحرافات مساوياً للصفر. ويكون الحل هنا إما إهمال الإشارات السالبة أو تربيع هذه الإنحرافات. وإهمال الإشارات السالبة ليس له مبرر رياضي، فيكون الحل هو تربيع هذه الإنحرافات، ثم نحسب متوسط الانحرافات المربعة فنحصل على التباين.

أي أن التباين يعرف كما يلي:

التباين : هو متوسط مربعات إنحرافات القيم عن وسطها الحسابي.

والمثال التالي يوضح كيفية حساب التباين.

مثال (٣) :

أحسب تباين دخول الشريحية الاجتماعية التالية : مثال رقم (١) :

78 76 76 74 75 71 75 70

الحل :

١ – نحسب أولاً الوسط الحسابي للدخول كما يلي :

$$\overline{X} = \frac{592}{8} = 74$$

٢ – نحسب إنحرافات القيم عن الوسط الحسابي (أي الفروق بينها وبين الوسط الحسابي) كما يلي :

$$4+, 1-, 2+, 0, 1+, 3-, 1+, 4-$$

(لاحظ أن مجموع الانحرافات يساوي صفر)

٣ – نربع هذه الانحرافات كما يلي :

16 1 4 0 1 9 1 16

٤ – التباين يساوي متوسط هذه المربعات، أي يساوي :

$$\frac{16+1+9+1+0+4+1+16}{8}=\frac{48}{8}=6$$

الإنحراف المعياري : Standard Deviation :

نلاحظ أن تمييز التباين سيكون " دخول مربعة " وبصفة عامة " وحدات مربعة " لأنه يتم تربيع الانحرافات (أو الفروق). لذلك فإنه يؤخذ الجذر التربيعي الموجب للتباين فنحصل على ما يسمى " الإنحراف المعياري " والذي سيكون تمييزه هو نفس الوحدات. ففي هذا المثال يكون الانحراف المعياري للدخول هو :

دولاراً $\sqrt{6}=2.449$

والانحراف المعياري – عادة – يستخدم بدلاً من التباين كأهم مقياس للتشتت. فالتباين ما هو إلا مربع الانحراف المعياري، والانحراف المعياري ما هو إلا الجذر التربيعي الموجب للتباين.

تعريف الانحراف المعياري : الانحراف المعياري هو. الجذر التربيعي الموجب للوسط الحسابي لمربعات انحرافات القيم عن وسطها السحابي.

أي هو الجذر التربيعي الموجب للتباين. وعادة يرمز للانحراف المعياري للعينة بالرمز S ولتباين العينة بالرمز S2 وللانحراف المعياري للمجتمع بالرمز اللاتيني σ (سيجما) ولتباين المجتمع بالرمز σ^2.

ويمكن التعبير عن كل من التابين والانحراف المعياري بالرموز كما يلي :

إذا فرضنا أن قيم العينة هي :

X1 , X2 , ... , Xn (أي عددها n) فإن الخطوات تكون كما يلي :

١ – حساب الوسط الحسابي أي حساب $\overline{X}$ حيث :

$$\overline{X} = \frac{\sum X}{n}$$

٢ – حساب إنحرافات القيم عن الوسط الحسابي، أي :

$$X_1 - \overline{X}, X_2 - \overline{X}, \ldots, X_n - \overline{X}$$

٣ – تربيع هذه الانحرافات، أي :

$$(X_1 - \overline{X})^2, (X_2 - \overline{X})^2, \ldots, (X_n - \overline{X})^2$$

٤ – التباين هو الوسط الحسابي لهذه المربعات، أي :

$$S^2 = \frac{(X_1 - \overline{X})^2 +, (X_2 - \overline{X})^2, + \ldots +, (X_n - \overline{X})^2}{n}$$

والتي يمكن كتابته (بإختصار البسط) كما يلي :

$$\boxed{S^2 = \frac{\sum_{i=1}^{n} (X_I - \overline{X})^2}{n}}$$

تباين العينة

٥ – الانحراف المعياري هو الجذر التربيعي الموجب للتباين، أي :

$$\boxed{S = \sqrt{\frac{\sum (X_I - \overline{X})^2}{n}}}$$

الانحراف المعياري للعينة

وبالطريقة نفسها إذا كانت لدينا قيم المجتمع الإحصائي. فإذا كانت قيم المجتمع هي : XN , ... , X2 , X1 أي عددها N فإن قانوني التباين والانحراف المعياري للمجتمع – بافتراض أن الوسط الحسابي للمجتمع هو μ (ميو) – يمكن كتابتها كما يلي:

تباين المجتمع

$$\boxed{\sigma^2 = \frac{\sum (x - \mu)^2}{N}}$$

الانحراف المعياري للمجتمع

$$\boxed{\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}}$$

ملاحظة مهمة :

يمكن حساب كل من التباين والانحراف المعياري بطريقة مختصرة – وتعطي النتائج نفسها بطبيعة الحال – كما يلي :

* تباين العينة بالطريقة المختصرة :

$$S^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$$

* الانحراف المعياري للعينة بالطريقة المختصرة :

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

وفي الطريقة المختصرة فإن كل المطلوب معرفته لحساب التباين أو الإنحراف المعياري هو $\sum x$ (أي مجموع القيم)، $\sum x^2$ (أي مجموع مربعات القيم) ثم التعويض في القانون.

مثال (٤) :

أحسب التباين والإنحراف المعياري بالطريقة المختصرة لبيانات المثال رقم (٣).

الحل :

يمكن تنظيم الحل في الجدول التالي :

X	X2
70	4900
75	5625
71	5041
75	5625
74	5476
76	5776
73	5329
78	6084
$\sum x = 592$	$\sum x^2 = 43856$

ويكون التباين (حيث $n = 8$) هو :

$$
\begin{aligned}
S^2 &= \frac{\sum x^2}{n} = (\frac{\sum x}{n})^2 \\
&= \frac{43856}{8} = \left(\frac{592}{8}\right)^2 \\
&= 5482 - 5476 \\
S^2 &= 6
\end{aligned}
$$

وبالتالي فإن الإنحراف المعياري هو : $S = \sqrt{6} = 2.449$ وهي النتائج نفسها التي حصلنا عليها سابقاً ولكن بطريقة أبسط.

المحاضرة الثامنة عشر

بعض خصائص الإنحراف المعياري :

١ – قيمة الانحراف المعياري دائماً موجبة أو أكبر من أو تساوي صفر. فأقل قيمة تساوي الصفر (وذلك عندما تكون جميع القيم متساوية، وفي هذه الحالة لا توجد فروق أو إنحرافات بينها وبين الوسط الحسابي وبالتالي لا يوجد أي تشتت بين القيم، وبالتالي فإن قيمة الإنحراف المعياري في حالة تساوي جميع القيم تساوي الصفر).

٢ – كلما كان التشتت كبيراً حول الوسط كلما كان الانحراف المعياري كبيراً، والعكس صحيح.

٣ – إذا أضفنا وطرحنا مقداراً ثابتاً من كل القيم فإن قيمة الانحراف المعياري (أو التباين) لا تتغير (أي لا تتأثر قيمة الانحراف المعياري بالطرح أو الجمع). ولتوضيح هذه الخاصية نأخذ المثال التالي :

مثال (٥) :

حل المثال السابق رقم ٤ بعد طرح 70 (على سبيل المثال من كل القيم لاحظ أن هذا سوف يسهل الحسابات).

الحل :

المربعات (بعد الطرح)	القيمة بعد طرح (70)
X2	X
0	0
25	5
1	1
25	5
16	4
36	6
9	3
64	8
ΣX 2 = 176	ΣX = 176

ويكون التباين

$$S^2 = \frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2$$

$$= \frac{176}{8} - \left(\frac{32}{8}\right)^2$$

$$= 22 - 16$$

$$= 6$$

والانحراف المعياري $S=\sqrt{6}=2.449$ **وهي النتائج نفسها مع ملاحظة أن العمليات الحسابية أسهل في هذه الحالة. وتجدر الإشارة إلى أنه لو طرحنا أي قيمة أخرى سنحصل على النتائج نفسها.**

٤- إذا ضربنا كل قيمة في مقدار ثابت ثم حسبنا الانحراف المعياري للقيم الجديدة فإنه يجب القسمة على هذا المقدار الثابت. وإذا قسمنا كل قيمة على مقدار ثابت ثم حسبنا الانحراف المعياري للقيم الجديدة فإنه يجب الضرب في هذا المقدار الثابت.

مثال (٧) : الجدول التالي يمثل توزيع مجموعة من الناخبين حسب أعمارهم وفي وقت معين، والمطلوب حساب الانحراف المعياري لأعمار الناخبين ؟

عدد الناخبين f	الأعمار X
3	20
4	25
5	30
2	35
14	المجموع

الحل :

ينظم الحل كما في الجدول التالي حيث يتم حساب كل من Xf، X2f للحصول على ΣXf، $\Sigma X^2 f$

حاصل ضرب	حاصل ضرب	أعداد	الأعمار

X	الناخبين f	Xf	X2f
20	3	60	1200
25	4	100	2500
30	5	150	4500
35	2	70	2450
المجموع	Σf = 14	Σxf = 380	Σx2f = 10650

ويحسب الانحراف المعياري باستخدام الصيغة التالية :

$$S=\sqrt{\frac{\sum x^2 f}{\sum f}-\left(\frac{\sum xf}{\sum f}\right)^2}$$

$$=\sqrt{\frac{10650}{14}-\left(\frac{380}{14}\right)^2}$$

$$=\sqrt{760.7-736.6}$$

$$=\sqrt{24.1}=4.9$$

أي أن الانحراف المعياري لأعمار الناخبين يساوي ٤.٩ سنة.

مثال (٨) شامل على التشتت :

البيانــات فــي الجــدول الخــاص بالمثــال الشــامل رقــم (٦) علــى المتوسطات التي تمثل أهم الحـروب التـي شـهدها العالم خـلال الفتـرة مـن عـام 1945، وحتـى عـام 1980م. احسـب المـدى، الانحـراف المعيـاري، ومعامل الاختلاف لعدد الحروب خلال الفترة ؟

الحل :

يمكن تلخيص بيانات الجدول رقم (١) في الجدول التالي :

الحروب X	التكرارات f
0	4
1	8
2	7
3	9
4	6
6	2
المجموع	36

والجـدول يوضـح أن هنـاك أربـع سـنوات لـم تـحـدث فيهـا حـروب، وثمـان سنوات حدثت في كل منـها حرب واحدة، وسبع سنوات حدثت في كل منـها حربان، وهكذا....

أولاً: المدى

المدى هو الفرق بين أكبر قيمة وأصغر قيمة. وأكبر قيمة هنا تساوي 6 حروب وأصغر قيمة تساوي صفر (لا توجد حروب)، وبالتالي فالمدى هو :

(6 = 0 − 6) أي أن المدى يساوي 6 حروب.

ثانياً: الانحراف المعياري :

يمكن تلخيص الحسابات كما في الجدول التالي :

حاصل ضرب X2f	حاصل ضرب X. f	التكرارات f	الحروب X
0	0	4	0
8	8	8	1
28	14	7	2
81	27	9	3
96	24	6	4
72	12	2	6
Σx2f = 285	Σxf = 85	Σf =36	المجموع

$$S = \sqrt{\frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2}$$

$$= \sqrt{\frac{285}{36} - \left(\frac{85}{36}\right)^2}$$

$$= \sqrt{7.917 - 5.569}$$

$$= \sqrt{2.348} = 1.53$$

أي أن الانحراف المعياري لعدد الحروب خلال تلك الفترة يساوي 1.53 حرباً

(لاحظ أن التباين يساوي 2.348 فهو القيمة قبل استخراج الجذر)

معامل الارتباط

Auto Correlation

لاحظنا في الفصول السابقة الطرق و الأساليب المختلفة في جمع البيانات و تصنيف و تبويب البيانات ، كذلك عملية استخراج بعض المقاييس التي تعطي فكرة أكثر و وضوحا عن تلك البيانات كالمتوسطات و مقاييس التشتت ، إن هذه الطرق و الأساليب استندت على البيانات المجمعة عن متغير واحد فقط سواء كانت هذه البيانات مبوبة في توزيع تكراري أم غير ذلك . وفي أحوال كثير يواجه الباحث حالات تتطلب دراسة متغيرين أو أكثر في آن واحد لبيان طبيعة و نوع العلاقة التي تربط بها هذه المتغيرات ، عليه فان هذا الفصل سوف يخصص لدراسة مقاييس أخرى تحدد درجة و نوع وشكل العلاقة بين متغيرين أو أكثر.

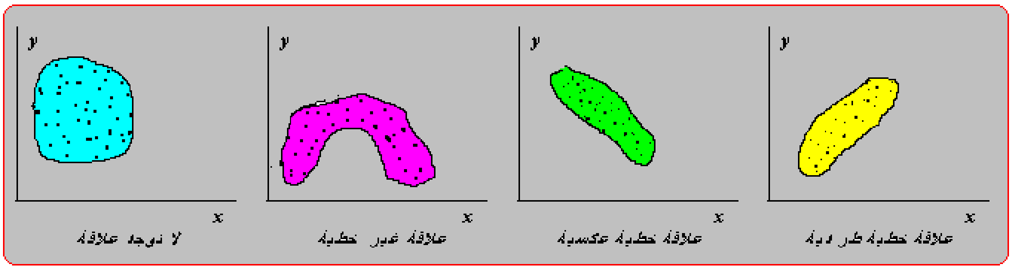
الارتباط الخطي Linear Correlation

أن مفهوم الارتباط الخطي يقترن بحالة وجود متغيرين أو أكثر ترتبط مع البعض بعلاقات خطية معينة على سبيل المثال العلاقة

بين طول الشخص و وزنه ، العلاقة بين تحصيل الطالب المتخرج من الكلية و معدل درجاته في الثانوية و مستوى ألمعاشي لأسرته ، العلاقة بين نسبة الشفاء من مرض معين و كمية الجرعة من الدواء المخصص للمريض و عمر المريض. فإذا كان التغير في إحدى المتغيرات يؤثر في تغير متغير أخر أو مجموعة متغيرات أخرى عندئذ يقال أن هذه المتغيرات مرتبطة فيما بينها و إذا كان المتغيرين المرتبطين (أو مجموعة من المتغيرات المترابطة) يتغيران (تتغير) بنفس الاتجاه أي زيادة أو نقصان في إحداهما تؤدي إلى الزيادة أو النقصان في الأخر (الأخرى) عندئذ يقال أن الارتباط ما بينهما هو ارتباط موجب ، على سبيل المثال زيادة طول الشخص يتوقع أن يقابلها زيادة في وزنه. انخفاض في دخل الفرد يتوقع عنه انخفاض في إنفاقه على بعض السلع. أما إذا كان المتغيرين المرتبطين (أ، مجموعة من المتغيرات المرتبطة) يتغيران (تتغير) باتجاه معاكس أي زيادة أ، نقصان في إحداهما تؤدي إلى نقصان (زيادة) في الأخر (الأخرى)، عندئذ يقال أن الارتباط ما بينهما هو ارتباط سالب على سبيل المثال زيادة سعر الوحدة من سلعة معينة يتوقع إن يؤدي إلى انخفاض في الطلب على تلك السلعة ، انخفاض في درجات الحرارة يتوقع أن ينجم عنه زيادة الطلب على الوقود.

و يقال أن الارتباط بين المتغيرين أو أكثر هو ارتباط تام perfect إذا كان التغير في إحداهما متناسب مع التغير في الأخر على سبيل المثال إن الارتباط بين درجة الحرارة المئوية و درجة الحرارة الفهرنهايتية هو ارتباط تام باعتبار إن التغير في الأول متناسب مع التغير في الثاني.

$$F=(9/5)C+32$$



الارتباط الخطى البسيط Simple Correlation

إذا كان الغرض من التحليل هو تحديد نوع وقوة العلاقة بين متغيرين ، يستخدم تحليل الارتباط ، وأما إذا كان الغرض هو دراسة وتحليل أثر أحد المتغيرين على الآخر ، يستخدم تحليل الانحدار، وفي هذا الفصل يتم عرض أسلوب تحليل الارتباط الخطي البسيط، أي في حالة افتراض أن العلاقة بين المتغيرين تأخذ الشكل الخطي .

الغرض من تحليل الارتباط الخطى البسيط

الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين، ويرمز له في حالة المجتمع بالرمز ρ (رو)، وفي حالة العينة بالرمز r، وحيث أننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع، سوف نهتم بحساب معامل الارتباط في العينة r كتقدير لمعامل الارتباط في المجتمع، ومن التحديد السابق للغرض من معامل الارتباط، نجد أنه يركز على نقطتين هما:

نوع العلاقة:. وتأخذ ثلاث أنواع حسب إشارة معامل الارتباط كما يلي:

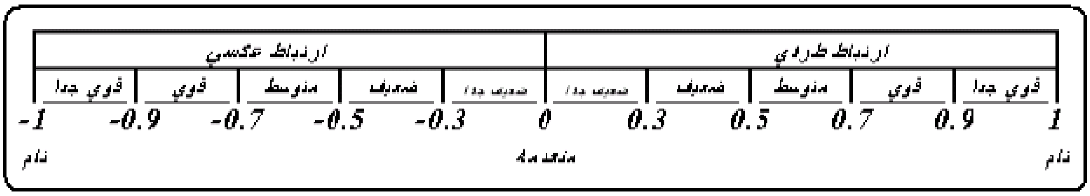
إذا كانت إشارة معامل الارتباط سالبة ($r < 0$) توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه انخفاض في المتغير الثاني، والعكس.

إذا كانت إشارة معامل الارتباط موجبة ($r > 0$)توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الثاني، والعكس .

إذا كان معامل الارتباط قيمته صفرا ($r = 0$) دل ذلك على انعدام العلاقة بين المتغيرين.

قوة العلاقة:. ويمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن (± 1)، حيث أن قيمة معامل الارتباط تقع في المدى $(-1 < r < 1)$، وقد صنف بعض الإحصائيين درجات لقوة العلاقة يمكن تمثيلها على الشكل التالي

درجات قوة معامل الارتباط



معامل الارتباط الخطي البسيط Simple Linear Correlation Coefficient

يعرف معامل الارتباط الخطي البسيط بأنه القيمة العددية للعلاقة الخطية بين متغيرين و تحسب من القانون الآتي

$$r_{xy} = \frac{n\sum_{i=1}^{n} X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{\left\{(n\sum X_i^2 - (\sum X_i)^2)(n\sum Y_i^2 - (\sum Y_i)^2)\right\}}}$$

معامل الارتباط الجزئي Partial Correlation Coefficient

يلاحظ في بعض الأحيان و جود ارتباط بين متغيرين يعزي جزئيا إلى ارتباط متغير ثالث مرتبط مع كلا المتغيرين . لنفرض إن المتغيرين X1,X2 مرتبطين و هناك متغير أخر X3 مرتبط مع كلا المتغيرين و إننا نرغب في قياس درجة الارتباط بين X1,X2 باستبعاد اثر المتغير الثالث X3 على كلا المتغيرين.إن المعامل الذي يقيس درجة ارتباط متغيرين باستبعاد اثر الثالث يدعى بمعامل الارتباط الجزئي . على سبيل المثال الارتباط بين طول الفرد X1 و وزنه X2 يتأثر بارتباط عمر الفرد X3 مع كل من X1,X2 ، كذلك الارتباط بين دخل الأسرة الشهري X1 و إنفاقها الشهري X2 يتأثر بارتباط عدد أفراد الأسرة X3 مع كل من X1,X2 ، على سبيل المثال عدد السكائر المدخنة X1 يوميا و الإصابة بنوع معين من أمراض الرئة X2 يتأثر بارتباط عمر المدخن X3 و عدد سنوات التدخين X4 مع كل من X1,X2. و يمكن إيجاد قيمة معامل الارتباط الجزئي وفق ما يلي: افرض إن X1,X2,X3 تمثل ثلاثة متغيرات عشوائية بحيث أن المتغيرين X1,X2 مرتبطين و أن X3 مرتبط مع كل من X1,X2 و افرض أن على أساس عينة من المفردات قوامها n تم الحصول على القياسات المتناظرة لهذه المتغيرات

و أن معامل الارتباط ((X1i , X2i , X3i), i=1,2,3........n) البسيط بين متغيرين منهما هو r12 , r13 , r23 عندئذ فان معامل الارتباط الجزئي بين X1,X2 باستبعاد اثر الثالث X3 هو

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} \qquad ....1$$

و في حالة وجود أربعة متغيرات و نرغب في حساب الارتباط الجزئي بين الأول و الثاني باستبعاد اثر الثالث و الرابع فان هذا المعامل هو

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}\, r_{24.3}}{\sqrt{(1 - r_{14.3}^2)(1 - r_{24.3}^2)}} \qquad ....2$$

المحاضرة الحادية والعشرون

معامل الارتباط الرتب لسبيرمان Spear mans rank correlation Coefficient

افرض إن X,Y متغيرين من النوع الوصفي و افرض إن البيانات المستحصل عليها من X,y على أساس عينة عشوائية من المفردات قوامها n هي بهيئة صفات غير قابلة للقياس الكمي .

افرض إن (Xi , Yi , i=1,2,3,.....n) ممكنة الترتيب تصاعديا أو تنازليا وفق معيار معين يمتاز به كل متغير (مثلا تقديرات درجات مجموعة من الطلبة يمكن ترتيبها تصاعديا على أساس معيار الأقل إلى

اعلي درجة أو العكس) عندئذ استنادا لهذا الترتيب و لكل متغير يمكن تخصيص قيم سلسلة الإعداد الطبيعية (1,2,3,......,n) لصفات الترتيب بحيث إن كل صفة يخصص لها احد إعداد هذه السلسة ، في حالة عدم تكرار أيه صفة منها (وقد يكون التخصيص تنازلي) و سوف نوضح حالة التكرار بعض الصفات و أسلوب علاج ذلك في أمثلة. و القانون المستخدم لإيجاد الارتباط لمثل هذه الحالات:

$$r_{xy} = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \qquad , \quad d_i = X_i - Y_i$$