
المحاضرة الاولى

الإحصاء (Statics)

هو العلم الي يستخدم الأسلوب العلمي في طرق جمع البيانات وتبويبها وتلخيصها وعرضها وتحليلها بهدف الوصول الى استنتاجات وقراءات معينة ومناسبة.

يقسم علم الإحصاء الى قسمين رئيسيين هما: -

١. **الإحصاء الوصفي: -**

ويعتمد الطرق الإحصائية في وصف مجموعة معينة من البيانات حيث تتضمن هذه الطرق أساليب جمع البيانات بصورة عددية وتبويبها وتنظيمها وتلخيصها وعرضها وحساب بعض المقاييس الإحصائية المختلفة لها.

٢. **الإحصاء الاستنتاجي او الاستدلالي: -**

ويعتمد الطرق الإحصائية التي تهدف ال عمل استنتاجات او استدلالات حول المصدر الذي جمعت منه البيانات. ويضم فرعين رئيسيين هما: -

ا- التقدير (Estimation): -

ويهتم بإيجاد قيم تقديرية للاستدلال منها على القيم الحقيقية للمصدر وجمع البيانات وهذه القيم اما تكون تقديرا محددا عند نقطة معينة او تقدير ضمن فترة او مدى معين.

ب- اختبار الفرضيات (Test of Hypothesis): -

يتضمن اختبار الفرضيات التي توضع كتقدير اولي للظاهرة المراد دراستها للوصول منها لقرار قبول الفرضية او رفضها.

أهمية علم الإحصاء

تبرز أهمية علم الإحصاء من خلال الوظائف التي يقوم بها والتي من أهمها: -

١. **الدقة (definiteness): -** حيث يعرض البيانات والحقائق بصورة واضحة ومحددة.
٢. **التلخيص (condensation): -** حيث يتم تلخيص البيانات الكثيرة بقيم قليلة ذات معنى.
٣. **المقارنة (comparison): -** أي وضع الأسس لمقارنة العوامل العائدة لنفس الظاهرة.
٤. **صياغة واختبار الفرضيات (formulating and testing of hypothesis): -** حيث تستخدم طرق إحصائية متنوعة لصياغة واختبار الفرضيات وتطوير نظريات جديدة.
٥. **التنبؤ او التكهن (predication): -** هو التنبؤ باتجاه وقيمة ظاهرة معينة خلال فترة زمنية مستقبلية.
٦. **التخطيط (planning): -** حيث يساعد في وضع الخطط والقرارات المناسبة من قبل مؤسسات الدولة لاتخاذ السياسة المناسبة للقطاعات المختلفة حيث يوفر البيانات اللازمة ويحدد حجمها واتجاه التغير فيها.

* بعض المصطلحات الإحصائية

١. **المشاهدة (observation)**

هي المادة الأولية التي يتعامل معها الباحث. مثلا إذا أردنا معرفة عدد اطوال الطلبة في كلية التربية الأساسية لفرع معين لقسم معين فإننا نختار عددا من الطلاب ونقيس طول كل طالب منهم فاذا كان عدد اطوالهم مثلا مساوي للـــ ١٦٥ سم فان هذا العدد يمثل مشاهدة واحدة وهكذا.

٢. المتغير (variable)

لو درسنا صفة ارتفاع شجرة لمجموعة معينة من الأشجار فسنجد اختلافا في الارتفاعات وفي هذه الحالة يطلق على صفة ارتفاع الشجرة بمصطلح متغير أي ان المتغير يرمز الى الصفة التي تتغير قيمتها من فرد لآخر ويمكن ان تصنف المتغيرات الى نوعين هما: -

ا. متغيرات وصفية او نوعية (qualitative variables)

هي متغيرات او قيم غير عددية عادة او بمصـــطلح اخر لا يمكن وصــــفها بأرقام مثل لون الزهرة، الجنس ذكر او انثى، صفة الشكل، مدى الإصابة (خفيفة، متوسطة، كبيرة).

ب. متغيرات كمية (quantitative variables)

هي متغيرات ذات قيم عددية او بمصــطلح اخر يمكن وصــفها وقياســها بأرقام مثل الطول، الوزن، الحجم وتأخذ وحدات قياس وتقسم الى قسمين: -

- **متغيرات مستمرة (continuous variables)**

هي المتغيرات التي تأخذ المشاهدة فيها أي قيمة رقمية ضمن مدى معين مثل صفة الطول وتقبل هذه القيم الكسور.

- **متغيرات غير مستمرة (discrete variables)**

هي المتغيرات التي تأخذ المشاهدة فيها قيم متباعدة مثل عدد الطلاب في مدرسة معينة ولا تقبل هذه القيم الكسور.

٣. القيمة الإحصائية (variate)

ويقصــد بها القيمة الخاصـــة بمفردة ما نسـبة الى صــفة معينة مثلا لو قسـنا اطوال أشـجار معينة ونرمز لمتغير الطول بالحرف (X) فان طول الشـجرة رقم (٥) يعتبر قيمة إحصـائية ويرمز لها (X_5) فاذا كان طول الشـجرة رقم (٥) يسـاوي (١٠) متر فيمكن التعبير عن هذه القيمة الإحصـائية رياضيا كالاتي $X_5=10$ وهكذا...

٤. المجتمع (population)

احصائيا يعرف المجتمع بانه جميع الافراد او العناصر التي تشترك في صفة متغيرة واحدة او أكثر تميزه تمييزا تاما عن المجتمعات الأخرى فمثلا حدائق كلية التربية الأسـاسـية تعتبر مجتمعا احصـائيا ولا يمكن تعميمها نتائج المشـاهدات فيها على حدائق الكليات الأخرى.

٥. العينة (sample)

هي جزء من المجتمع مأخوذة بطريقة معينة في حالة عدم إمكانية الحصــول على قيم جميع افراده لأســباب مادية او فنية مثل قيم اوزان جميع طلبة مدرسة معينة.

***بعض الرموز الإحصائية**

- $\sum$ هو حرف اغريقي يسمى **sigma** ويعني الجمع **(summation).**
- $\sum_{i=1}^{n}$ يعني الجمع للمفردات من ١ ال n حيث n عدد العينات.

$\sum_{i=1}^{n}$ i= 1, 2, 3......... n

- $\sum_{i=1}^{n} Xi$ يعني الجمع للمفردات من رقم ١ ال العينة رقم n.

$\sum_{i=1}^{n} Xi = (X_1 + X_2 + X_3 + \dots X_n)$

- $\sum_{i=5, i\neq 7}^{8} Xi$ يعني الجمع للمفردات من المفردة رقم ٥ الى المفردة رقم ٨ باستثناء المفردة رقم ٧.

$\sum_{i=5, i\neq 7}^{8} Xi = X_5 + X_6 + X_8$

- $\sum_{i=1}^{n} X^2 i$ يعني جمع مربع المفردات ابتداء من رقم ١ الى المفردة رقم n.

$\sum_{i=1}^{n} X^2 i = (X^2{}_1 + X^2{}_2 + X^2{}_3 + \dots X^2{}_n)$

- $(\sum_{i=1}^{n} Xi)^2$ يعني مربع المجموع الكلي للمفردات.

$(\sum_{i=1}^{n} Xi)^2 = (X_1 + X_2 + X_3 + \dots X_n)^2$

- $\sum_{i=1}^{n} Xi * Yi$ يعني مجموع قيم حاصل ضرب قيم المتغيرين X,Y

$\sum_{i=1}^{n} Xi * Yi = X_1 * Y_1 + X_2 * Y_2 + X_3 * Y_3 + \dots X_n * Y_n$

- $(\sum_{i=1}^{n} Xi) * (\sum_{j=1}^{m} Yj)$ ويعني حاصل ضرب مجموع قيم المتغيرين X,Y

$(\sum_{i=1}^{n} Xi) * (\sum_{j=1}^{m} Yj) = (X_1 + X_2 + X_3 + \dots X_n) * (Y_1 + Y_2 + Y_3 + \dots Y_m)$

<u>*بعض القواعد المفيدة في عملية الجمع</u>

<u>قاعدة (١)</u> اذا كانت (c) أي عدد ثابت فأن

$\sum_{i=1}^{n} c = c_1 + c_2 + c_3 + \dots + c_n = c*n$

<u>قاعدة (٢)</u> اذا كانت (c) أي عدد ثابت فأن

$\sum cyi = c \sum yi$

$\sum cyi = c\, y_1 + c\, y_2 + c\, y_3 + \dots + c\, y_n = c\,(y_1 + y_2 + y_3 \dots + y_n) = c \sum yi$

<u>قاعدة (٣)</u> جمع قيمة متغيرين او اكثر هو مجموع جمعهم

$\sum (Xi + yi) = \sum Xi + \sum yi$

$\sum (Xi + yi) = (X_1 + Y_1) + (X_2 + Y_2) + (X_3 + Y_3) + \dots + (X_n + Y_n)$

$= (X_1 + X_2 + X_3 + \dots + X_n) + (Y_1 + Y_2 + Y_3 + \dots + Y_n)$

$= \sum Xi + \sum yi$

مثال ١ / إذا كانت اعمار خمسة طلاب كالاتي: 20, 18, 24, 22, 16

اوجد كل مما يأتي: -

1) $(\sum_{i=1}^{5} Xi)^2$ 2) $\sum_{i=1}^{5} (Xi)^2$ 3) $\sum_{i=1}^{3} Xi$

Solution: -

$X_1=20,$ $X_2=18$ $X_3=24$ $X_4=22$ $X_5=16$

1-$(\sum_{i=1}^{5} Xi)^2 = (X_1+X_2+X_3+X_4+X_5)^2$
$= (20+18+24+22+16)^2$
$= (100)^2$
$=10000$

2- $\sum_{i=1}^{5} (Xi)^2 = (X_1)^2+(X_2)^2+(X_3)^2+(X_4)^2+(X_5)^2$
$= (20)^2+(18)^2+(24)^2+(22)^2+(16)^2$
$= 2040$

3- $\sum_{i=1}^{3} Xi = (X_1+X_2+X_3)$
$= 20 + 18 + 24$
$= 62$

مثال ٢ / إذا كانت قيم $(X_i = 3, 4, 1)$ و $(Y_i = 5, 2, 6)$

اوجد كلا مما يلي: -

1- $\sum_{i=1}^{3} Xi * \sum_{i=1}^{3} Yi$
2- $\sum_{i=1}^{3} Xi * Yi$

Solution: -

1- $\sum_{i=1}^{3} Xi * \sum_{i=1}^{3} Yi = (X_1+X_2+X_3) * (Y_1+Y_2+Y_3)$
$= (3+4+1) *(5+2+6)$
$= (8) * (13) = 104$

2- $\sum_{i=1}^{3} Xi * Yi = (X_1*Y_1+X_2*Y_2+X_3*Y_3)$
$= (3)(5) + (4)(2) + (1)(6)$
$= 15 + 8 + 6 =29$

مثال ٣/ إذا علمت ان

$X_i=2, 6, 3, 1$ و $Y_i=3, 9, 6, 2$

اوجد قيمة كل مما يأتي: -

1- $\sum_{i=1}^{4}(Yi - Xi)^2$ 2- $\sum_{i=1}^{4}(Xi - 3)(Yi - 5)$ 3- $\sum_{i=1}^{3} Xi * Yi^{\,2}$

Solution: -

1- $\sum_{i=1}^{4}(Yi - Xi)^2$ $= (Y_1 - X_1)^2 + (Y_2 - X_2)^2 + (Y_3 - X_3)^2 + (Y_4 - X_4)^2$

$= (3 - 2)^2 + (9 - 6)^2 + (6 - 3)^2 + (2 - 1)^2$

$= (1)^2 + (3)^2 + (3)^2 + (1)^2$

$= 1 + 9 + 9 + 1$

$= 20$

2- $\sum_{i=1}^{4}(Xi - 3)(Yi - 5)$ $= (X_1 - 3)(Y_1 - 5) + (X_2 - 3)(Y_2 - 5) + (X_3 - 3)(Y_3 - 5) + (X_4 - 3)(Y_4 - 5)$

$= (2 - 3)(3 - 5) + (6 - 3)(9 - 5) + (3 - 3)(6 - 5) + (1 - 3)(2 - 5)$

$= (-1)(-2) + (3)(4) + (0)(1) + (-2)(-3)$

$= 2 + 12 + 0 + 6$

$= 20$

3- $\sum_{i=1}^{3} Xi * Yi^{\,2}$ $= X_1 (Y_1)^2 + X_2 (Y_2)^2 + X_3 (Y_3)^2$

$= 2(3)^2 + 6(9)^2 + 3(6)^2$

$= 18 + 486 + 108 \qquad = 612$

المصدر

كتاب المدخل الى الإحصاء

كلية الزراعة والغابات

جامعة الموصل

المؤلف خاشع محمود الراوي

المحاضرة الثانية

عرض البيانات Data Presentation

مقدمة: بعد جمع البيانات يكون من الصعب دراستها وفهمها دون تنظيمها وجدولتها وعرضها بشكل يسهل على الباحث الوصول الى النتائج المطلوبة وكثيرا ما يكون هدف الباحث من عرض البيانات هو جذب انتباه القارئ نحو اتجاه هذه البيانات بالزيادة او النقصان او العلاقة بين المتغيرات التي يدرسها او المقارنة بين المجاميع من البيانات لذا يقوم الباحث بتبسيطها وذلك بعرضها بأشكال معبرة وهادفة.

*طرق عرض البيانات Methods of Data Presentation

هناك عدة طرق لعرض البيانات منها: -

أولا) العرض الجدولي (Tabular Presentation)

يعرف الجدول **(table)** بانه طريقة منظمة لعرض البيانات العددية بشكل أعمدة راسية و صفوف افقية حسب عدد الفئات المطلوبة في تصنيف البيانات ويعبر الجدول عن الكثير من المعلومات الت يمكن ملاحظتها مباشرة ولا داعي لتكرارها في النص اللغوي بل يكتفي بالإشارة الى محتوى الجدول ونوع البيانات التي يتضمنها ولابد ان يتوفر في الجدول الوضوح من حيث عدد الصفوف والاعمدة ونوع البيانات والرموز المستخدمة فيه وكذلك كتابة عناوين راسية تحدد نوع البيانات في كل عمود وكذلك كتابة عناوين تحدد نوع البيانات في كل صف افقي إضافة الى كتابة عنوان رئيسي فوق او تحت الجدول وترقيم الجدول ان كان هناك اكثر من جدول. وهناك نوعان رئيسيان من الجداول هي: -

١- **الجدول البسيط (simple table)**

تعتبر هذه الجداول أكثر أنواع الجداول بدائية وسهولة اذ يحتوي فيها الجدول على عمودين حيث يمثل العمود الأول تقسيمات الصفة او الظاهرة او الأسماء او المفردات او العناصر ويبين العمود الثاني عدد المفردات الثابتة لكل فئة او مجموعة ذات العلاقة بكل مفردة او عنصر. مثلا جدول يبين درجات الطلبة او تصنيف دول العالم حسب عدد سكانها وهكذا...

مثال (١) جدول يمثل توزيع (١٠٠) طالب في صف معين في مدرسة معينة حسب اوزانهم

عدد الطلبة	فئات الوزن
5	**50-52**
15	**53-55**
45	**56-58**
27	**59-61**
8	**62-64**
100	**المجموع**

٢- الجدول المركب (Complex Table)

هو نوع من الجداول تأتي فيه البيانات حسب صفتين او ظاهرتين او أكثر في نفس الوقت. ويتألف الجدول المركب من صفتين من:

- **الصفوف: -** تمثل فئات او مجاميع احدى الصفتين.
- **الاعمدة: -** تمثل فئات او مجاميع الصفة الأخرى.

فالخلايا التي تقابل الصفوف والاعمدة تمثل عدد المفردات او التكرارات المشتركة بين الفئات او المجاميع لكلا الصفتين مثال على ذلك الجدول الاتي الذي يمثل العلاقة بين طول الطلاب واوزانهم في مدرسة معينة: -

الوزن كغم / الطول سم	51-60	61-70	71-80	المجموع
121-140	20	6	4	30
141-160	10	2	40	52
161-180	2	6	10	18
المجموع	32	14	54	100

*جدول التوزيع التكراري (frequency distribution table)

هو جدول بسيط يتكون من عمودين حيث: -

- **العمود الأول: -** يقسم فيه قيم المتغير الى اقسام او مجموعات تدعى الفئات (classes)
- **العمود الثاني: -** يبين اعداد كل فئة متكررة ويسمى بالتكرار(frequency)

ومثال على ذلك الجدول التالي الذي يمثل العلاقة بين عدد النباتات واطوالهن. حيث يمثل الجدول أدناه توزيع تكراري لأطوال (٨٠) نباتا من القطن بالسنتمترات.

فئات الطول (سم)	التكرار (عدد النباتات)
31-40	1
41-50	2
51-60	5
61-70	15
71-80	25
81-90	20

*بعض التعاريف المهمة

١- البيانات غير المبوبة (ungrouped data)

هي البيانات الأولية او الاصلية التي جمعت ولم توضع في جدول توزيع تكراري.

٢- **البيانات المبوبة (grouped data)**

هي البيانات التي نظمت في جدول توزيع تكراري.

٣- **الفئات (classes): -**

هي الفترات او المديات التي قسمت اليها قيم المتغير حيث كل فئة تأخذ مدى معين من قيم المتغير.

٤- **حدود الفئات (class limits): -**

هي نقطة بداية المدى ونقطة نهايته لكل فئة من الفئات. حيث لكل فئة حد أدنى للفئة (lower class limit) وحد اعلى للفئة (upper class limit).

٥- **الحدود الحقيقية للفئات (true class limit): -**

لكل فئة حدان حقيقيان، حد أدنى حقيقي (lower class boundary) وحد اعلى حقيقي (upper class boundary).

٦- **طول الفئة (class length): -**

هو مقدار المدى بين حدي الفئة الأعلى والادنى ويفضل ان تكون اطوال الفئات متساوية لتسهيل العملية الحسابية.

٧- **مركز الفئة (class mid-point): -**

لكل فئة مركز وهو عبارة عن منصف المدى بين حدي الفئة الأعلى والادنى.

٨- **تكرار الفئة (class frequency): -**

هي عدد المفردات او القيم التي تقع ضـــمن مدى تلك الفئة ويرمز لها بالرمز (f_i) وان مجموع التكرارات يجب ان يكون مساويا دائما للعدد الفعلي لقيم الظاهرة او المتغير او المفردة.

*حساب طول الفئة

اولا) عندما تكون حدود الفئات اعداد صحيحة فقط

طول الفئة (L) = الحد الأعلى للفئة – الحد الأدنى للفئة + 1

مثلا إذا كانت لدينا الفئة التالية 61-70 فان طول الفئة هو

طول الفئة = 70 – 61 + 1 = 10

ثانيا) عندما تكون حدود الفئات اعداد حقيقية

طول الفئة (L) = الحد الحقيقي الأعلى للفئة – الحد الأدنى الحقيقي للفئة

مثلا إذا كانت لدينا الفئة التالية 60.5 - 70.5 فان طول الفئة هو

طول الفئة = 70.5 – 60.5 = 10

ثالثا) طول الفئة = الفرق بين الحدين الأدنى او الحدين الأعلى لفئتين متتاليتين:

مثلا إذا كانت لدينا الفئتين المتتاليتين: - 51 – 60

---

61 – 70

10 = 60 – 70= طول الفئة (L) او 10 = 51 – 61 = طول الفئة (L)

رابعا) طول الفئة = الفرق بين الحدين الحقيقيين الأدنى او الحدين الحقيقيين الأعلى لفئتين متتاليتين:

مثلا إذا كانت لدينا الفئتين المتتاليتين: - 40.5 – 50.5

50.5 – 60.5

10 = 50.5 – 60.5 = طول الفئة (L) او 10 = 40.5 – 50.5 = طول الفئة (L)

خامسا) طول الفئة = الفرق بين مركزي فئتين متتاليتين

مثلا لتكن مركزي فئتين متتاليتين كالاتي: - 75.5

85.5

فان طول الفئة يحسب كالتالي: 10 = 75.5 - 85.5 = L طول الفئة

*حساب الحدود الحقيقية لأي فئة: -

يمكن حساب الحدود الحقيقية لأي فئة بطرق عدة منها: -

(طول تلك الفئة) ½ - مركز تلك الفئة = الحد الأدنى الحقيقي لأي فئة

او

(طول تلك الفئة) ½ + مركز تلك الفئة = الحد الاعلى الحقيقي لأي فئة

او

$$\text{الحد الأدنى الحقيقي لأي فئة} = \frac{\text{الحد الاعلى للفئة السابقة + الحد الادنى تلك الفئة}}{2}$$

او

$$\text{الحد الأعلى الحقيقي لأي فئة} = \frac{\text{الحد الادنى للفئة التالية + الحد الاعلى تلك الفئة}}{2}$$

مثال / في الجدول التالي احسب الحد الأدنى والحد الأعلى وطول الفئة للفئة الثالثة

الفئات	التكرار
31-40	5
41-50	21
51-60	12
61-70	17

الحل /

1 + الحد الأدنى للفئة الثالثة – الحد الأعلى للفئة الثالثة = طول الفئة الثالثة (L)

L (3) = 60 - 51 +1 =10

الحد الأدنى الحقيقي للفئة الثالثة $= \dfrac{\text{الحد الاعلى للفئة الثانية + الحد الادنى الفئة الثالثة}}{2}$

$= \dfrac{51+50}{2} = 50.5$

الحد الأعلى الحقيقي للفئة الثالثة $= \dfrac{\text{الحد الادنى الفئة الرابعة + الحد الاعلى الفئة الثالثة}}{2}$

$= \dfrac{60+61}{2} = 60.5$

اذن الحد الأدنى والاعلى الحقيقيان للفئة الثالثة هما: - **50.5 – 60.5**

واجب: - اوجد الحدود الدنيا والعليا لباقي الفئات واعد صياغة الجدول التكراري.

*حساب مركز الفئة

1- **مركز الفئة** $= \dfrac{\text{الحد الاعلى للفئة + الحد الادنى للفئة}}{2}$

2- **مركز الفئة** $= \dfrac{\text{الحد الاعلى الحقيقي للفئة + الحد الادنى الحقيقي للفئة}}{2}$

مثال / في الجدول السابق اوجد مركز الفئة الثالثة؟

مركز الفئة الثالثة $= \dfrac{\text{الحد الاعلى للفئة الثالثة + الحد الادنى للفئة الثالثة}}{2}$

$= \dfrac{51 + 60}{2} = 55.5$

او

مركز الفئة الثالثة $= \dfrac{\text{الحد الاعلى الحقيقي للفئة الثالثة + الحد الادنى الحقيقي للفئة الثالثة}}{2}$

$= \dfrac{50.5 + 60.5}{2} = 55.5$

***ملاحظة: - تكرار الفئة يعني جميع قيم المتغير الواقعة في مدى الفئة، فاذا كان تكرار الفئة (15) والفئة هي 60 - 80 فهذا يعني ان هناك (15) قيمة من قيم المتغير واقعة في المدى 60 – 80.**

*** التوزيع التكراري النسبي (Relative Frequency Distribution): -**

ويبين هذا التوزيع الأهمية النسبية لكل فئة من الفئات ويتم حساب التكرار النسبي لكل فئة كالاتي: -

$$\textbf{التكرار النسبي لأي فئة} = \frac{\text{تكرار تلك الفئة } fi}{\text{المجموع الكلي للتكرارات } \varepsilon}$$

مثال / إذا كانت مجموع التكرارات في جدول تكراري هو (80) وان تكرار الفئة الرابعة (15)، اوجد التكرار النسبي للفئة الرابعة؟

الحل: -

$$\textbf{التكرار النسبي للفئة الرابعة} = \frac{\text{تكرار الفئة الرابعة}}{\text{المجموع الكلي للتكرارات } \varepsilon}$$

$$= \frac{\mathbf{15}}{\mathbf{80}} \quad = 0.1875$$

وعادة يوضع التكرار النسبي **كنسبة مئوية** وذلك بضرب التكرار النسبي **بـ 100 %** وكما هو موضح في الجدول التالي: -

الفئات	التكرار (*fi*)	التكرار النسبي (*R.fi*)	التكرار النسبي المئوي %
31 – 40	1	0.0125	1.25
41 – 50	2	0.025	2.50
51 – 60	5	0.0625	6.25
61 – 70	15	0.1875	18.75
71 – 80	25	0.3125	31.25
81 – 90	20	0.25	25
91 – 100	12	0.15	15
	80	1	100%

ملاحظة: - يجب ان يكون مجموع التكرار النسبي لجميع الفئات = (1) عدد صحيح، والتكرار المئوي = (100) %

المصدر

كتاب المدخل الى الإحصاء

كلية الزراعة والغابات

جامعة الموصل

المؤلف خاشع محمود الراوي

المحاضرة الثالثة

ثالثا: -العرض البياني (Graphical Presentation)

ان الرسوم والصور والاشكال الهندسية ماهي الا تعبير وتوضيح للبيانات بطريقة جذابة وسهلة وفعالة تساعد القارئ على فهم واستيعاب قيم الظاهرة ومقارنتها مع بعضها.

ان وسائل العرض او التمثيل البياني كثيرة ومتنوعة وسنكتفي بشرح العرض البياني للتوزيعات التكرارية فقط.

عادة يخصص **المحور السيني (x)** لتمثيل قيم او فئات المتغير بينما يخصص **المحور الصادي (y)** لتمثيل التكرارات لهذا المتغير ويجب دائما ان يبدأ بتدرجه من الصفر كما انه ليس من الضروري ان يكون مقياس او تدرج المحورين من نفس المقياس ويشمل العرض البياني الأنواع التالية: -

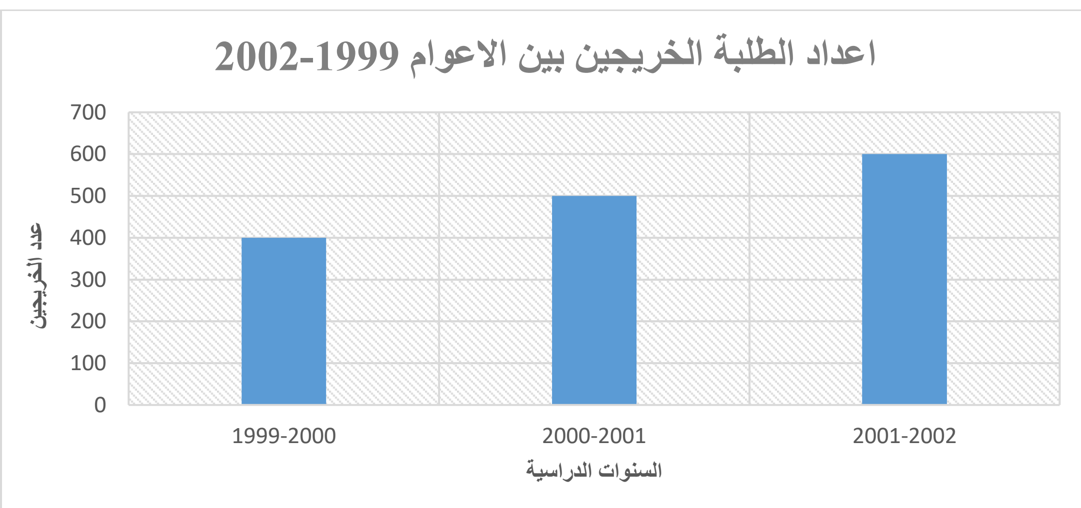
١) المدرج التكراري. (Histogram)

تعتبر العروض البيانية بطريقة المدرج التكراري من أكثر الأنواع التي تستخدم من قبل المؤسسات التي تهتم بعرض الاحصائيات إضافة الى الاستخدام في البحوث والدراسات والتقارير المختلفة.

والمدرج التكراري هو عبارة عن **مســــتطيلات راســـية او عمودية تمتد قواعدها على المحور الافقي** لتمثل اطوال الفئات بينما **ارتفاعاتها تمثل تكرارات الفئات.**

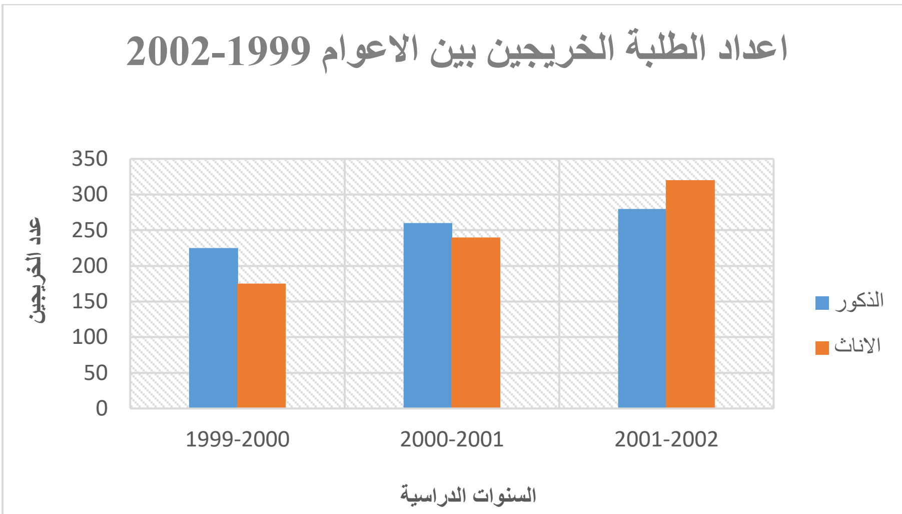
ويلاحظ على هذه الطريقة انها **تختلف حســب الهدف الذي يراد تحقيقه وطبيعة البيانات المتوفرة** فمثلا إذا كان الهدف هو **بيان التطور الزمني للحدث او للظاهرة** فيمكن ان تمثل البيانات على شكل **مستطيلات يمثل المحور السيني متغير الزمن** وان **ارتفاعاتها تمثل تطور الحدث.**

مثال: - ارسم المدرج التكراري لأعداد خريجي احدى الكليات خلال السنوات ١٩٩٩-٢٠٠٢.



اما إذا كان الهدف **المقارنة بين ظاهرتين او فئتين وأكثر** فيفضل رسم أعمدة او مستطيلات متلاصقة للظواهر التي يراد مقارنتها وفقا لتطورها الزمني ويفضل استخدام ألوان متباينة لكل ظاهرة او فئة من الظواهر لغرض تمييزها بسهولة.

مثال: - التوزيع التالي يمثل اعداد الخريجين لإحدى الكليات من الذكور والاناث خلال السنوات ١٩٩-٢٠٠٢.

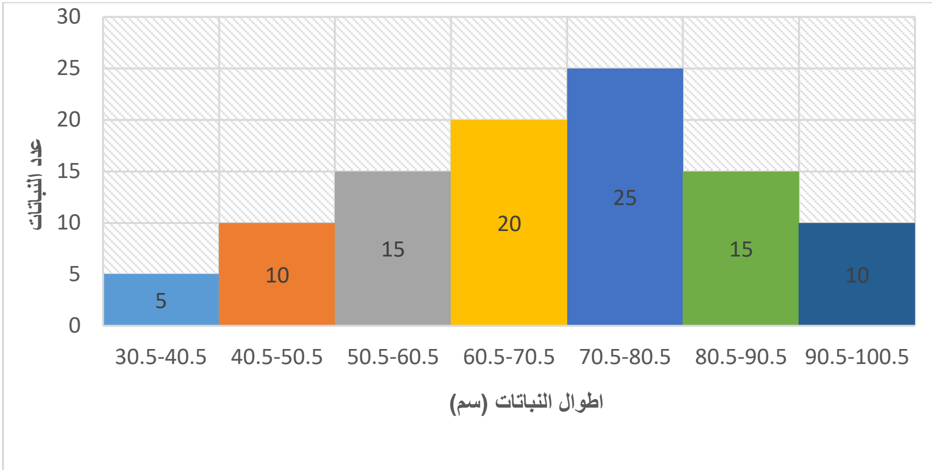


ولرسم المدرج التكراري نتبع الخطوات التالية: -

١- رسم المحور الافقي والعمودي
٢- يقسم المحور الافقي الى اقسام متساوية بمقياس رسم مناسب يشمل **الحدود الحقيقية للفئات** ويفضل ترك **مسافة صغيرة بين نقطة الصفر والحد الأدنى للفئة الأولى.**
٣- يقسم المحور العمودي **بقيم التكرارات** ويجب ان يكون التدريج يتضمن **اقل قيمة الى اعلى قيمة.**

مثال/ ارسم المدرج التكراري للجدول الاتي: -

الفئة (اطوال النباتات بالسنتمتر)	عدد النباتات (التكرار f_i)
30.5 - 40.5	5
40.5 - 50.5	10
50.5 - 60.5	15
60.5 - 70.5	20
70.5 - 80.5	25
80.5 - 90.5	15
90.5 – 100.5	10



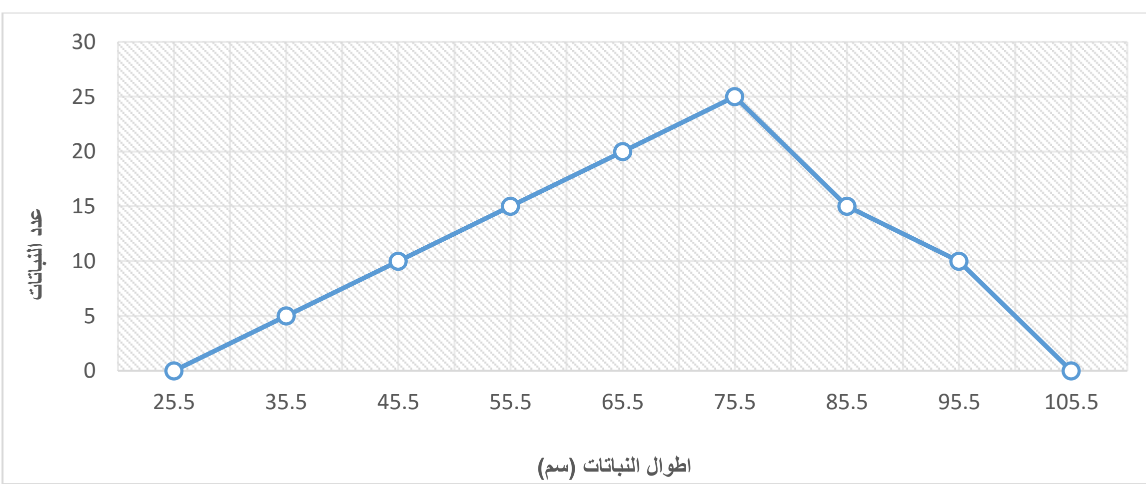
٢) المضلع التكراري (Frequency Polygon)

هو عبارة عن **خطوط مستقيمة تصل بين نقاط** كل منها واقع **فوق مركز فئة** على **ارتفاع يمثل تكرار الفئة** وعادة **يقفل المضلع التكراري** بتوصيل بداية المضلع بالمحور الافقي **بمركز فئة واقعة الى يسار اول فئة يكون تكرارها صفرا**، وتوصيل نهاية المضلع بالمحور الافقي **بمركز فئة واقعة الى يمين اخر فئة يكون تكرارها صفرا** وبذلك تكون مساحة المضلع التكراري **مساوية لمساحة المدرج التكراري**.

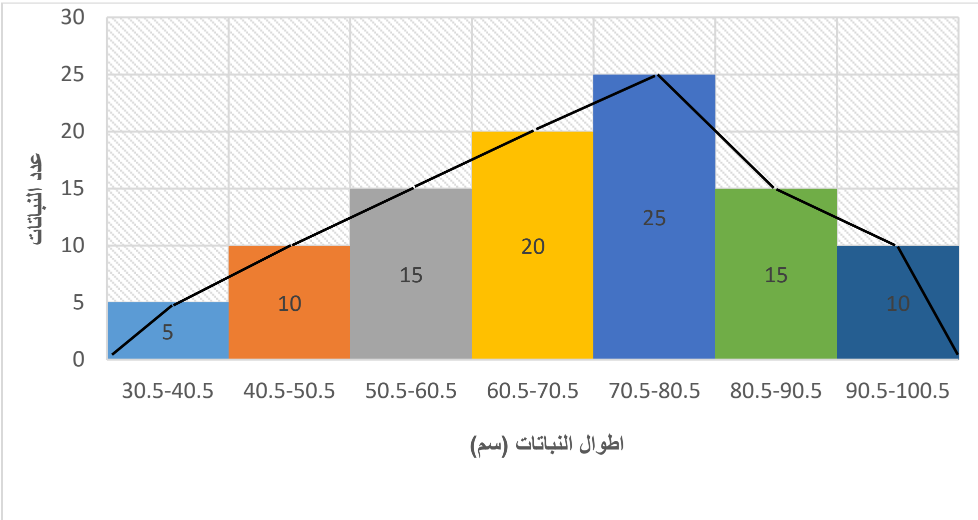
ولرسم المضلع التكراري نتبع الخطوات التالية: -

١- رسم المحورين الافقي والعمودي.
٢- تقسيم المحور الافقي الى اقسام متساوية بحيث يشمل **جميع مراكز الفئات** ويقسم المحور العمودي الى اقسام متساوية بحيث يشمل التكرارات جميعها.
٣- وضع نقطة في موضع تقاطع المحورين **لمراكز الفئات وتكراراتها.**
٤- توصيل تلك النقاط **بخطوط مستقيمة**.

مثال/ ارسم المضلع للجدول في المثال السابق.



ملاحظة: - يمكن رسم المضلع التكراري باستعمال المدرج التكراري وذلك **بتنصيف القواعد العليا للمستطيلات** (والتي تمثل **مراكز الفئات**) بنقاط ثم **توصيل هذه النقاط بمستقيمات** والمثال التالي يبين ذلك.



٣) العرض البياني الدائري (The Pie Graphic)

في هذه الطريقة للعرض البياني يجري تمثيل جميع الأجزاء في دائرة كاملة **وذلك لتحديد نسبة كل جزء الى الكل**. وان هذه الطريقة تختلف عن أسلوب المستطيلات او الاعمدة وذلك لان القيم التي تحملها هذه الاعمدة **هي قيم حقيقية** بينما تمثل الأجزاء التي تنقسم فيها الدائرة **نسبة كل منها للمجموع الكلي** لذلك تعتبر هذه الطريقة مهمة لأننا نستطيع باستخدامها ان **نقارن الأجزاء** مع بعضها.

والطريقة التي تتبع في تقسيم الدائرة تتضمن الخطوات الاتية: -

١- تحديد قيم الأجزاء والمجموع الكلي لقيم الأجزاء.
٢- استخدام زاوية القطاع لدائرة (زاوية الجزء) من خلال المعادلة الاتية: -

$$\text{زاوية الجزء} = \frac{\text{قيمة الجزء}}{\text{قيمة المجموع الكلي للاجزاء}} \times 360^{\circ}$$

مثال/ في احدى الجامعات كانت الدرجات الاكاديمية لأعضاء هيئتها التدريسية موزعة اعدادها كما في الجدول التالي: -

الدرجة الاكاديمية (الفئات)	العدد (التكرار)
أستاذ	100
أستاذ مساعد	300
مدرس	600
المجموع	1000

المطلوب تمثيل هذه البيانات **بعرض دائري**

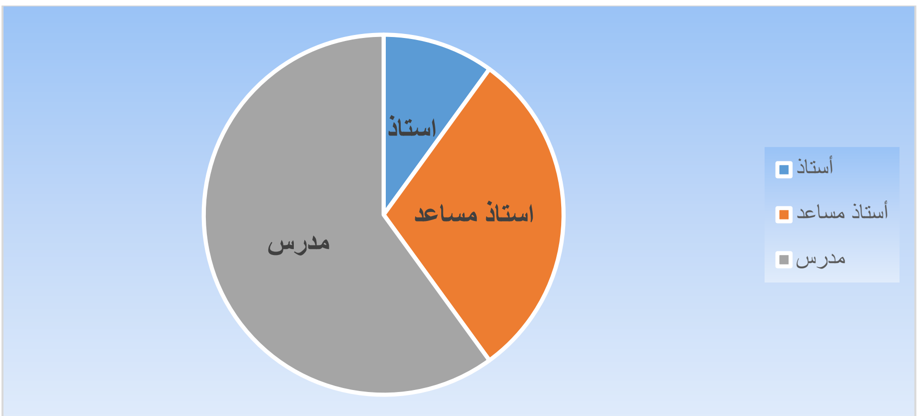
الحل: - **أولا)** نستخرج زوايا أجزاء الدائرة: -

زاوية القطاع لدرجة **أستاذ** = $\frac{\text{قيمة الجزء}}{\text{قيمة المجموع الكلي للاجزاء}} * 360^\circ = \frac{100}{1000} * 360^\circ = 36^\circ$

زاوية القطاع لدرجة **أستاذ مساعد** = $\frac{\text{قيمة الجزء}}{\text{قيمة المجموع الكلي للاجزاء}} * 360^\circ = \frac{300}{1000} * 360^\circ = 108^\circ$

زاوية القطاع لدرجة **مدرس** = $\frac{\text{قيمة الجزء}}{\text{قيمة المجموع الكلي للاجزاء}} * 360^\circ = \frac{600}{1000} * 360^\circ = 216^\circ$

ثانيا) نرسم الدائرة بنصف قطر معين ونجري عليه عملية التقسيم للدائرة وذلك برسم زوايا متجاورة لكل منها موافق لقيمتها.



٤) العرض بالخطوط البيانية (the graphs lines)

ان هذه الطريقة لا تختلف في جوهرها عن **طريقة العرض باستخدام المدرج التكراري**، اذ ان كلاهما يوضح **العلاقة بين ظاهرتين او متغيرين حيث تعرض الظاهرة الأولى على المحور الافقي وقيم الظاهرة الثانية على المحور العمودي** وذلك باستخدام مقياس رسم مناسب. يكمن الاختلاف بينهما بان الباحث في هذه الطريقة يقوم بتوصيل **كل نقطتين متجاورتين بخط مستقيم بدلا من استخدام الاعمدة او المستطيلات**.

مثال/ ارسم المثال السابق على شكل خطوط بيانية: -

المحاضرة الرابعة

أولا) مقاييس النزعة المركزية (Measures of Central Tendency)

بعد عملية جمع البيانات وتبويبها وعرضها في جداول بيانية فيجب بعدها دراسة خصائص البيانات واستخلاص النتائج باستخدام مجموعة من المقاييس ومنها النزعة المركزية وهي تعني ميل المفردات او المشاهدات نحو التمركز او التجمع حول قيمة رقمية معينة في التوزيع التكراري وبالتالي فان هذه القيمة التي تتمركز حولها البيانات تكون ممثلة لباقي القيم ووسيلة لوصف البيانات ولإظهار الخصائص المهمة للظاهرة المرصودة من قبل الباحث.

هناك عدة مقاييس خاصة بقياس النزعة المركزية للبيانات حول الاحداث او الظواهر وفيما يأتي اهم هذه المقاييس من حيث خصائصها وطريقة حسابها: -

١- الوسط الحسابي (The Mathematician Mean)

ويسمى أيضا بالمتوسط الحسابي وهو اكثر أنواع المقاييس استخداما ويعرف على انه متوسط القيم لمتغير ما وهي القيمة الناتجة من قسمة مجموع القيم على عددها ويرمز له بالرمز $\bar{X}$ ويقرا(**اكس بار**) وهناك طرق لحسابه حسب نوع البيانات وهي: -

أ- **البيانات غير المبوبة:** اذا كان لدينا (n) من القيم او المشاهدات $(X_1,X_2,X_3,\dots,X_n)$ فان الوسط الحسابي لها هو: -

$$\bar{X} = \frac{\sum_{i=1}^{n} Xi}{n} \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\ (1)$$

مثال / إذا كانت اعمار معلمين في مدرسة معينة كالاتي 35 ,30 ,23 ,22 ,20 ما هو المتوسط الحسابي لأعمار المعلمين في تلك المدرسة؟

الحل: -

$$\bar{X} = \frac{\sum_{i=1}^{n} Xi}{n} = \frac{\sum_{i=1}^{5} Xi}{5} = \frac{20+22+23+30+35}{5} = \frac{130}{5} = 26$$

أي ان متوسط اعمار المعلمين في تلك المدرسة هو 26 سنة.

ب- البيانات المبوبة: وهي على نوعين: -

- **بيانات مبوبة حسب القيم وتكراراتها** وتحسب من القانون التالي: -

$$\bar{X} = \frac{\sum_{i=1}^{n} fiXi}{\sum_{i=1}^{n} fi} = \frac{f1X1+f2X2+\dots+fn\,Xn}{f1+f2+\dots+fn} \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\ (2)$$

حيث ان: -

$\sum_{i=1}^{n} fiXi$ = **هو مجموع حاصل ضرب القيمة في تكرارها.**

$\sum_{i=1}^{n} fi$ = **مجموع التكرارات.**

مثال/ جد الوسط الحسابي للقيم في الجدول التالي

تكرارها(fi)	القيم(Xi)
2	10
5	15
3	20
1	30

<u>الحل: -</u>

نوجد حاصــل ضــرب القيمة في تكرارها ونضــعها في عمود يضــاف للجدول ونوجد مجموع التكرارات ومجموع القيمة في تكرارها ونضعها في نهاية الجدول وكما مبين كالاتي: -

fi * Xi	تكرارها(fi)	القيم(Xi)
2×10= 20	2	10
5×15= 75	5	15
3×20= 60	3	20
1×30= 30	1	30
185	11	المجموع

نعوض في القانون: -

$$\bar{X} = \frac{\sum_{i=1}^{n} fiXi}{\sum_{i=1}^{n} fi} = \frac{185}{11} = 16.8$$

- **بيانات مبوبة حسب الفئات وتكراراتها** لحساب الوسط الحسابي في هذه الحالة نتبع الاتي: -

١- **نحسب ونحدد مراكز الفئات.**

٢- **ضرب كل مركز فئة بمقدار تكرارها.**

٣- **تقسيم (حاصل ضرب مركز كل فئة في تكرارها) على (مجموع التكرارات).**

مثال/ جد الوسط الحسابي للقيم في الجدول الاتي:

الفئات	التكرار fi
31 – 40	1
41 – 50	2
51 – 60	5
61 – 70	15
71 – 80	25
81 – 90	20
91 – 100	12

الحل: -

- نوجد مجموع التكرارات.
- نوجد مراكز الفئات حسب القانون التالي مركز الفئة $= \frac{\text{الحد الاعلى+الحد الادنى}}{2}$ بالنسبة لهذا الجدول ونضعها في عمود يضاف للجدول.
- نوجد حاصل ضرب مراكز الفئات في تكراراتها ونحسب مجموعها ونضعها في عمود يضاف للجدول

الفئات	التكرار fi	مراكز الفئات	مركز الفئة × التكرار
31 – 40	1	35.5	35.5
41 – 50	2	45.5	91
51 – 60	5	55.5	277.5
61 – 70	15	65.5	982.5
71 – 80	25	75.5	1887.5
81 – 90	20	85.5	1710
91 – 100	12	95.5	1146
المجموع	$\sum fi = 80$		$\sum fiXi = 6130$

نحسب الوسط الحسابي من القانون الاتي: -

$$\overline{X} = \frac{\sum_{i=1}^{n} fiXi}{\sum_{i=1}^{n} fi} = \frac{6130}{80} = 76.62$$

خواص الوسط الحسابي: -

١- قيمة الوسط الحسابي تتأثر بشكل كبير بقيم المشاهدات المتطرف الكبيرة منها والصغيرة وبالتالي فان الوسط الحسابي قد لا يكون معبرا بشكل حقيقي عن متوسط قيم المشاهدات بسبب القيم المتطرفة.

٢- مجموع انحرافات القيم عن توسطها الحسابي يساوي صفرا.

أي ان: -

- **بالنسبة للبيانات غير مبوبة: -**

$$\sum(Xi - \overline{X}) = 0$$

- **بالنسبة للبيانات المبوبة: -**

$$\sum fi(Xi - \overline{X}) = 0$$

٣- **إذا اضفنا عدد ثابت (K)** الى كل قيمة من قيم المشاهدات فان الوسط الحسابي للقيم الجديدة يساوي **الوسط الحسابي للقيم الاصلية + العدد الثابت (K)**. اما إذا **طرحنا عدد ثابت (K)** من كل قيمة من قيم المشاهدات فان الوسط الحسابي للقيم الجديدة يساوي **الوسط الحسابي للقيم الاصلية - العدد الثابت (K).**

مثال/ اذا كان لدينا القيم التالية $X_i = 8, 3, 2, 12, 10$ **فان الوسط الحسابي**

$$\overline{X_1} = \frac{\sum_{i=1}^{n} Xi}{n} = \frac{\sum_{i=1}^{5} Xi}{5} = \frac{8+3+2+12+10}{5} = \frac{35}{5} = 7$$

فاذا اضفنا 3 كعدد ثابت للقيم الاصلية للمجتمع تصبح $X_i = 11, 6, 5, 15, 13$ **فان الوسط الحسابي**

$$\overline{X_2} = \frac{\sum_{i=1}^{n} Xi}{n} = \frac{\sum_{i=1}^{5} Xi}{5} = \frac{11+6+5+15+13}{5} = \frac{50}{5} = 10$$

ويمكن حساب الوسط الحسابي مباشرة من **القاعدة رقم ٣** بدون القيام بعملية الحساب وكما يلي: -

$$\overline{X_2} = \overline{X_1} + k = 7 + 3 = 10$$

٤- **إذا ضربنا عدد ثابت (*K*)** في كل قيمة من قيم المشاهدات فان الوسط الحسابي للقيم الجديدة يساوي **الوسط الحسابي للقيم الاصلية × العدد الثابت (*K*)**. اما إذا **قسمنا عدد ثابت (*K*)** من كل قيمة من قيم المشاهدات فان الوسط الحسابي للقيم الجديدة يساوي **الوسط الحسابي للقيم الاصلية ÷ العدد الثابت (*K*)**.

٥- الوسط الحسابي لمجموع قيم متغيرين = مجموع الوسطين الحسابيين للمتغيرين

$$\bar{X} = \bar{Y} + \bar{Z}$$

٦- إذا كان لكل قيمة من المشاهدات () وزن يتناسب مع أهميتها فاذا رمزنا لهذه الوزن بالرمز () فان الوسط الحسابي الموزون لهذه القيم هو: -

$$\bar{X} = \frac{\sum_{i=1}^{n} wiXi}{\sum_{i=1}^{n} wi} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textbf{(3)}$$

مثال/ القيم التالية تمثل النتائج النهائية لكل الدروس لاحد الطلبة لأربعة أعوام في كلية التربية الاساسية علما ان لكل نتيجة وزنا واهمية او نسبة معينة. اوجد الوسط الحسابي الموزون لنتيجة هذا الطالب للأعوام الأربعة.

الأهمية النسبية (*Wi*)	معدل النتيجة النهائية (*Xi*)	سنة النتيجة
10%	70	المرحلة الاولى
20%	60	المرحلة الثانية
30%	75	المرحلة الثالثة
40%	55	المرحلة الرابعة

الحل: - ١) نوجد حاصل **ضرب معدل النتيجة النهائية في أهميتها النسبية** لكل سنة من السنوات. ونضعها في عمود

٢) نوجد **مجموع حاصل ضرب معدل النتيجة النهائية في أهميتها** النسبية **ومجموع الأهمية النسبية**.

Wi × Xi	الأهمية النسبية (*Wi*) %	معدل النتيجة النهائية (*Xi*)	سنة النتيجة
700	10%	70	المرحلة الاولى
1200	20%	60	المرحلة الثانية
2250	30%	75	المرحلة الثالثة
2200	40%	55	المرحلة الرابعة
6350	100%		المجموع

اذن **الوسط الحسابي او معدل الطالب لأربع سنوات** هو: -

$$\bar{X} = \frac{\sum_{i=1}^{n} wiXi}{\sum_{i=1}^{n} wi} = \frac{6350}{100} = 63.5\%$$

المحاضرة الخامسة

ثانيا) الوسيط (The Median)

هو القيمة الوسطية التي تقع في منتصف مجموع من البيانات المرتبة تصاعديا او تنازليا إذا كان عدد المتغيرات فرديا. ومتوسط القيمتين الوسطيتين إذا كان عدد المتغيرات زوجيا ويرمز له بالرمز **($\overline{\text{Me}}$) ويلفظ (ام أي بار).**

كيفية حساب الوسيط

- **حساب الوسيط إذا كانت القيم غير مبوبة: - وهي على نوعين**

١- إذا كان عدد قيم المشاهدات **فردي** فيمكن اتباع الخطوات التالية لاستخراج الوسيط.

أ- نرتب قيم المشاهدات **تصاعديا** او **تنازليا**.

ب- نستخرج **ترتيب الوسيط** وذلك من خلال المعادلة التالية: -

$$\text{ترتيب الوسيط } (\overline{\text{Me}}) = \frac{n+1}{2}$$

حيث (n) هي **عدد قيم المشاهدات.**

مثال/ جد الوسيط للقيم التالية 16, 12, 8, 11, 9, 10, 17

الحل: -

نرتب القيم تصاعديا: -

8, 9, 10, 11, 12, 16, 17

$$\text{ترتيب الوسيط } (\overline{\text{Me}}) = \frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4$$

اذن **الوسيط** هو القيمة الرابعة في الترتيب التصاعدي = **11**

او نرتب القيم تنازليا: -

17, 16, 12, 11, 10, 9, 8

$$\text{ترتيب الوسيط } (\overline{\text{Me}}) = \frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4$$

اذن **الوسيط** هو القيمة الرابعة في الترتيب التنازلي = **11**

٢- إذا كان عدد قيم المشاهدات **زوجي** فيمكن اتباع الخطوات التالية لاستخراج الوسيط.

أ- نرتب قيم المشاهدات **تصاعديا او تنازليا**.

ب- نحدد **ترتيب الوسيطين**:

$$\text{ترتيب الوسيط الأول} = \frac{n}{2} \quad , \quad \text{ترتيب الوسيط الثاني} = \frac{n}{2} + 1$$

٣- نحدد **القيم المناظرة** لكل من **الوسيط الأول والوسيط الثاني** في الترتيب **التصاعدي** او **التنازلي.**

٤- نستخرج **الوسيط** من القانون التالي: -

$$\textbf{الوسيط } (\overline{\mathbf{Me}}) = \frac{\text{قيمة الوسيط الثاني} + \text{قيمة الوسيط الاول}}{2}$$

مثال/ احسب الوسيط للقيم الاتية: -

18, 9, 16, 12, 14, 4, 6, 8

الحل

نرتب القيم تصاعديا

4, 6, 8, 9, 12, 14, 16, 18

نوجد ترتيب الوسيط الأول والثاني

ترتيب الوسيط الأول $= \frac{n}{2} = \frac{8}{2} = 4$, **قيمة الوسيط الأول** = 9

ترتيب الوسيط الثاني $= (\frac{n}{2} + 1) = (4+1) = 5$, **قيمة الوسيط الثاني** = 12

اذن ...

$$\textbf{الوسيط } (\overline{\mathbf{Me}}) = \frac{\text{قيمة الوسيط الثاني} + \text{قيمة الوسيط الاول}}{2} = \frac{9+12}{2} = \frac{21}{2} = \mathbf{10.5}$$

وبنفس الطريقة يمكن حساب الوسيط بترتيب القيم تنازليا.

- **حساب الوسيط إذا كانت البيانات المبوبة: -**

يمكن تلخيص خطوات إيجاد الوسيط لبيانات مبوبة كالاتي: -

١- عمل جدول **توزيع تكراري تجميعي تصاعدي**

٢- إيجاد **ترتيب الوسيط** $= \frac{\sum fi}{2}$, حيث $\sum fi$: -هي مجموع التكرارات.

٣- نحدد **فئة الوسيط** وهي القيمة التي تقع قيمة الوسيط بين حديها وذلك عن طريق **إيجاد قيمتين متتاليتين في التكرار التجميعي التصاعدي يقع بينها قيمة الوسيط** .... ويتضمن: -

أ- **إيجاد حدوده الحقيقية**

ب- **كتابة التكرار التجميعي التصاعدي** امام كل منها.

٤- تطبيق القانون: -

$$(\overline{\mathbf{Me}}) = \mathbf{L_i} + \left[\frac{\frac{\sum fi}{2} - \mathbf{F_i}}{fi}\right] \times w$$

حيث: - $\mathbf{L_i}$ = **الحد الأدنى الحقيقي لفئة الوسيط.**

$\sum fi$ = **مجموع التكرارات.**

$\mathbf{F_i}$ = **التكرار المتجمع عند بداية فئة الوسيط.**

---

f_i = تكرار فئة الوسيط.

w = طول الفئة

مثال/ اوجد الوسيط للتوزيع التكراري التالي: -

فئات الطول (سم)	التكرار (fi)
60 – 62	5
63 – 65	18
66 – 68	42
69 – 71	27
72 – 74	8

الحل: -

١- **نوجد التكرار التجميعي التصاعدي ومجموع التكرارات والحدود الحقيقية للفئات: -**

فئات الطول (سم)	التكرار (fi)	التكرار المتجمع الصاعد		الحدود الحقيقية للفئات
			Fi	
60 – 62	5	اقل من 60	0	59.5 – 62.5
63 – 65	18	اقل من 63	5	62.5 – 65.5
66 – 68	42	اقل من 66	**23**	65.5 – 68.5
69 – 71	27	اقل من 69	**65**	68.5 – 71.5
72 – 74	8	اقل من 72	92	71.5 – 74.5
		اقل من 74	100	
	$\sum fi = 100$			

٢- **إيجاد ترتيب الوسيط: -**

$$\text{ترتيب الوسيط} = \frac{\sum fi}{2} = \frac{100}{2} = 50$$

أي ان **قيمة الوسيط** هو طول الشخص الذي ترتيبه **(50)** (**بعد ترتيب القيم تصاعديا او تنازليا**).

وفي **جدول التوزيع التكراري التجميعي التصاعدي** نرى ان **(50)** هي واقعة بين الرقمين **(23)** و **(65)**.

اذن **الحد الأدنى الحقيقي لفئة الوسيط** $L_i = 65.5$

التكرار المتجمع عند بداية فئة الوسيط $F_i = 23$

تكرار فئة الوسيط $f_i = 65 - 23 = 42$

طول فئة الوسيط $w = 68.5 - 65.5 = 3$

نعوض في القانون: -

$$(\overline{Me}) = L_i + \left[\frac{\frac{\sum fi}{2} - Fi}{fi}\right] \times w = 65.5 + \left[\frac{50 - 23}{42}\right] \times 3 = 67.43 \text{ cm}$$

المحاضرة السادسة

ثالثا) المنوال او القمة (The Mode)

تعريف: -

١- بيانات غير مبوبة: -

اذا كان لدينا n من المشاهدات $(y_1, y_2, y_3, \ldots y_n)$ فان المنوال لهذه المشاهدات هو القيمة **الأكثر تكرارا** بين هذه المشاهدات ويرمز له بالرمز $(\overline{Mo})$ ويقرا (**ام او بار**).

ومن هذا يتضح انه قد يكون هناك **منوال واحد** لهذه المشاهدات وعندها يسمى التوزيع وحيد القمة **(Unimodal)** او يكون هناك **منوالان** وعندها يسمى التوزيع ذو قمتين **(Bimodal)** وقد يكون هناك **أكثر من منوالين** كما انه قد **لا يوجد منوال** للمشاهدات.

مثال/ اوجد **المنوال** للبيانات التالية:

a- **3, 5, 2, 6, 5, 9, 5, 2, 8, 6**

b- **51.6, 78.7, 50.3, 49.5, 48.9**

الحل

a) المفردة (5) هي أكثر المفردات تكرارا فهي المنوال.

$$\overline{Mo} = 5$$

b) لا توجد مفردات متكررة ...اذن لا يوجد منوال لهذه المفردات.

٢- بيانات مبوبة

إذا كانت القيم $(y_1, y_2, y_3, \ldots. y_n)$ تمثل **مراكز الفئات** في جدول التوزيع التكراري

مع تكراراتها $(f_1, f_2, f_3, \ldots. f_n)$ على التوالي.

فان **المنوال** هو: -

$$\overline{Mo} = L_i + \left(\frac{d1}{d1+d2}\right)\times w$$

حيث ان: -

فئة المنوال هي الفئة التي تملك أكبر عدد من التكرار ات.

L_i = الحد الأدنى الحقيقي لفئة المنوال.

d_1 = الفرق بين تكرار فئة المنوال والفئة السابقة لها.

d_2 = الفرق بين تكرار فئة المنوال والفئة اللاحقة لها.

w = طول الفئة.

مثال / اوجد المنوال لجدول التوزيع التكراري التالي: -

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

الحل

أولا) إيجاد فئة التكرارات: -

ان الفئة **(66 – 68)** لها **أكبر التكرارات (42)** فهي **فئة المنوال**

ثانيا) إيجاد الحدود الحقيقية للفئات: -

الحدود الحقيقية	التكرار f_i	الفئات
59.5 – 62.5	5	60 – 62
62.5 – 65.5	18	63 – 65
65.5 – 68.5	42	66 – 68
68.5 – 71.5	27	69 – 71
71.5 – 74.5	8	72 – 74

ثالثا) تطبيق القانون: -

$$\overline{Mo} = L_i + \left(\frac{d1}{d1+d2}\right) \times w$$

$L_i = 65.5$

$d_1 = 42 - 18 = 24$

$d_2 = 42 - 27 = 15$

$w = 68 - 66 + 1 = 3$

$$\overline{Mo} = 65.5 + \left(\frac{24}{24+15}\right) \times 3$$

$$= 65.5 + 1.85$$

$$= 67.35$$

المحاضرة السابعة

ثانيا) مقاييس التشتت او الاختلاف (Deviation Measurement Or Variance)

يقصد **بالتشتت او الاختلاف بانه التباعد او التقارب بين قيم المشاهدات التابعة لمتغير ما**، ويمكن تعريف مقاييس التشتت بانها **مقياس تشتت قيم المشاهدات عن وسطها الحسابي**. حيث كلما كان مقياس التشتت **قيمته كبيره** فهذا يدل على **عدم التجانس بين القيم للمشاهدات**. وتكون قيمة مقياس التشتت **صغيرة** عندما تكون **الفروق بين قيم المشاهدات قليلة**.

اذن فان **مقياس التشتت** يعطينا فكرة عن **مدى تجانس او تباين قيم المشاهدات لظاهرة معينة حول وسطها الحسابي** او بمعنى اخر مقدار تشتتها او انتشارها او بعدها عن الوسط الحسابي لها.

ان **لمقاييس التشتت أهمية في وصف التوزيعات التكرارية ومقارنتها مع بعضها البعض** حيث ان مقاييس التوسط التي درسناها سابقا لا تكفي لهذا الغرض، **فقد يتساوى الوسط الحسابي لمجموعتين من القيم مثلا بينما يختلف مدى تجانس او تشتت قيم المجموعة الأولى عن تجانس وتشتت القيم للمجموعة الثانية** وكما هو موضح في المثال التالي: -

المجموعة الأولى $X_1 = 22, 21, 19, 18, 23, 20, 17$

المجموعة الثانية $X_2 = 13, 20, 45, 5, 7, 15, 35$

فعند حساب **الوسط الحسابي** لكلا المجموعتين نجد انه يساوي **20** ولكن من خلال المشاهدة نجد ان **قيم المجموعة الأولى أكثر تجانسا من قيم المجموعة الثانية** على الرغم من تساوي وسطيهما الحسابي.

أنواع مقاييس التشتت

هناك عدة أنواع من مقاييس التشتت أهمها: -

أولا- مقاييس التشتت المطلق: - حيث تكون وحدات مقاييس التشتت نفس وحدات القيم الاصلية فمثلا إذا كان القياس بالسنتمتر للقيم الاصلية فان وحدة مقاييس التشتت تكون بالسنتمتر ايضا وهكذا ... ومن أهمها: -

١- المدى (The Range): -

يعرف **المدى لمجموعة معينة من القيم** بانه **الفرق بين اعلى قيمة واقل قيمة** في تلك المجموعة ويرمز له بالرمز **(R)**.

$$R = X_{max.} - X_{min.} \qquad \text{--------} \ (1)$$

حيث ان: -

X_{max} = اعلى قيمة في البيانات لمجموعة معينة.

X_{min} = ادنى قيمة في البيانات لمجموعة معينة.

مثال/ احسب المدى للقيم في المجموعات التالية: -

$$X_i = 12, 6, 7, 3, 15, 10, 18, 5$$

$$Y_i = 9, 3, 8, 9, 8, 9, 8, 18$$

Sol.

- $R_{xi} = 18 - 3 = 15$
- $R_{Yi} = 18 - 3 = 15$

نلاحظ ان **المدى في كلا المجموعتين متساوي** ولكن ... **عدم التجانس في المجموعة الأولى أكبر من عدم التجانس في المجموعة الثانية** التي تتألف معظمها من العددين 8, 9 لذلك فان المدى يكون أحيانا مضللا لان يعتمد فقط على القيمتين الطرفيتين للمجموعة التي من الممكن ان تكون متساوية وكذلك من الصعب حساب المدى الحقيقي للتكرارات في جدول توزيع تكراري بسبب عدم امكانية معرفة القيمتين الطرفيتين.

٢- الانحراف المتوسط (The Mean Deviation)

أولا - بيانات غير مبوبة: -

إذا كان لدينا (n) من المشاهدات $(X_1, X_2, X_3, \ldots, X_n)$ فان الانحراف المتوسط لها هو متوسط الانحرافات المطلقة (أي اهمال الإشارة) عن وسطها الحسابي ويرمز له بالرمز (M.D)

أي ان: -

$$M.D = \frac{\sum|xi - \bar{x}|}{n} \text{ ------------ } (2)$$

ملاحظة: - السبب في اخذ **الانحرافات المطلقة** هو ان **إبقاء الإشارات الموجبة والسالبة** يجعل **مجموع انحرافات القيم عن الوسط الحسابي مساويا للصفر** كما ذكرنا سابقا في خواص الوسط الحسابي.

$$\sum xi - \bar{x} = 0$$

مثال/ اوجد الانحراف المتوسط للقيم التالية: -

$$Xi = 9, 8, 6, 5, 7$$

الحل: -

أولا) نحسب الوسط الحسابي للقيم

$$\bar{x} = \frac{9+8+6+5+7}{5} = 7$$

ثانيا) نضع القيم والوسط في جدول ونضيف لها العمودين التاليين: -

Xi	$\bar{x}$	$Xi - \bar{x}$	$\lvert Xi - \bar{x}\rvert$
9	7	2	2
8	7	1	1
6	7	-1	1
5	7	-2	2
7	7	0	0
n = 5		$\sum xi - \bar{x} = 0$	$\sum \lvert Xi - \bar{x}\rvert = 6$

ثالثا) نعوض في القانون: -

$$M.D = \frac{\sum |xi - \bar{x}|}{n}$$

$$= \frac{6}{5}$$

$$= 1.2$$

ثانيا) بيانات مبوبة: -

إذا كانت لدينا $(X_1, X_2, X_3, \ldots, X_n)$ تمثل مراكز الفئات في جدول توزيع تكراري مع تكراراتها $(f_1, f_2, f_3, \ldots. f_n)$ فان الانحراف المتوسط لجدول توزيع تكراري هو: -

$$M.D = \frac{\sum fi|xi - \bar{x}|}{\sum fi} \text{ ------------ (3)}$$

مثال/ اوجد الانحراف المتوسط **لجدول التوزيع التكراري التالي: -**

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

الحل: - نعمل الجدول التالي: -

الفئات	التكرار f_i	مركز الفئة xi	$fixi$	$\overline{x}$	$\|xi - \overline{x}\|$	$fi\|xi - \overline{x}\|$
60 – 62	5	61	305	67.45	6.45	32.25
63 – 65	18	64	1152	67.45	3.45	62.10
66 – 68	42	67	2814	67.45	0.45	18.90
69 – 71	27	70	1890	67.45	2.55	68.85
72 – 74	8	73	584	67.45	5.55	44.40
المجموع	100		6745			226.50

خطوات الحل: -

١- نحسب مراكز الفئات ونضعها في عمود يضاف للجدول.

٢- نضرب التكرارات في مراكز الفئات ونوجد المجموع لها ونضعها في عمود يضاف للجدول.

٣- نحسب الوسط الحسابي حسب القانون التالي ونضعه في عمود يضاف للجدول.

$$\overline{x} = \frac{\sum fixi}{\sum fi} = \frac{6745}{100} = 67.45$$

٤- نحسب الانحراف المطلق $|xi - \overline{x}|$ ونضعه في عمود يضاف للجدول.

٥- نضرب التكرارات في الانحراف المطلق $fi|xi - \overline{x}|$ ونوجد المجموع لها ونضعها في عمود يضاف للجدول.

٦- نعوض في القانون: -

$$M.D = \frac{\sum fi|xi - \overline{x}|}{\sum fi}$$

$$= \frac{226.50}{100} = 2.265$$

المحاضرة الثامنة

ثالثا) التباين (variance)

ان مجموع انحرافات أي عينة عن وسـطها الحسـابي يسـاوي صـفرا. وبدلا من اخذ القيمة المطلقة للانحرافات (أي بدون اشارات) فإننا نستطيع ان نتغلب على ذلك بتربيع قيم الانحرافات وبذلك تصبح جميعها موجبة أي نحصل على مجموع الانحرافات (summation of squares) والتي يرمز لها بالرمز ($S\ S$) وعلى ذلك فان: -

$$S\ S = \sum (Xi - \bar{X})^2 \quad \text{----------------- (1)}$$

ولكي نأخذ حجم العينة في الاعتبار حتى نتمكن من مقارنة العينات المختلفة فإننا **نقسـم مجموع الانحرافات على درجات الحرية** (degree of freedom) والتي تمثل بـ (n – 1) وبذلك نحصل على ما يسمى التباين (S^2).

مما سبق يمكن تعريف التباين كما يلي: -

1- بيانات غير مبوبة: -

إذا كانت لدينا (n) من المشاهدات **($X_1, X_2, X_3, \dots, X_n$) فان التباين والذي يرمز له بالرمز (S^2) يكون كالاتي:**

$$S^2 = \frac{\sum_{i=1}^{n}(xi - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{i=1}^{n} Xi)^2}{n}}{n-1} \quad \text{----------------(2)}$$

ويلاحظ ان هذا القانون هو لحساب تباين العينة اما إذا كانت قيم المشاهدات تمثل المجتمع كله فان التباين للمجتمع ويرمز له (σ^2) ويلفظ (سيكما سكوير) ويحسب كالاتي: -

$$\sigma^2 = \frac{\sum_{i=1}^{n}(xi - \bar{M})^2}{N} \quad \text{---------------(3)}$$

حيث ان: -

M = الوسط الحسابي للمجتمع

N = عدد المفردات للمجتمع.

ومن الملاحظ في حالة إيجاد تباين العينة نقسم على (n – 1) أي على درجة الحرية وهو ما يعني ان مجموع الانحرافات عن الوسط الحسابي يساوي صفرا. لذلك فعند سحب عينة واحدة فان (n – 1) من المشاهدات هي قيم حرة.

رابعا) الانحراف القياسي (standard deviation)

ويسمى أيضا بالخطأ القياسي او الانحراف المعياري وهو اخذ الجذر التربيعي للتباين وذلك لإرجاع قيمة التباين الى وحداته الاصلية أي ان: -

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n}(xi-\bar{X})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{i=1}^{n} Xi)^2}{n}}{n-1}} \qquad \text{--------------------- (4)}$$

اما الانحراف المعياري للمجتمع (σ) فهو: -

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{n}(xi-\bar{M})^2}{N}} \qquad \text{-----------------(5)}$$

مثال/ البيانات الاتية تبين عدد الأساتذة في خمسة اقسام. احسب التباين والانحراف القياسي لهذه القيم.

Xi = 9, 8, 6, 5, 7

الحل: -

يمكن حل هذا السؤال بطريقتين وكما يلي: -

أولا) الطريقة الأولى: -

١- نحسب الوسط الحسابي للقيم. حسب القانون: -

$$\bar{X} = \frac{\sum_{i=1}^{n} Xi}{n} = \frac{9+8+6+5+7}{5} = 7$$

٢- نطرح كل قيمة من قيم المشاهدات من وسطها الحسابي.

٣- نربع القيم المطروحة من الوسط. وكما مبين في الجدول التالي.

Xi	$\bar{X}$	$(Xi - \bar{X})$	$(Xi - \bar{X})^2$
9	7	2	4
8	7	1	1
6	7	-1	1
5	7	-2	4
7	7	0	0
$\sum_{i=1}^{5} Xi = 35$			$\sum_{i=1}^{5}(Xi - \bar{X})^2 = 10$

٤- نطبق القانون: -

$$S^2 = \frac{\sum_{i=1}^{n}(xi-\bar{X})^2}{n-1} = \frac{10}{5-1} = \frac{10}{4} = 2.5$$ (التباين)

$$S = \sqrt{S^2} = \sqrt{2.5} = 1.58$$ (الانحراف القياسي)

ثانيا) الطريقة الثانية

١- نربع قيم المشاهدات مباشرة ونحسب مجموعهما. وكما مبين في الجدول التالي: -

Xi	Xi^2
9	81
8	64
6	36
5	25
7	49
$\sum_{i=1}^{5} Xi = 35$	$\sum_{i=1}^{5} Xi^2 = 255$

٢- نطبق القانون الاتي: -

$$S^2 = \frac{\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{i=1}^{n} Xi)^2}{n}}{n-1} = \frac{255 - \frac{(35)^2}{5}}{4} = 2.5$$ (التباين)

$$S = \sqrt{S^2} = \sqrt{2.5} = 1.58$$ (الانحراف القياسي)

المحاضرة التاسعة

ب- بيانات مبوبة: -

إذا كانت لدينا $(X_1, X_2, X_3, \ldots, X_n)$ تمثل مراكز الفئات في جدول توزيع تكراري مع تكراراتها $(f_1, f_2, f_3, \ldots.\ f_n)$ فان الانحراف القياسي هو: -

$$S = \sqrt{\frac{\sum fi(Yi-\bar{Y})^2}{\sum fi-1}} = \sqrt{\frac{\sum fiyi^2-\frac{(\sum fiyi)^2}{\sum fi}}{\sum fi-1}} \quad \text{------------ 1}$$

(الانحراف القياسي)

$$S^2 = \frac{\sum fi(Yi-\bar{Y})^2}{\sum fi-1} = \frac{\sum fiyi^2-\frac{(\sum fiyi)^2}{\sum fi}}{\sum fi-1} \quad \text{------------- 2}$$

(التباين)

مثال / احسب الانحراف القياسي والتباين للجدول الاتي: -

التكرار fi	الفئات
5	60 – 62
18	63 – 65
42	66 – 68
27	69 – 71
8	72 – 74
$\sum fi = 100$	

الحل: -

ا- الطريقة المطولة: - نعمل الجدول الاتي

$f_i(y_i-\bar{y})^2$	$(y_i-\bar{y})^2$	$y_i-\bar{y}$	y_i	التكرار fi	الفئات
208.0125	41.6025	- 6.45	61	5	60 – 62
214.2450	11.9025	- 3.45	64	18	63 – 65
8.5050	0.2025	- 0.45	67	42	66 – 68
175.5675	6.5025	2.55	70	27	69 – 71
246.4200	30.8025	5.55	73	8	72 – 74

852.7500				$\sum fi = 100$	

اذن مجموع المربعات (SS) هو: -

$$SS = \sum fi(yi - \bar{y})^2 = 852.7500$$

وان التباين هو: -

$$s^2 = \frac{SS}{\sum fi - 1} = \frac{852.7500}{99} = 8.6$$

اما الانحراف القياسي فهو: -

$$S = \sqrt{s^2} = \sqrt{8.6} = 2.9$$

٢- الطريقة المختصرة: -

الفئات	التكرار f_i	y_i	$f_i y_i$	y_i^2	$f_i y_i^2$
60 – 62	5	61	305	3721	18605
63 – 65	18	64	1152	4096	73728
66 – 68	42	67	2814	4489	188538
69 – 71	27	70	1890	4900	132300
72 – 74	8	73	584	5329	42632
	$\sum fi = 100$		6745		455803

اذن مجموع المربعات (SS) هو: -

$$SS = \sum fiyi^2 - \frac{(\sum fiyi)^2}{\sum fi} = 455803 - \frac{(6745)^2}{100} = 852.75$$

وان التباين هو: -

$$s^2 = \frac{SS}{\sum fi - 1} = \frac{852.75}{99} = 8.6$$

اما الانحراف القياسي فهو: -

$$S = \sqrt{s^2} = \sqrt{8.6} = 2.9$$

اهم خواص التباين والوسط الحسابي: -

١- عند إضــافة (او طرح) عدد ثابت (K) الى كل قيمة من قيم المشــاهدات، فان قيمة التباين والانحراف القياسي لا يتغيران أي ان: -

التباين للقيم الجديدة = التباين للقيم الاصلية.

الانحراف القياسي للقيم الجديدة = الانحراف القياسي للقيم الاصلية.

أي ان

$$x_i = (yi + k)\ , \qquad x_i = (y_i - k)$$

$$S^2_x = S^2_y$$

$$\therefore \quad S_x = S_y$$

مثال/ احسب التباين والانحراف القياسي للقيم التالية ثم أ ضف (3) لكل منها واحسب التباين والانحراف القياسي للقيم الجديدة.

<u>الحل</u>

(القيم الاصلية)　　$y_i = 8, 3, 2, 12, 10$

(القيم الجديدة بعد إضافة (٣) لكل قيمة)　　$x_i = 11, 6, 5, 15, 13$

y_i	y^2_i		x_i	x^2_i
8	64		11	121
3	9		6	36
2	4		5	25
12	144		15	225
10	100		13	169
35	321		50	576

$$SS_y = \sum yi^2 - \frac{(\sum yi)^2}{n}$$

$$= 321 - \frac{(35)^2}{5}$$

$$= 76$$

$$SS_x = \sum xi^2 - \frac{(\sum xi)^2}{n}$$

$$= 576 - \frac{(50)^2}{5}$$

$$= 76$$

$$S^2_y = \frac{SSy}{n-1} = \frac{76}{4} = 19$$

$$S^2_x = \frac{SSx}{n-1} = \frac{76}{4} = 19$$

$S_x = \sqrt{S^2x} = \sqrt{19}$ $S_y = \sqrt{S^2y} = \sqrt{19}$

نلاحظ ان التباين او الانحراف القياسي لم يتأثر بإضافة العدد الثابت الى كل قيمة من القيم الاصلية.

٢- إذا ضربت كل قيمة من قيم المشاهدات بعدد ثابت (K) فان: -

التباين للقيم الجديدة = التباين للقيم الاصلية × مربع العدد الثابت.

الانحراف القياسي للقيم الجديدة = الانحراف القياسي للقيم الاصلية × العدد الثابت.

٣- إذا كان كل من (x) و (y) متغيرين مستقلين وكان المتغير (z) يساوي مجموعهما فان: -

تباين (z) = تباين (x) + تباين (y)

* العلاقة بين الانحراف القياسي والانحراف المتوسط: -

إذا كان التوزيع غير متماثل (ملتوي التواء بسيطا) فان: -

الانحراف المتوسط = $\frac{4}{5}$ الانحراف القياسي

$$M.D = \frac{4}{5} S$$

ملاحظة: -

عند قياس مدى تشتت متوسطات العينات التابعة لمجتمع ما فانه يستخدم ما يسمى بالخطأ القياسي (Standard Error) او الانحراف القياسي للمتوسطات (Standard Deviation Of The Mean) ويرمز له بالرمز ($S_{\bar{y}}$) ويحسب من القانون التالي: -

$$S_{\bar{y}} = \sqrt{\frac{S^2}{n}} = \frac{S}{\sqrt{n}}$$

ثانيا) مقاييس التشتت النسبي: -

ان مقاييس التشتت النسبي لها أهميتها عند مقارنة تشتت مجموعتين او أكثر تختلف في وحدات القياس لقيمها، لان مقاييس التشتت النسبي تكون خالية من وحدات القياس ومن اهم أنواعها: -

١- معامل الاختلاف (coefficient of variance)

إذا كان (S) و ($\bar{y}$) هما الانحراف القياسي و الوســط الحســابي لمجموعة من القيم على التوالي فان معامل الاختلاف ويرمز له (C. V.) هو: -

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

مثال/ نتائج الامتحانات النهائية لدرسي الإحصاء والكيمياء للصف الأول كانا كالاتي: -

	الإحصاء	الكيمياء
الوسط الحسابي	78	73
الانحراف القياسي	8	7.6

ففي أي الموضوعين كان تشتت الدرجات أكثر ؟

الحل

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

$$= \frac{8}{78} \times 100 = 10.25\% \qquad \text{(للإحصاء)}$$

$$= \frac{7.6}{73} \times 100 = 10.41\% \qquad \text{(للكيمياء)}$$

ملاحظة: - يمكن استخدام مقاييس أخرى للتشتت النسبي لأنه في الحقيقة: -

معامل الاختلاف = $\dfrac{\text{مقياس من المقاييس المطلقة}}{\text{مقياس من مقاييس التوسط}}$ × 100 %

مثلا في حالة استخدام الانحراف المتوسط يمكن حساب معامل الاختلاف كالاتي: -

$$C.V. = \frac{M.D}{\bar{y}} \times 100$$

$$C.V. = \frac{M.D}{Me} \times 100$$

٢- الدرجة القياسية (Standardized Scores)

في كثير من الأحيان نحتاج الى مقارنة مفردتين من مجموعتين مختلفتين. وفي هذه الحالة يجب تحويل وحدات كل مفردة الى وحدات قياســية حتى تكون المقارنة ذات معنى وذلك باســتخدام الوســط الحســابي والانحراف القياسي لكل مجموعة.

تسمى القيمة (Z_i) درجة قياسية إذا كانت تساوي: -

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

مثال/ حصـل طالب على درجة 84 في الامتحان النهائي بالرياضـيات علما ان الوسـط الحسـابي في امتحان الرياضـيات لجميع الطلبة كان = 76 وبانحراف قياسي = 10 , اما في امتحان الفيزياء فقد حصـل نفس الطالب على درجة ٩٠ حيث كان الوسط الحسابي في امتحان الفيزياء لجميع الطلبة = 82 والانحراف القياسي = 16 ففي أي الموضوعين كانت قابلية الطالب اعلى؟

الحل: -

عند مقارنة درجة الامتحانين مباشرة نجد ان درجته في الفيزياء (90) اعلى من درجته في الرياضـيات (84).

ولكن عند تحويل هاتين الدرجتين الى درجات قياسية نجد ان: -

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

$$Z_i = \frac{84 - 76}{10} = 0.8 \quad \text{(بالنسبة للرياضيات)}$$

$$Z_i = \frac{90 - 82}{16} = 0.5 \quad \text{(بالنسبة للفيزياء)}$$

ومن هذا يتضح ان قابليته في الرياضيات اعلى مما في الفيزياء وهو عكس ما توصلت اليه المقارنة السابقة.

المحاضرة العاشرة

مقاييس الارتباط (Measures of Correlation)

سبق ان درسنا بعض المقاييس الإحصائية مثل مقاييس النزعة المركزية ومقاييس التشتت والتي كانت تستخدم في وصف توزيع واحد بصورة منفصلة في علاقته بالتوزيعات الأخرى. بينما مقاييس الارتباط تتطلب طرق إحصائية لدراسة العلاقة بين متغيرين او توزيعين بدلا من دراسة خصائص توزيع واحد ومن هذه الدراسات: -

١- علاقة مستوى الذكاء بالميول المهنية.
٢- علاقة تحصيل الطالب بذكائه.
٣- علاقة المستوى الاقتصادي بالتطور الاجتماعي لمجتمع معين.

وهذه الدراسات تأخذ الابعاد الآتية: -

- إذا كانت قيمة المتغير الأول عالية وقيمة المتغير الثاني عالية تكون العلاقة موجبة او طردية.
- إذا كانت قيمة المتغير الأول واطئة وقيمة المتغير الثاني واطئة تكون العلاقة ايضا موجبة او طردية.
- إذا كانت قيمة المتغير الأول عالية وقيمة المتغير الثاني واطئة او العكس تكون العلاقة سالبة او عكسية.
- إذا كانت قيم المتغيرين غير واضحة الاتجاه لا توجد علاقة بين المتغيرين.

وهناك طريقة لتفسير تلك العلاقات بين أي متغيرين تقاس بمعامل هو معامل الارتباط ويرمز له بالرمز (r) ويتخذ قيم عددية محصورة ضمن المدى (١,-١) او كالاتي: -

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

فاذا كانت قيمة (r) أكبر او أصغر من هذه الحدود فهذا يدل على وجود خطأ حسابي.

ملاحظات: -

١- إذا كانت قيمة $(r = 1)$ فان العلاقة موجبة تامة او طردية تامة.
٢- إذا كانت قيمة $(r = -1)$ فان العلاقة سالبة تامة او عكسية تامة.
٣- إذا كانت قيمة $(r = 0)$ فانه لا توجد علاقة بين المتغيرين.
٤- إذا كانت قيمة $(-1 \leq r \leq 0)$ فان العلاقة سالبه او عكسية.

٥- إذا كانت قيمة $(-1 \le r \le 0)$ فان العلاقة موجبه او طردية تزداد قوتها كلما اقتربنا من واحد صحيح.

أنواع معاملات الارتباط: -

١- معامل ارتباط بيرسون (معامل الارتباط الخطي البسيط): -

يستخدم إذا كان (x, y) متصلين او مستمرين على شكل ارقام او قيم عددية والعلاقة بينهما علاقة خطية.

ويحسب من العلاقة التالية: -

$$r = \frac{n\sum_{i=1}^{n} XiYi - \sum_{i=1}^{n} Xi * \sum_{i=1}^{n} Yi}{\sqrt{[n\sum (Xi)^2 - (\sum Xi)^2] * [n\sum (Yi)^2 - (\sum Yi)^2]}}$$

مثال/ احسب معامل ارتباط بيرسون للبيانات التالية: -

$Xi = 1, 2, 3, 4, 5$　　　$Yi = 3, 6, 9, 12, 15$

الحل: -

Xi	Yi	Xi^2	Yi^2	$Xi * Yi$
1	3	1	9	3
2	6	4	36	12
3	9	9	91	27
4	12	16	144	48
5	15	25	225	75
15	45	55	495	165

نطبق القانون: -

$$r = \frac{n\sum_{i=1}^{n} XiYi - \sum_{i=1}^{n} Xi * \sum_{i=1}^{n} Yi}{\sqrt{[n\sum (Xi)^2 - (\sum Xi)^2] * [n\sum (Yi)^2 - (\sum Yi)^2]}}$$

$$r = \frac{5*165 - 15*45}{\sqrt{[5*55 - (15)^2][5*495 - (465)^2]}}$$

$$r = \frac{825-675}{\sqrt{[270-225][2475-2025]}}$$

$$r = \frac{150}{\sqrt{50*450}}$$

$$r = \frac{150}{\sqrt{22500}} = \frac{150}{150} = 1$$

(أي ان الارتباط إيجابي تام او طردي تام)

٢- معامل ارتباط سبيرمان (معامل ارتباط الرتب): -

اشـتق هذا القانون لمعالجة حالات خاصـة تعتمد على رتب القيم بدلا من اسـتخدام القيم العددية الاصـلية التي تجري معالجتها باستخدام معامل ارتباط بيرسون. والسبب في استخدام معامل ارتباط الرتب هو سهولة حسابه ولتعذر التعامل مع القيم الاصلية بدقة كافية وخاصـة عندما تكون حسـاباتها طويلة ومعقدة ويشـيع اسـتخدام هذا المعامل عندما يكون عدد أزواج البيانات للمتغيرين قليلة نسبيا بحيث لا تزيد عن ثلاثين زوجا لهذ يهتم الباحث بالرتب للبيانات أكثر من اهتمامه بقيمها الحقيقية. ويرمز له بالرمز (r_s) ويحسب من المعادلة التالية: -

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

حيث: -

d_i = الفرق بين رتب المتغير الأول ورتب المتغير الثاني

مثال/ اوجد معامل ارتباط سبيرمان لقيم المتغيرين (x, y)

$Xi = 5, 3, 1, 4, 2$ $Yi = 2, 1, 4, 5, 3$

الحل: -

Xi	*Yi*	*di*	*(di)²*
5	2	3	9
3	1	2	4
1	4	-3	9
4	5	-1	1
2	3	-1	1
			$\sum(di)^2 = 24$

نطبق القانون: -

$$r_s = 1 - \frac{6*\sum(di)^2}{n(n^2-1)} = 1 - \frac{6*24}{5(5^2-1)} = 1 - \frac{144}{5*24} = 1 - 1.20 = -0.20$$

اذن العلاقة عكسية او سالبة

مقارنة بين معامل ارتباط بيرسون ومعامل ارتباط سبيرمان: -

١- لا يشـترط تسـاوي قيمة معامل ارتباط بيرسـون مع معامل ارتباط سـبيرمان وذلك لان هذين المعاملين مختلفان تماما من حيث الأهداف في استخدامهما والفروقات بينهما تكون طفيفة على الاغلب.

٢- معامل ارتباط بيرسـون ادق من معامل ارتباط سـبيرمان بسـبب ان معامل ارتباط بيرسـون يسـتخدم القيم الاصلية بينما سبيرمان يستخدم الرتب المشتقة من القيم الاصلية.

٣- سهولة حساب معامل ارتباط سبيرمان يجعله مفضلا على ا ستخدام معامل ارتباط بيرسون الذي يتضمن عمليات حسابية معقدة.

٤- يستخدم سبيرمان البيانات سواء كانت كمية ام نوعية ترتيبية حتى وان كان أحد المتغيرين كميا والأخر نوعيا بينما لا يمكن حساب معامل الارتباط لبيرسون الا إذا كان المتغيرين كميين.

واجب / احسب معاملي ارتباط بيرسون وسبيرمان من الجدول الاتي: -

Xi	*Yi*
8	2
3	2
6	5
6	5
1	2
3	7
6	5
3	8