الاحصاء الرياضي

م.م ريم طلال كامل

للــــعام الـــدراسي

(2023-2022)

مفردات المادة في الاحصاء الرياضي
م.م ريم طلال كامل

الفصل الاول // ١. مفاهيم الاحصاء

٢. انواع التوزيعات التكرارية

٣. العرض الهندسي للبيانات

الفصل الثاني // مقاييس النزعة المركزية

١. الوسط الحسابي

٢. الوسيط

٣. المنوال

الفصل الثالث // اولاً مقاييس التشتت المطلق

١. المدى

٢. الانحراف المتوسط

٣. الانحراف المعياري

٤. التباين

٥. الانحراف الربيعي

ثانياً // مقاييس التشتت النسبي

① معامل الاختلاف C.V

② الدرجة المعيارية Z

الفصل الرابع // الارتباط الخطي Linear correlation

١. الارتباط البسيط simple correlation

أ. pearson

ب. spearman

٢. الارتباط المتعدد multiple correlation

الفصل الخامس // اختبار الفرضيات

أ) اختبار الفرق بين وسطين حسابيين لعينتين مستقلتين

ب) اختبار الفرق بين وسطين حسابيين لعينتين غير مستقلتين

الفصل السادس // تحليل التباين

الفصل السابع // اختبار الاستقلالية

2021 /10/

المحاضرة الاولى :
م.م ريم طلال كامل

(1)

ج الاحصاء :

١) تعريف علم الاحصاء : الطريقة الاحصائية التي تمكننا من جمع وتنظيم الحقائق عن ظواهر متنوعة في صورة أرقام وبالتالي تحليلها لتفسير اتجاهات هذه الظواهر والعلاقات فيما بينها.

ويعرفه علماء آخرون على انه :-

(2) علم الاحصاء : مجموعة النظريات والقوانين والقواعد والاساليب التي تستخدم في دراسة ووصف وتحليل الظواهر المعبر عنها تعبيراً كمياً لاستخلاص واستنتاج معايير لهذه الظواهر وتفسيرها وربما التنبؤ لما ستؤول اليه حالة الظاهرة مستقبلاً.

ينقسم الاحصاء الى قسمين

(1) الاحصاء الوصفي — (2) الاحصاء الاستدلالي

أولاً: الاحصاء الوصفي Descriptive statistics

ويشمل الاساليب الاحصائية التي تستعمل في وصف مجموعة معينة من البيانات وتتضمن جمع البيانات والحقائق وكذلك تبويبها وتلخيصها وعرضها بأساليب عديدة وحساب بعض المقاييس والمؤشرات الاحصائية المختلفة فمثلاً :

١. عندما تقوم إحدى الجامعات أو الكليات بتسجيل أعداد

أساتذتها وموظفيها وخريجيها وفقاً للأقسام العلمية والوحدات الادارية مبوبة بجداول واستعمال أساليب العرض البياني وغيره فهذا هو إستعمال الاحصاء الوصفي.

ع. حينما تقوم إحدى الاندية بتسجيل اعداد لاعبيها وعدد مرات مبارياتهم وعدد مرات فوزهم على الاندية الأخرى وعرض هذه الاعداد برسم بياني بأحد أنواع العرض البياني فهذا استعمال للاحصاء الوصفي.

ثانياً :- الاحصاء الاستدلالي أو الاستنتاجي
Statistical Inference

ويشمل الطرق الاحصائية التي تهدف الى عمل أستنتاجات أو استدلالات للمجتمع عن طريق العينة أو المصدر التي جمعت منه البيانات وتشمل :-

أ. طرائق التقدير Estimation Methods
ب. اختبار الفرضيات Test of Hypotheses

خ) الطريقة الاحصائية في البحث العلمي :-

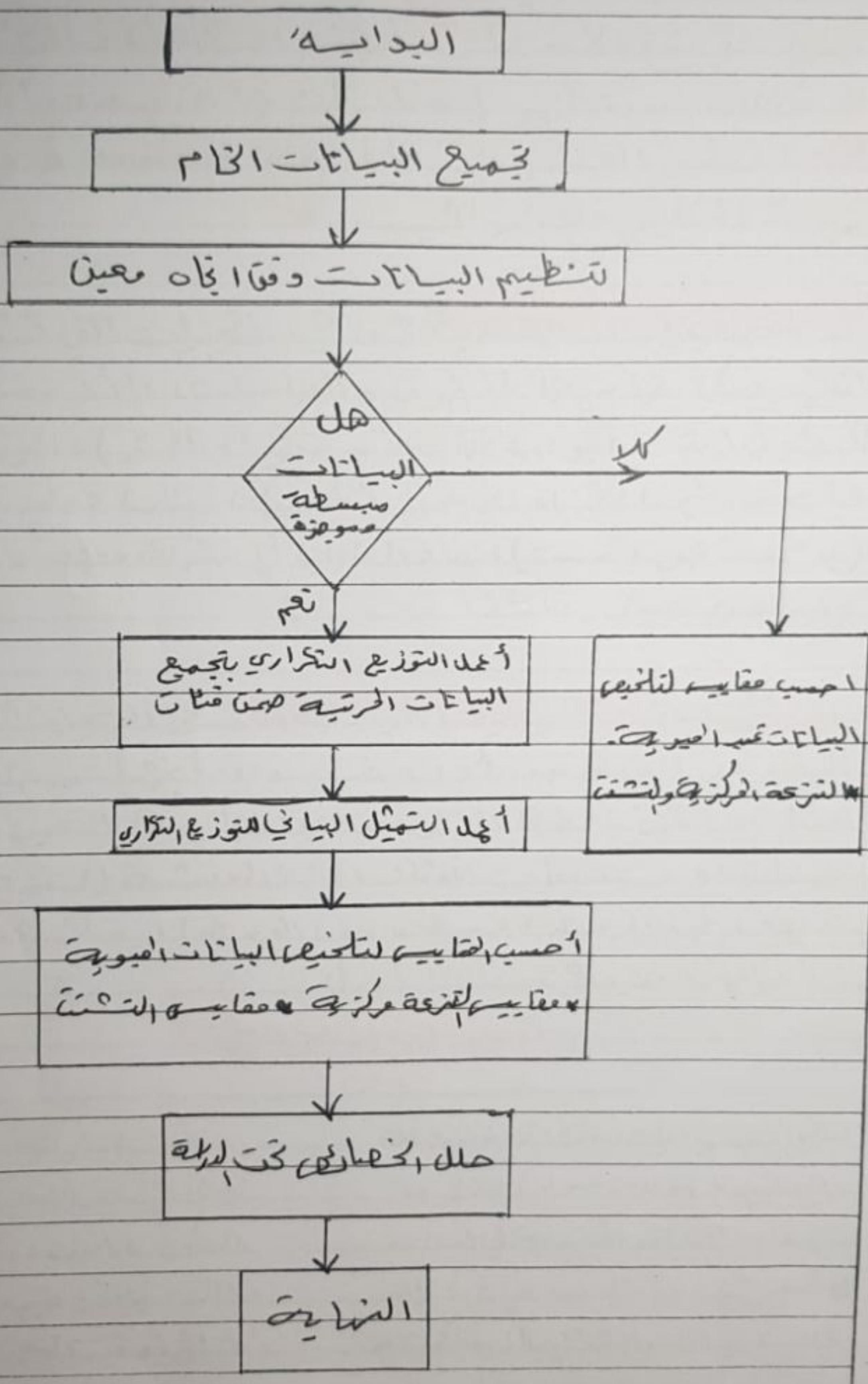
حينما ندرس علم الاحصاء لا نعني بذلك الاحصائية وإنما الطريقة الاحصائية وهي الطريقة العلمية المرهونة تطبيقها بإمكانية التعبير عن الظواهر المدروسة تعبيراً كمياً أو رقمياً، فهي أذن الطريقة التي تمكننا من جمع الحقائق عن الظواهر في صورة قياسات رقمية أو كمية.

والان الطريقة الاحصائية لها مراحل رئيسة هي :-

① تحديد مشكلة البحث ② جمع البيانات ③ تصنيف وتبويب البيانات
④ حساب المؤشرات او المعلمات ⑤ التفسير والتنبؤ ⑥ اتخاذ القرار المناسب

المخطط الانسيابي (1-1)

يوضح مراحل الطريقة الاحصائية



المتغيرات العشوائية وأنواعها

إن الظاهرة التي يراد دراستها إحصائياً عادة ما يرمز لها بأحد الرموز كأن يكون X وعليه فإن x_i تعني المفردة i ضمن المجتمع الإحصائي لهذه الظاهرة وبالتأكيد فإن كل ظاهرة ستظهر اختلافات في قيم مفرداتها وتدعى X بالمتغير العشوائي Random variable و x_i تدعى بقيمة المتغير العشوائي .

أ. المتغيرات الكمية :- هي المتغيرات التي تمثل الظواهر التي يمكن قياسها بالطرق والقياسات المألوفة مثل // الطول (يقاس بوحدة قياس سنتمتر او المتر) ، الوزن (بوحدة قياس غم ، كغم) ، الأجور (فلس ، دينار ، دولار) ، العمر (يوم ، شهر ، سنة) ، عدد العاملين في شركة معينة ، عدد اللاعبين في نادي معين وهكذا ..

ب. المتغيرات النوعية (الوصفية) :- هي المتغيرات التي تمثل الظواهر التي توصف أو تصنف ، بمعنى أنها غير خاضعة للقياس الكمي بالطرق المألوفة أي انها يمكن التعبير عنها بشكل صفات أو تصنيفات فمثلاً الحالة المعيشية (ممتازة ، جيدة جداً ، جيدة ، فوق الوسط ، وسط ، دون الوسط) او مثلاً الحالة الاجتماعية (أعزب ، متزوج ، مطلق ، ارمل ...) .

وتقسم المتغيرات الكمية الى

- متغيرات عشوائية متقطعة او منفصلة وهي قيم قابلة للعد سواء كانت محدودة مثل عدد الاهداف المسجلة في الساعة وغير محدودة مثل عدد حبات الحنطة في سنابل المحصول الزراعي لمزرعة كبيرة جداً.
- متغيرات عشوائية مستمرة او متصلة وهي قيم غير قابلة للعد مثل الطول ، الوزن ، الحجم ، الارتفاع وذلك لانها تمثل بقيم رقمية في مدى أو مجال أو فترة معين أي انها متصلة

مثل اذا كانت X تمثل عدد الابناء الذكور في كل أسرة
لمجموعة مكونة من 6 أسر فأن مجموعة القيم الممكنة الى
X يمكن ان يعبر عنها :-

$$F = \{ X ; x = 0, 1, 2, 3, 4, 5 \}$$

اذ يمكن عد عناصر هذه المجموعة اي انها قابلة للعد بالرغم
انها محدودة اي لها بداية العدد 0 ولها نهاية العدد 5
وبذلك يقال لـ X انها متغير عشوائي متقطع او منفصل
او مثلاً عدد الطلبة الاناث في الاقسام العلمية في كليات الشريعة ،
وقد تكون المجموعة غير محدودة مثل اذا كانت X متغير عشوائي
يشير الى حبات القمح (الحنطة) في سنابل هذا المحصول في مزرعة
كبيرة جداً ، فيمكن التعبير عن مجموعة القيم الممكنة الى X في
المجموعة

$$X : x = \{0, 1, 2, 3, 4, 5\}$$

حيث يمكن عد عناصر هذه المجموعة بالرغم من كونها مجموعة غير
محدودة (لها بداية وليس لها نهاية) فان X متغير عشوائي
متقطع او يسمى منفصل .

أما المتغيرات العشوائية مستمرة أو متصلة وهي المتغيرات غير
القابلة للعد مثل // اذا كان X متغير عشوائي يشير
الى وزن الموظف بالكغم فيمكن التعبير عن مجموعة القيم X
عن طريق :-

$$B = \{ X : 60 \leq X < 95 \}$$

أو مثلاً اذا كان X متغير عشوائي يمثل كمية النفط بالغالون
المستهلك من قبل مجموعة الاسر في مدينة معينة ، فيمكن
التعبير عن مجموعة قيم المتغير الآتي :-

$$C = \{ X : 0 \leq X \leq \infty \}$$

or

حيث b, a اعداد حقيقية $a \leq X \leq b$

مثل الوزن يكون ما بين

$$60 \leq X \leq 66$$

العينة : Sample

تعريف العينة :- مجموعة جزئية من مجتمع الدراسة يتم اختيارها بطريقة مناسبة وإجراء الدراسة عليها ومن ثم استخدام تلك النتائج وتعميمها.

وكذلك تعرف على أنها تمثل المجتمع الأصلي وتحقق أغراض البحث وتغني الباحث عن المشقات التي تواجهه أثناء البحث.

ويرمز للعينة بالرمز n والمجتمع N .

أنواع العينات

الاحتمالية

1. العينة العشوائية البسيطة
2. العينة العشوائية الطبقية
3. العينة الأسلوبية العشوائية
4. العينة المتعددة المراحل

غير الاحتمالية

1. العمدية
2. الحصصية

المحاضرة الثانية

7

عرض البيانات وفق الجداول

أولاً :- التوزيع التكراري Frequency Distribution
وهو جدول يلخص ويرتب البيانات للمتغير العشوائي التي سبق لنا ان جمعناها و صنفناها و قسمت الى عدد من المجاميع المرتبة تصاعدياً او تنازلياً. ويتكون الجدول من عمودين هما عمود الفئات (class) وعمود التكرارات (Frequencies)

وسيتم توضيح مكونات التوزيع التكراري لتكوين جدول توزيع تكراري كما في المثال التالي :-

لتكن $X_1, X_2, ..., X_n$ تمثل بيانات متغير عشوائي من عينة عشوائية بحجم (n) مفردة و نرغب في ترتيب هذه البيانات في جدول توزيع تكراري عدد فئاتها هو m ، بفرض ان X_L أصغر قيمة و X_u اكبر قيمة في مجموعة البيانات عندها يمكن ان نعرف مكونات التوزيع التكراري بالشكل التالي :-

1. المدى الكلي للتوزيع Total Range
ويمثل الفرق بين اكبر قيمة و اصغر قيمة في المجموعة مضافاً له العدد واحد . أو الفرق بين قيمة الحد الاعلى و قيمة الحد الادنى + 1 . ويرمز له بالرمز T.R

$$T.R = X_u - X_L + 1$$

* ويسمى طول أو مجال أو سعة الفئة ،

* بداية الفئة : القيمة التي تمثل الحد الادنى للفئة Lower Limit
* نهاية الفئة : القيمة التي تمثل الحد الاعلى للفئة Upper Limit

(8) Number of classes ② عدد فئات التوزيع

وتعني عدد المجاميع التي يتألف منها التوزيع التكراري وهنالك صيغ تقريبية يمكن من خلالها تحديد عدد فئات التوزيع أهمها :

a: صيغة Yule وهي

$$m = 2.5\sqrt[4]{n}$$

b: صيغة sturges وهي

$$m = 1 + 3.322 \log_{10} n$$

اذ تمثل n : حجم العينة

ملاحظة // عند تطبيق الصيغ المذكورة اعلاه (sturges و Yule) يتم تقريب الناتج لاقرب عدد صحيح .

③ طول الفئة Length of class

ويمكن حساب طول الفئة من خلال قسمة المدى الكلي للتوزيع (T.R) على عدد الفئات (m)

$$L = \frac{T.R}{m}$$

④ الحد الادنى والحد الاعلى للفئة

Lower and upper bound of class

لكل فئة من فئات التوزيع التكراري بداية ونهاية .

وتعني البداية الحد الادنى للفئة والنهاية تعني الحد الاعلى لها

⑤ مركز الفئة   Center of class

ويمثل مركز الفئة قيمة من قيم المتغير العشوائي التي تتوسط المسافة بين الحد الادنى والحد الاعلى للفئة. ويرمز للحد الادنى بالرمز L.L ويرمز للحد الاعلى L.U ومركز الفئة X ويتم حساب مركز الفئة بالصيغة الرياضية الآتية :-

$$X = \frac{L.U + L.L}{2}$$

L.L   مركز الفئة   L.U

⑥ تكرار الفئة   Class frequency

يمثل جزء من مفردات العينة التي تتصف بكونها تقع من حيث القيمة العددية ما بين حدي الفئة. حيث ان مجموع هذه الاجزاء يشكل عدد مفردات العينة n. $f_1+f_2+f_3+\dots+f_m=n$
ويرمز للتكرار بالرمز (f_i)

ملاحظة // ان منتصف الفئة والذي يمثل القيمة التي تتوسط مدى او طول الفئة يسمى بمركز الفئة مصطلح (center of class)

$$\text{مركز الفئة } X = \frac{\overset{\text{الحد الاعلى}}{\text{upper Limit}} + \overset{\text{الحد الادنى}}{\text{Lower Limit}}}{2}$$

(10) (في حالة البيانات متقطعة) المثال الأول //

كون جدول توزيع تكراري لعلامات (30) لاعب من أحد الاختبارات وكانت كما يلي :

46	49	48	58	54	50
40	62	37	48	54	75
54	48	59	45	34	58
47	61	49	44	68	39
63	56	43	57	40	45

الحل //

① نحدد أكبر قيمة X_U وأصغر قيمة X_L لإيجاد المدى الكلي

∴ $X_U = 75 \quad , \quad X_L = 34$

$$T.R = X_U - X_L + 1$$
$$= 75 - 34 + 1$$
$$= 42$$

∴ المدى 42

② نحدد قيمة (m) أي عدد الفئات

$$m = 2.5\sqrt[4]{n} = 2.5\sqrt[4]{30} = 5.85 \simeq 6$$

∴ عدد الفئات 6

③ نجد طول الفئة

$$L = \frac{T.R}{m} = \frac{42}{6} = 7$$

∴ طول الفئة 7

∴ يكتب جدول التوزيع التكراري بالشكل الآتي : (11)

ت	التكرارات f_i	الفئات class	الحد الاعلى X_u	الحد الادنى X_L	مركز الفئة X
1	5	34-40	40 X_L+L-1	34 X_{L_1}	37
2	6	41-47	47 X_L+2L-1	41 X_L+L	44
3	9	48-54	54 X_L+3L-1	48 X_L+2L	51
4	6	55-61	61 X_L+4L-1	55 X_L+3L	58
5	3	62-68	68 X_L+5L-1	62 X_L+4L	65
6	1	69-75	75 X_L+6L-1	69 X_L+5L	72

∴ عدد اللاعبين الذين درجاتهم مابين (34-40) هم 5 لاعبين وكذلك لبقية الدرجات لبقية اللاعبين.

ملاحظة // الحد الاعلى = الحد الادنى + طول الفئة − 1 :
الحد الادنى = الحد الادنى السابق + طول الفئة

ملاحظة // ان مجموع التكرارات يكون مساوي لحجم العينة

أي بمعنى $\sum f_i = n$

∴ $\sum f_i = 30$ = حجم العينة

مركز الفئة و تحسب بالصيغة الاتية :-

$$X_1 = \frac{L.L + L.U}{2} = \frac{34 + 40}{2} = 37$$

$$X_2 = \frac{41 + 47}{2} = 44$$

$$\vdots$$

$$X_6 = \frac{69 + 75}{2} = 72$$

المثال الثاني // (في حالة المتغيرات المستمرة)

تمثل البيانات التاليـة أوزان 30 لاعب متقدم لاختيار رياضة رفع الاثقال. المطلوب تبويب البيانات في صورة جدول توزيع تكراري.

~~66.2~~	~~81.5~~	~~89~~	~~87~~	~~103~~	~~71.3~~
~~60~~	~~85~~	~~74~~	~~106~~	~~69.4~~	~~85~~
~~97~~	~~66.2~~	~~76~~	~~106~~	~~65~~	~~88.1~~
~~84~~	~~67~~	~~101~~	~~86~~	~~89~~	~~101.3~~
~~82~~	~~93~~	~~99.5~~	~~86~~	~~87~~	~~92~~

خطوات الحل //

1. نحدد أكبر قيمة أي أكبر وزن مذكور في عينة البحث وكذلك أصغر قيمة.

$X_U = 106$ ، $X_2 = 55$

(13) نجد المدى الكلي $T.R = X_U - X_L + 1$

$= 106 - 55 + 1 = 52$

(2) نجد قيمة (m) اي عدد الفئات
وذلك باستخدام صيغة Yule

$m = 2.5 \sqrt[4]{n}$

$= 2.5 \sqrt[4]{30} \simeq$ يجب التقريب

$= 5.85 \simeq 6$

(3) نجد طول الفئة

$L = \frac{T.R}{m} = \frac{52}{6} = 8.6 \simeq 9$

(4) ترتيب جدول التوزيع التكراري

ت	التكرار f_i	الفئات class	الحد الاعلى للفئة	الحد الادنى للفئة	مركز الفئة X
1	2	55 اقل من 64	$55+9-1$ $X_L + L - 1 = 64$	X_{L_1} 55	59.5
2	4	64 اقل من 73	73	$X_{L_1} + L$ 64	68.5
3	3	73 اقل من 82	82	$X_L + 2L$ 73	77.5
4	12	82 اقل من 91	91	$X_L + 3L$ 82	86.5
5	4	91 اقل من 100	100	91	95.5
6	5	100 - 109	109	100	104.5
	$\therefore \sum f_i = n = 30$				

العرض الهندسي للبيانات (17)

تعتبر الاشكال والرسوم البيانية من المعالم الاحصائية المؤثرة في توضيح توزيع وانتشار ونمو البيانات وهذه الاشكال والصور تعرض بعدة صيغ تعطي جمالية اكثر للقارئ في فهم مايدور حوله من الظواهر الاجتماعية المختلفة الموضحة في تلك الاشكال والرسوم.

ولهذا النوع من الاشكال أهمية كبيرة في دراسة خصائص التوزيعات التكرارية فضلاً عن حساب بعض المؤشرات الاحصائية.

أولاً // المدرج التكراري

عبارة عن مجموعة من المستطيلات قاعدة كل منها تمثل طول الفئة في التوزيع التكراري وارتفاع كل منها يمثل قيمة التكرار المقابل لتلك الفئة.

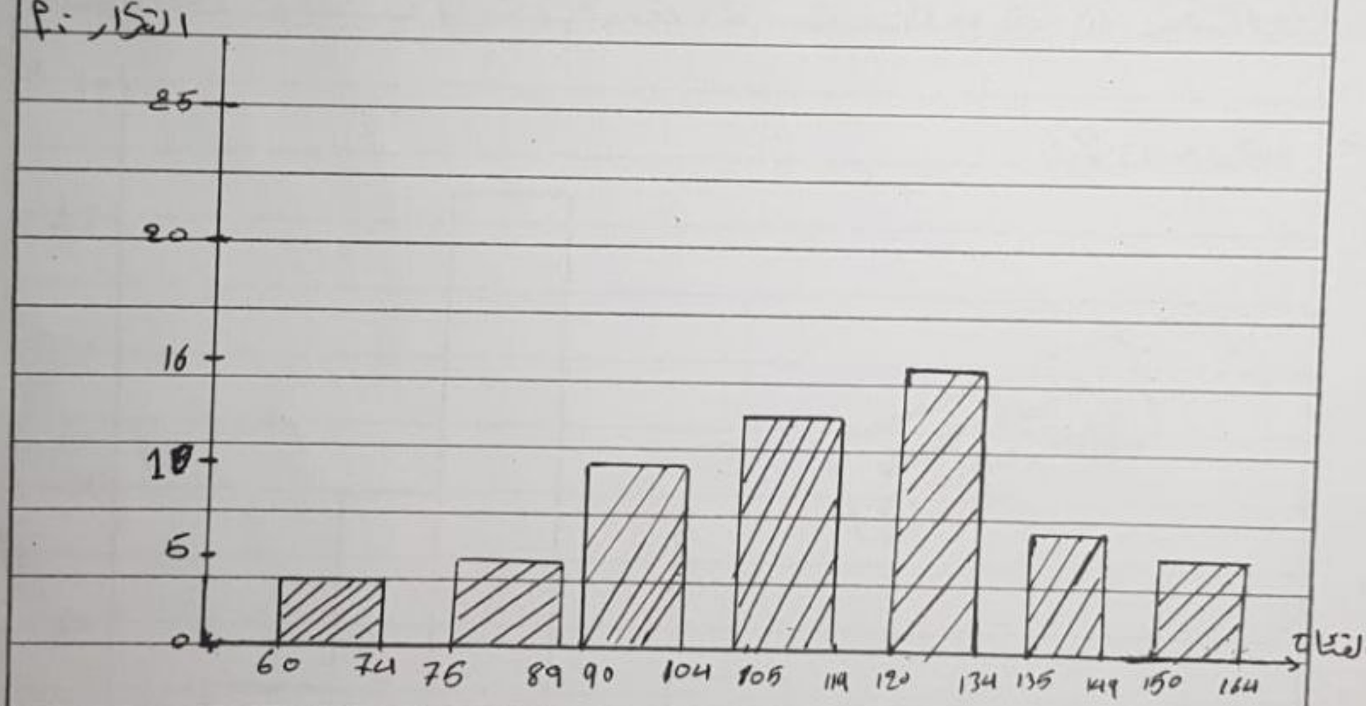
هذه المستطيلات تكون منفصلة عن بعضها في حالة المتغيرات المتقطعة وتكون متصلة مع بعضها في حالة المتغيرات المستمرة وحسب تسلسل فئات التوزيع.

مثال // (1) الآتي توزيع تكراري يمثل 60 عداءاً حسب المسافات التي قطعوها خلال اسبوع يطلب رسم مدرج تكراري لهذا التوزيع.

(18)

المسافات المقطوعة (كم) الفئات Classes	عدد العدائين التكرارات f_i
60 - 74	4
75 - 89	5
90 - 104	10
105 - 119	12
120 - 134	16
135 - 149	7
150 - 164	6
	$\sum f_i = n = 60$

∴ المتغير العشوائي من النوع المتقطع وعليه فإن المدرج التكراري سيكون بالشكل التالي :

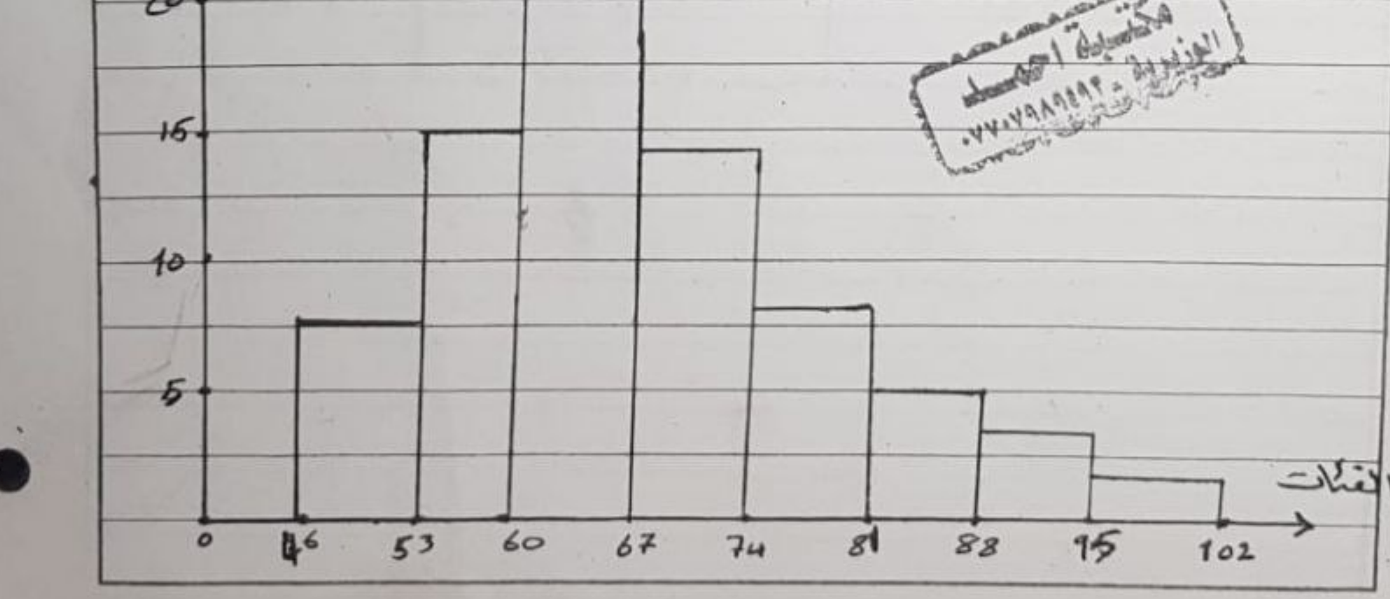


مثال (2)

الجدول التالي يبين التوزيع التكراري لأوزان عينة من طلبة أحدى الكليات قوامها 100 طالب.
المطلوب / رسم المدرج التكراري لهذا التوزيع.

عدد الطلاب (التكرار f_i)	الاوزان (الفئات Classes)
7	46-
15	53-
27	60-
~~21~~	~~67-~~
14	74-
8	81-
5	88-
3	95-102

∴ التغير من النوع المستمر فإن المدرج التكراري سيكون بالشكل التالي:



ثانياً // المضلع التكراري

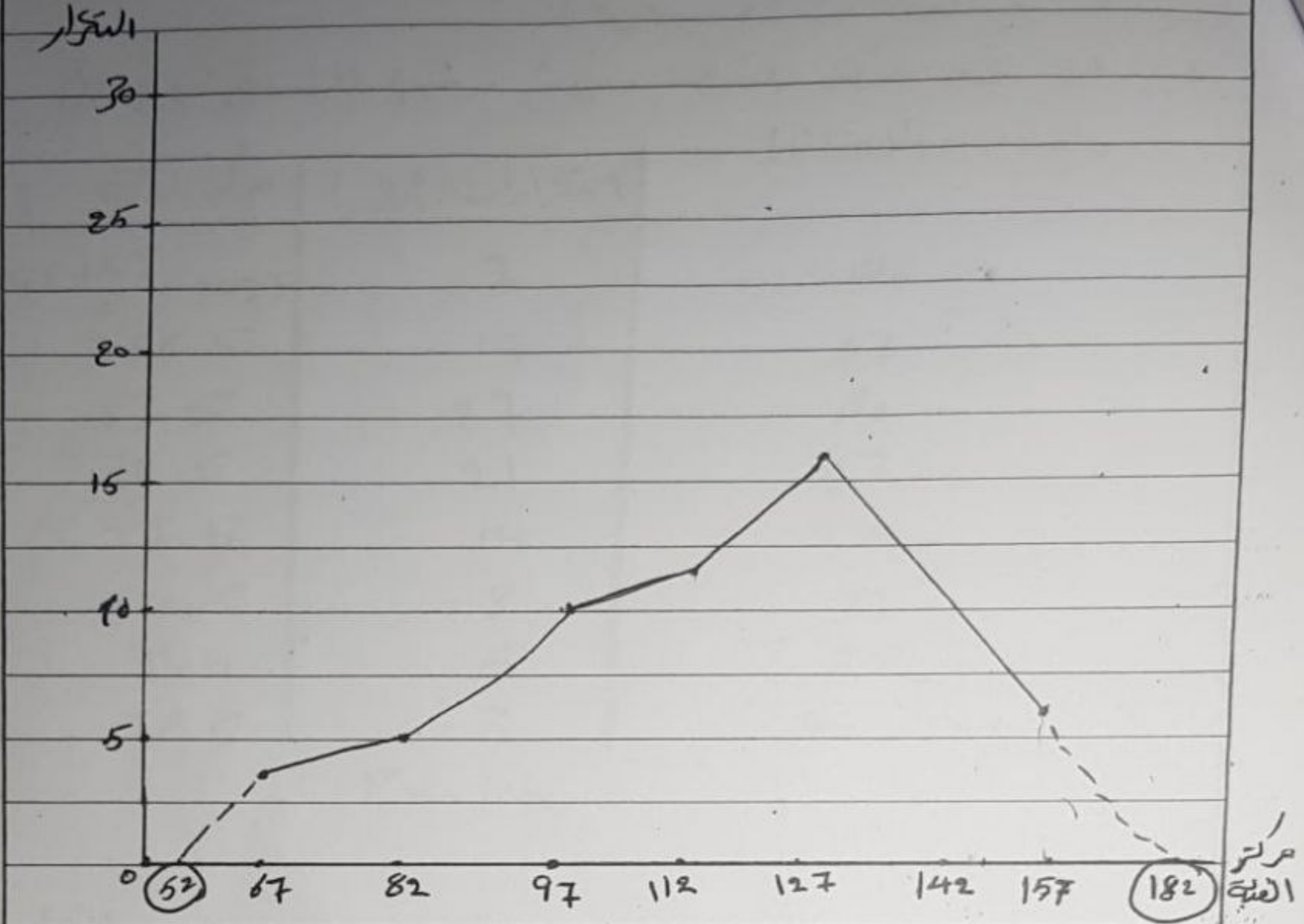
عبارة عن عدد من المستقيمات المتصلة مع بعضها على شكل سلسلة ونقطة اتصال المستقيم بالآخر تقابل مركز الفئة وهذا يعني انك عند رسم مضلع تكراري يستوجب الامر ايجاد مراكز الفئات ومن ثم رسم المضلع.

أرسم المضلع التكراري لكل من المثالين السابقين :-

مثال ص 17 (في حالة البيانات المتقطعة)
و ص 18

مركز الفئة (عدد العمائين)	التكرار f_i	الفئات (المسافات المقطوعة)
$\frac{60+74}{2}=67$	4	60 - 74
82	5	75 - 89
97	10	90 - 104
112	12	105 - 119
127	16	120 - 134
142	7	135 - 149
157	6	150 - 164
	$\sum f_i = 60$	

21



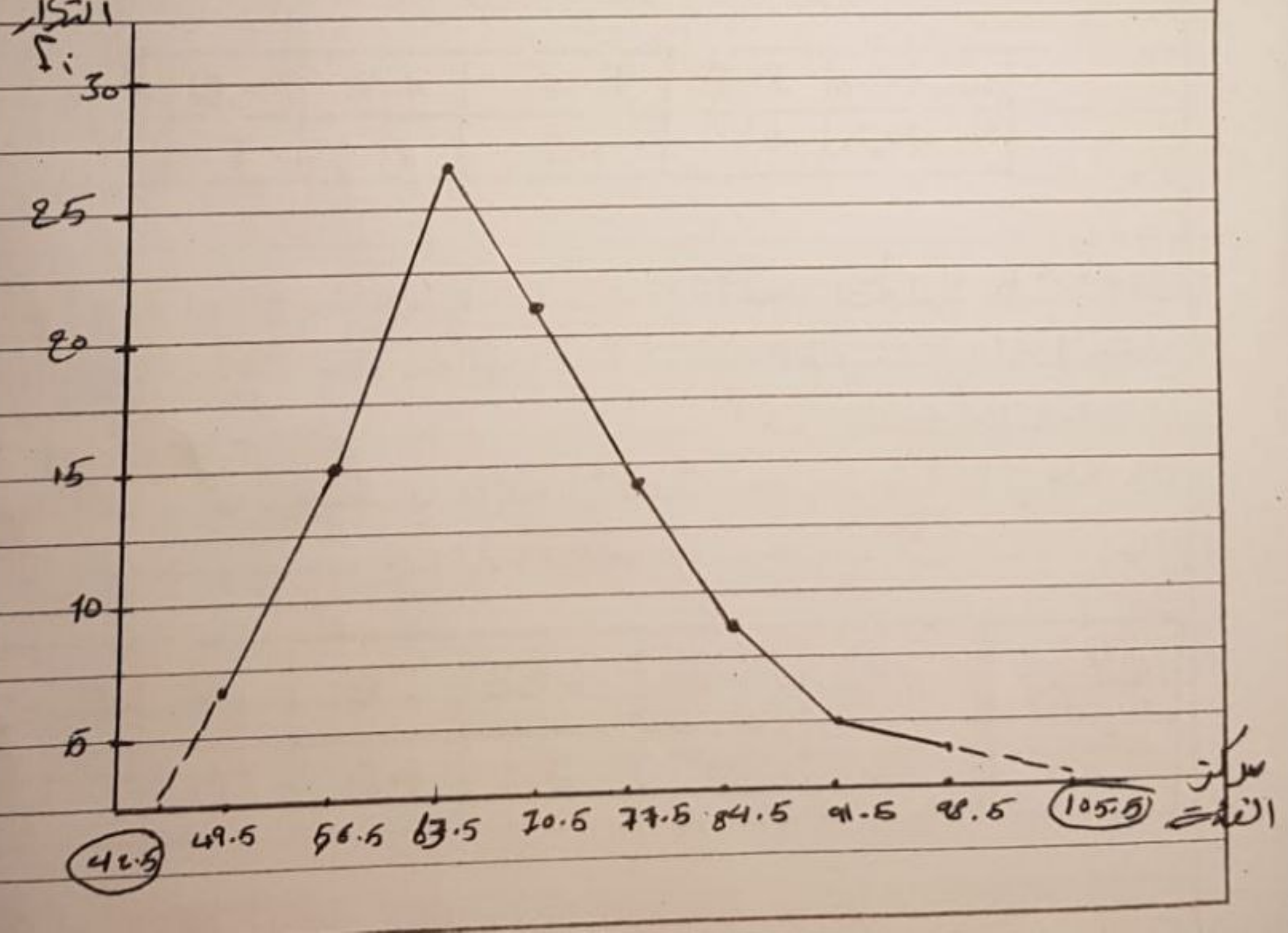
تم الغلق عند المركزين الوهميين 182 , 52

ارسم المضلع التكراري للمثال المذكور ص 19
(في حالة البيانات المستمرة)

يجب ان نجد مركز الفئة ومن ثم ترتيب النتائج في جدول

الفئات (الاوزان)	التكرارات (عدد الطلاب)	مركز الفئة
46-	7	$\frac{46+53}{2}=49.5$
53-	15	56.5
60-	27	63.5
67-	21	70.5
74-	14	77.5
81-	8	84.5
88-	5	91.5
95-102	3	98.5

$\Sigma n = 100$



H.W

اولاً// سجلت عدد مباريات لمنتخبات مختلفة في عموم العراق وكانت بالاعداد الاتية :-

97, 88 76 108 66 78
73 49 53 70 101 90

المطلوب// ① كون جدول توزيع تكراري
② ارسم المدرج التكراري

ثانياً// في الجدول التالي يوضح عدد الاطفال في 55 عائلة
المطلوب// ① ارسم المدرج التكراري.
② ارسم المضلع التكراري.

عدد الاطفال	0-3	4-7	8-11	12-15	16-19
عدد العائلات	10	24	11	7	3

ثالثاً// فيما يلي جدول التوزيع التكراري لدرجات التي حصل عليها طلاب المرحلة الثالثة البالغ عددهم 169 طالب في مادة الاحصاء الرياضي.
المطلوب// ① ارسم المدرج التكراري.
② ارسم المضلع التكراري.

الفئات الدرجات	30-	40-	50-	60-	70-	80-90
التكرار f_i	18	13	55	54	19	10

الفصل الثاني

مقاييس النزعة المركزية
Measures of central Tendency

سميت بمقاييس النزعة المركزية لامكانية تمثيلها لبيانات الدراسة (اي الظاهرة المدروسة) بقيمة واحدة تتركز في وسطها.

⇐ أهم مقاييس النزعة المركزية هي

اولاً // الوسط الحسابي Arithmetic Mean

أحد ابرز وأهم مقاييس النزعة المركزية لامتيازه بخصائص جيدة أهمها سهولة حسابه ويرمز له $\bar{X}$.

1) حساب الوسط للبيانات غير المبوبة
Mean for ungrouped Data

لتكن $X_1, X_2, X_3, \ldots, X_n$ تمثل البيانات المستحصل عليها عن المتغير العشوائي X على أساس عينة من المفردات قوامها n. فان الوسط الحسابي لهذه البيانات ما هو الا مجموع قياسات مفردات هذه العينة على عددها:

وتحسب وفق الصيغة الرياضية الآتية :-

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\sum_{i=1} X_i}{n}$$

$$\bar{X} = \frac{\sum^{n} X_i}{n} \qquad i = 1, 2, \ldots, n$$

حيث

n : حجم العينة

مثال/ البيانات التالية تمثل كمية المطر الساقط سنوياً بالمليمترات على مدينة اربيل خلال فترة خمس سنوات

520 350 450 380 400

فما هو معدل او متوسط او الوسط الحسابي لسقوط الامطار في مدينة اربيل خلال هذه الفترة ؟

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{520 + 350 + 450 + 380 + 400}{5} =$$

$\therefore \bar{X} = 420$ مليمتر

مثال/ البيانات التالية تمثل اوزان عينة من اللاعبات قوامها 15 لاعبة أوجد معدل أوزان اللاعبات.

50.2 , 60.9 , 68.3 , 59.2 , 58.1

62.3 65.3 52.9 61.5 63.2

59.1 69.3 64.2 65.2 56.6

(26)

② أحتساب الوسط الحسابي لبيانات مبوبة
Mean for grouped Data

البيانات المبوبة هي عبارة عن قيم او تكرارات مصنفة حسب فئات معينة وكل فئة يقابلها عدد معين من التكرارات التي تمثل الظاهرة.

لتكن $X_1, X_2, ..., X_m$ تمثل مراكز الفئات لجدول توزيع تكراري عدد فئاته m وان $f_1, f_2, ..., f_m$ تمثل التكرارات المقابلة لهذه الفئات، عندئذ يستخرج الوسط الحسابي لهذا التوزيع وفق ما يلي :-

$$\bar{X} = \frac{\sum_{i=1}^{m} f_i X_i}{\sum_{i=1}^{m} f_i}$$

أذ ان :-

f_i = التكرارات
X_i = مركز الفئة
m = عدد الفئات

خطوات ايجاد الوسط الحسابي $\bar{X}$

1. يتم حساب مراكز الفئات X_i
2. تكوين عمود يمثل حاصل ضرب مركز الفئة بمقدار تكرارها (اي ايجاد $f_i X_i$)

3. قسمة مجموع ($f_i X_i$) على التكرارات الكلية

اي مجموع عمود $f_i X_i$ يقسم على مجموع التكرارات.

مثال/ الجدول للتوزيع التكراري الآتي يبين عينة من انواع الرياضة قوامها 75 نوع من الرياضة حسب عدد اللاعبين ضمن كل رياضة، يطلب حساب متوسط او معدل عدد اللاعبين في هذه العينة.

عدد اللاعبين	2 - 4	5 - 7	8 - 10	11 - 13	14 - 16	17 - 19	20 - 22
عدد انواع الرياضة	8	12	20	13	10	8	4

خطوات الحل // ∵ البيانات مبوبة ∴ نعمل الجدول التالي ∴

الفئات (عدد اللاعبين)	التكرارات f_i (انواع الرياضة)	مراكز الفئات	$f_i X_i$
2 - 4	8	3	24
5 - 7	12	6	72
8 - 10	20	9	180
11 - 13	13	12	156
14 - 16	10	15	150
17 - 19	8	18	144
20 - 22	4	21	84
	75		810

$$\therefore \quad \overline{X} = \frac{\sum_{i=1}^{m} f_i X_i}{\sum_{i=1}^{m} f_i} = \frac{810}{75} = 10.8$$

مخطط بسيط يوضح مقاييس النزعة المركزية

الوسط الحسابي $\bar{X}$

① في حالة البيانات غير المبوبة

$$\bar{X} = \frac{\sum x_i}{n}$$

x_i = القيم
n = حجم العينة

② في حالة البيانات المبوبة

$$\bar{X} = \frac{\sum f_i x_i}{\sum f_i}$$

x_i : مركز الفئة
f_i : التكرارات

الوسيط Me

① اذا كانت عدد البيانات فردي

⇓

ويقابل القيمة التي تسلسلها $\frac{n+1}{2}$

② اذا كانت عدد البيانات زوجي

⇓

الوسط الحسابي للقيمتين التي تسلسلها $\frac{n}{2}$ ، $\frac{n}{2}+1$

* يتم في حالة ترتيب البيانات المعطاة في السؤال ترتيباً تصاعدياً أو تنازلياً.

المنوال M_o

① في حالة البيانات المبوبة

هو العدد الاكثر تكراراً $= M_o$ من بين كل الاعداد

② في حالة البيانات المبوبة

مركز الفئة الذي يقابل أكبر تكرار $= M_o$

29

HW

اولاً // اوجد معدل اعمار الاطفال المذكورة بالشكل الاتي:- 5 , 6 , 13 , 2 , 7 , 9 , 4

ثانياً // في الجدول الاتي نتائج لعبة كرة السلة لعدد من اللاعبات. اوجد الوسط الحسابي.

الفئات (النقاط)	10-19	20-29	30-39	40-49	50-59	60-69
التكرار (عدد اللاعبات)	10	7	4	4	3	2

ثالثاً // الجدول الاتي يبين اوزان تم تسجيلها من قبل لاعبي رفع الاثقال في البطولات المحلية والقارية احسب المتوسط.

الفئات (الاوزان)	100-	120-	140-	160-	180-	200-220
التكرار (عدد اللاعبين)	4	5	6	7	10	8

رابعاً:- البيانات التالية اوجد الوسط الحسابي والوسيط والمنوال:

26, 33, 29, 30, 32, 32, 28, 32, 27

خامساً // اوجد الوسط الحسابي والمنوال والوسيط.

6, 5, 8, 5, 7, 6, 4

سادساً // فيما يلي جدول التوزيع التكراري لعدد الاطفال في 55 عائلة اوجد كلاً من (الوسط الحسابي، المنوال)

الفئات عدد الاطفال	0-3	4-7	8-11	12-15	15-19
عدد العائلات f	10	24	11	7	3

(31)

ثانياً :- الوسيط Median

يعرف الوسيط على انه القيمة التي تتوسط قيم مجموعة معينة مرتبة تصاعدياً او تنازلياً. بحيث ان عدد القيم قبله تساوي عدد القيم بعده.
ويعرف ايضاً بانه القيمة التي تقسم مجموعة مرتبة من البيانات تصاعدياً او تنازلياً الى نصفين.

~~1) الوسط الحسابي~~

1) حساب الوسيط في حالة البيانات غير المبوبة :-

a) اذا كان عدد القيم n عدد فردي فان الوسيط هو القيمة التي ترتيبها $\left(\frac{n+1}{2}\right)$ اي ان الوسيط الذي يرمز له بالرمز M_e يكون كما يلي :-

$$Me = X_{\underbrace{\left(\frac{n+1}{2}\right)}_{\text{رتبة المتغير}}}$$

مثال// أوجد الوسيط لدرجات طالب في 5 امتحانات لمادة الاحصاء اذا كانت الدرجات كالآتي :-

80, 82, 76, 87, 84

الحل// نرتب الدرجات تصاعدياً او تنازلياً (اختياري)

87, 84, $\underline{\underline{82}}$, 80, 76

وبما ان عدد الدرجات فردي يعني $n=5$ فالوسيط يحسب بالقانون او وفق الصيغة الرياضية التالية :-
وهو القيمة التي ترتيبها $\frac{n+1}{2} = \frac{5+1}{2} = \frac{6}{2} = 3$

$\therefore Me = X_3 = 82$

b) اذا كانت n تمثل عدد زوجي فان الوسيط هو الوسط الحسابي للقيمتين اللتين ترتيبهما

$\frac{n}{2}+1$ و $\frac{n}{2}$

اي ان الوسيط يمثل الوسط الحسابي لقيمتي X بعد الترتيب التصاعدي او التنازلي اللتين تسلسلهما على التوالي هو

$$\boxed{\frac{n}{2}+1 \quad , \quad \frac{n}{2}} \therefore Me = \frac{X_{\frac{n}{2}+1} + X_{\frac{n}{2}}}{2}$$

مثال/ جد الوسيط لعمر الفرد لمجموعة من اعمار عينة افرادها 12 فرد.

20, 22, 19, 26, 24, 27, 28, 29, 29, 18, 20, 23, 25

الحل//

∴ البيانات غير مبوبة
نرتب القيم ترتيباً تصاعدياً أو تنازلياً (اختياري)

18, 19, 20, 20, 22, <u>23</u>, <u>24</u>, 25, 26, 27, 28, 29

القيم زوجية فان ترتيب الوسيط هو

$\frac{n}{2}+1 = \frac{12}{2}+1 = 7$ و $\frac{12}{2} = 6$

وهذا يعني ان القيمتان السادسة والسابعة هما اللتان تحددان قيمة الوسيط.
وبذلك فان الوسيط لعمر الفرد لهذه المجموعة يمثل الوسط الحسابي لهاتين القيمتين

الوسيط ⇐ $Me = \frac{23 + 24}{2} = 23.5$

مثال 3 // أحرز لاعب كرة سلة في مباريات الدوري العام الأهداف الآتية :

15 , 12 , 21 , 16 , 18

أوجد الوسيط .

الحل // ∵ البيانات غير مبوبة ∴ ترتب ترتيباً تصاعدياً

12 , 15 , $\underline{16}$, 18 , 21

∵ عدد n فردي = 5

$$X\left(\frac{n+1}{2}\right) = X\left(\frac{5+1}{2}\right) = X_3$$

∴ قيمة التسلسل الثالث هو الوسيط

∴ الوسيط هو $Me = 16$

مثال 4 // جد الوسيط للبيانات التالية

51 , 52 , 62 , 65 , 68 , 75 , 79 , 81

وبما أن عدد القيم هو عدد زوجي $n=8$ فالوسيط هو الوسط الحسابي للقيمتين اللتين ترتيبهما

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

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

$$Me = \frac{X_{(\frac{n}{2}+1)} + X_{(\frac{n}{2})}}{2} = \frac{X_5 + X_4}{2} = \frac{68+65}{2}$$

$$\therefore Me = \frac{133}{2} = 66.5$$

ثالثاً :. المنوال Mode

ويعرف على انه تلك القيمة التي تتكرر اكثر من غيرها من بين مجموعة من القيم او انها القيمة الشائعة من بين مجموعة من القيم .

١. حساب المنوال في حالة البيانات غير المبوبة

لتكن لدينا مجموعة من البيانات $X_1, X_2, \ldots, X_n$ ، عندئذ فأن المنوال هو تلك القيمة التي تتكرر اكثر من غيرها من بين هذه المجموعة .

مثال // للبيانات التالية أوجد المنوال

2, 3, 2, 4, 2, 5, 4, 4, 5, 4, 6, 8, 9, 4, 7, 3, 7, 6

الحل // :- البيانات غير مبوبة :. يتبين من هذه المجموعة بأن العدد (4) قد تكرر اكثر من غيره من الاعداد وعليه فان المنوال لهذه المجموعة = 4

$$\therefore M_o = 4$$

مثال // جد المنوال للبيانات الاتية :

2, 4, 3, 6, 8, 7, 10, 12

الحل //

البيانات غير مبوبة

:. واضح من هذه المجموعة انه لا يوجد عدد مكرر اكثر من غيره وعليه لا يوجد منوال لهذه المجموعة

:. المنوال = $M_o = Zero$

مثال // جد المنوال للبيانات التالية :-

5, 4, 7, 8, 10, 7, 3, 5, 9, 7, 5, 3

الحل // ∵ البيانات غير مبوبة
∴ واضح بهذه المجموعة ان العددين (5, 7)
قد تكررت اكثر من غيرها من الاعداد.

اذ من الممكن ايجاد اكثر من منوال واحد لمجموعة من البيانات

وعليه فأن المنوال لهذه المجموعة سيكون

$$Mo = (5, 7)$$

② حساب المنوال في حالة البيانات المبوبة

ليكن لدينا توزيع تكراري عندئذ فأن المنوال يمثل قيمة مركز الفئة التي تقابل اكبر تكرار.

مثال // في جدول التوزيع التكراري الاتي بيانات تمثل توزيع 40 لاعب حسب اوزان الاثقال التي رفعها اللاعب المطلوب تحديد المنوال لهذا التوزيع .

الفئات (اوزان الاثقال)	60-74	75-89	90-104	105-119	120-134
التكرارات f_i	2	6	14	10	8

الحل // ∵ البيانات مبوبة ∴ نبحث عن اكبر تكرار موجود في هذا التوزيع وهو العدد (14) الذي يقابل الفئة (90-104) وعليه فان المنوال في هذا التوزيع هو مركز هذه الفئة اي ان :-

المنوال $Mo = \frac{90+104}{2} = \frac{194}{2} = 97$

الفصل الثالث //

مقاييس التشتت :- يعرف التشتت بأنه تباعد أو انتشار قيم مجموعة من المفردات عن بعضها البعض أو عن قيمة معينة ثابتة (كالوسط الحسابي مثلاً).

ان الهدف من معرفة ودراسة التشتت البيانات هو معرفة وتكوين فكرة عن مدى تجانس قيم مجموعة من المفردات وهذا يعني ان دراسة التشتت أمر مفيد في اجراء المقارنة بين قيم مجموعتين أو أكثر من البيانات عن ظاهرة معينة.

أولاً // مقاييس التشتت المطلق

① المدى Range

يعتبر المدى ابسط أنواع مقاييس التشتت المطلقة ويعرف على انه يحسب الفرق بين اكبر قيمة في مجموعة من البيانات واصغر قيمة فيها. ويرمز له بالرمز R

a - حساب المدى في حالة البيانات غير المبوبة

$$R = X_u - X_L$$

حيث ان :-

X_u : أكبر قيمة

X_L : اصغر قيمة

مثال // أوجد المدى للبيانات التالية !

2, 5, 3, 8, 7, 10, 9, 12, 15

الحل // بما ان البيانات غير مبوبة :. نجد اكبر قيمة واصغر قيمة

$$R = 15 - 2 = 13$$

مثال // جد المدى لمجموعة من القيم
14, 15, 13, 7, 12, 10, 7, 13, 10, 8, 16, 11

الحل //

∵ البيانات غير مبوبة
∴ المدى
نجد ان اعلى قيمة هي
16 واقل قيمة هي 7

$$R = X_u - X_L$$
$$R = 16 - 7$$
$$R = 9$$

b) حساب المدى في حالة البيانات المبوبة

R = اكبر فئة - اصغر فئة

مثال / جد المدى للتوزيع التكراري التالي :

فئات	(1-3)	(4-6)	(7-9)	(10-12)	(13-15)
تكرارات	2	5	10	6	3

الحل // ∵ البيانات مبوبة

∴ اكبر فئة هي 15 واصغر فئة هي 1 ضمن جدول التوزيع التكراري.

R = اكبر قيمة - أصغر قيمة
$= 15 - 1$
$= 14$

② الانحراف المتوسط Mean deviation

ويعرف الانحراف المتوسط بانه مجموع الانحرافات المطلقة لقيم المتغير العشوائي عن نقطة اختيارية مثل A مقسوماً على عدد هذه القيم. غالباً ما تختار النقطة A لان تكون احد مقاييس النزعة المركزية الثلاث (الوسط الحسابي الوسيط، المنوال)

a) حساب الانحراف المتوسط في حالة البيانات غير المبوبة

$$M.D = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

حيث

$\bar{x}$: يمثل الوسط الحسابي في حالة البيانات غير المبوبة

x_i : القيم.

n : حجم العينة.

مثال // للبيانات التالية جد الانحراف المتوسط

2, 3, 4, 5, 5, 6, 7, 10, 13, 14, 19

الحل // :- البيانات غير مبوبة

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

١. يجب استخراج قيمة $\bar{x}$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{2+3+4+5+5+6+7+ \cdots +19}{11}$$

$$\bar{x} = \frac{88}{11} = 8$$

نعمل الجدول التالي لايجاد قيمة $|x_i - \bar{x}|$

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$
2	$-6 = 2-8$	6
3	$-5 = 3-8$	5
4	$-4 = 4-8$	4
5	$-3 = 5-8$	3
5	$-3 = 5-8$	3
6	$-2 = 6-8$	2
7	$-1 = 7-8$	1
10	$2 = 10-8$	2
13	$5 = 13-8$	5
14	$6 = 14-8$	6
19	$11 = 19-8$	11
		48

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

$\therefore$

$$M.D = \frac{48}{11} = 4.36$$

b) حساب الانحراف المتوسط في حالة البيانات المبوبة

$$M.D = \frac{\sum_{i=1} f_i |x_i - \bar{x}|}{\sum_{i=1} f_i}$$

حيث ان :-

x_i : تمثل مركز الفئة .

$\bar{x}$: الوسط الحسابي في حالة البيانات المبوبة .

f_i : التكرار .

$n = \sum f_i$: حجم العينة .

مثال // الآتي توزيع لعينة بعدد 19 من النوادي الرياضية حسب عدد افراد لاعبيها ، المطلوب حساب الانحراف المتوسط.

عدد اللاعبين الفئات	(4-8)	(9-13)	(14-18)	(19-23)	(24-28)
النوادي الرياضية (f_i)	3	4	6	2	4

الحل // ∵ البيانات مبوبة

∴ يكتب قانون او الصيغة الرياضية الخاصة بالانحراف المتوسط في حالة البيانات المبوبة

$$M.D = \frac{\sum_{i=1}^{n} f_i |x_i - \bar{x}|}{\sum_{i=1}^{n} f_i}$$

1. نحتاج الى استخراج قيمة مركز الفئة x_i
2. نحتاج الى استخراج قيمة الوسط الحسابي $\bar{x}$ في حالة البيانات المبوبة.

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{304}{19} = 16$$

$$M.D = \frac{\sum f_i |x_i - \bar{x}|}{\sum f_i} = \frac{100}{19} = 5.26$$

42 (المطلوب ايجاده

معطى معطى

$f_i\lvert x_i-\bar{x}\rvert$	$\lvert x_i-\bar{x}\rvert$	$x_i-\bar{x}$	$f_i x_i$	مركز الفئة x_i	الفئات	التكرار f_i
30	10	-10	18	6	4 - 8	3
20	5	-5	44	11	9 - 13	4
0	0	0	96	16	14 - 18	6
10	5	5	42	21	19 - 23	2
40	10	10	104	26	24 - 28	4
100			304			

③ الانحراف المعياري Standard devation

ويدعى في بعض الأحيان بالانحراف القياسي . اذ يعتبر هذا المقياس أفضل مقاييس التشتت على الاطلاق.

وكذلك يعرف بأنه الجذر التربيعي الموجب لمتوسط مجموع مربعات انحرافات القيم للمتغير العشوائي عن وسطها الحسابي .

a) حساب الانحراف المعياري في حالة البيانات غير مبوبة ويحسب بالصيغة الرياضية التالية :-

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

حيث ان :-

x_i : القيم .

$\bar{x}$: الوسط الحسابي للقيم في حالة البيانات غير مبوبة

n : حجم العينة .

مثال // في البيانات التالية اوزان عينة من طلبة قوامها عشرة طلاب يطلب حساب قيمة الانحراف المعياري .

56 , 56 , 68 , 72 , 63 , 65 , 68 , 71 , 69 , 62

الحل // ∵ البيانات غير مبوبة

∴ نكتب الصيغة الرياضية للانحراف المعياري في حالة البيانات غير مبوبة .

$$S = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}}$$

نحتاج الى استخراج قيمة $\bar{X}$ في حالة البيانات غير مبوبة .

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{56+56+68+72+63+65+68+71+69+62}{10}$$

$$\bar{X} = \frac{650}{10} = 65$$

$(x_i - \bar{x})^2$	$x_i - \bar{x}$	x_i
81	-9	56
81	-9	56
9	3	68
49	7	72
4	-2	63
0	0	65
9	3	68
36	6	71
16	4	69
9	-3	62
294		

$$\therefore S = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}}$$

$$S = \sqrt{\frac{294}{10}} = \sqrt{29.4}$$

$$S = 5.422$$

b) حساب الانحراف المعياري في حالة البيانات المبوبة

- يحسب بالصيغة الرياضية التالية:-

$$S=\sqrt{\frac{\sum f_i\,(x_i-\bar{x})^2}{\sum f_i}}$$

حيث ان :-

f_i : تمثل التكرار. $\boxed{n=\sum f_i}$

x_i : تمثل مركز الفئة.

$\bar{x}$: الوسط الحسابي في حالة البيانات المبوبة.

مثال // البيانات التالية تمثل توزيع مجموعة من اللاعبين لرياضة دفع الاثقال وعددهم 120 وفقاً لفئات اوزانهم. المطلوب حساب الانحراف المعياري

الفئات	التكرارات	مركز الفئة	$f_i x_i$	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$f_i(x_i-\bar{x})^2$
60-	25	65	1625	-13.17	173.46	4336.25
70-	46	75	3450	-3.17	10.05	462.3
80-	35	85	2975	6.83	46.65	1632.75
90-100	14	95	1330	16.83	283.25	3966.5
	$\sum f_i=120$		9380			10396.8

١. بما ان البيانات مبوبة :.

٢. يجب استخراج $\bar{x}$ ومركز الفئة x_i

$$\bar{x}=\frac{\sum f_i x_i}{\sum f_i}=\frac{9380}{120}$$

$$\bar{x}=78.17$$

$$S=\sqrt{\frac{\sum f_i\,(x_i-\bar{x})^2}{\sum f_i}}$$

$$=\sqrt{\frac{10396.8}{120}}=\sqrt{86.64}$$

$$\therefore S=9.31$$

مثال // لجدول التوزيع التكراري أوجد الانحراف المعياري .

الفئات (معطى)	التكرارات (معطى)	x_i (تم إيجاده)	$f_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$
15-	10	20	200	-19	361	3610
25-	8	30	240	-9	81	648
35-	5	40	200	1	1	5
45-	10	50	500	11	121	1210
55-65	7	60	420	21	441	3087
	$\Sigma f_i = 40 = n$		1560			8560

بما أن البيانات مبوبة $\therefore$ نحتاج حساب مركز الفئة x_i وكذلك $\bar{x}$ للحصول أخيراً على الانحراف المعياري

$$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1560}{40} = 39$$

$$S = \sqrt{\frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i}} = \sqrt{\frac{8560}{40}} = \sqrt{214}$$

$S = 14.63 =$ الانحراف المعياري

④ التباين Variance

يعرف التباين بانه متوسط مجموع مربعات انحرافات قيم x عن وسطها الحسابي.
وهذا يعني ان التباين ما هو الا مربع الانحراف المعياري لتلك المجموعة من القيم، ويرمز للتباين بالرمز S^2 الذي يشير الى تباين العينة (في حالة البيانات المبوبة وغير المبوبة).

مثال // اذا كان الانحراف المعياري لمجموعة قيم هو 4
$S=4$
فان التباين لتلك المجموعة هو $S^2=16$

مثال // للمثال السابق ص 43 أوجد التباين.

$\because$ الانحراف المعياري $S=5.422$

$\therefore$ التباين يساوي (مربع الانحراف المعياري)

$$\therefore S^2=(5.422)^2$$

$$=29.398$$

مثال // للمثال السابق ص 44 أوجد التباين

$S=9.31 \quad \therefore S^2=86.676$

$\therefore$ التباين هو مربع الانحراف المعياري

مخطط يلخص الصيغ الرياضية الخاصة بمقاييس التشتت المطلق

① المدى R

في حالة البيانات غير مبوبة

$$R = X_U - X_L$$

أكبر قيمة ، أصغر قيمة

في حالة البيانات المبوبة

$R = X_U$ أكبر فئة $- X_L$ أقل فئة

② الانحراف المتوسط $M.D$

في حالة البيانات غير مبوبة

$$M.D = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

x_i :

$\bar{x}$: الوسط الحسابي $\frac{\sum x_i}{n}$

n : حجم العينة

في حالة البيانات المبوبة

$$M.D = \frac{\sum_{i=1}^{n} f_i |x_i - \bar{x}|}{\sum_{i=1}^{n} f_i}$$

x_i : مركز الفئة

$\bar{x}$: الوسط الحسابي

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

f_i : التكرارات

③ الانحراف المعياري (S)

④ التباين (S^2)

في حالة البيانات غير المبوبة

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

x_i : القيم

$\bar{x}$: الوسط الحسابي

n : حجم العينة

في حالة البيانات المبوبة

$$S = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}}$$

x_i : مركز الفئة

$\bar{x}$: الوسط الحسابي

f_i : التكرارات

* (التباين هو مربع الانحراف المعياري)

عزيزاتي الطالبات، هذا الواجب البيتي سيساعدكم على فهم المادة بشكل أسهل. امنياتي لكم بالتوفيق

م.د ريم طلال كامل

اولاً // للبيانات التالية أحسبي مقاييس التشتت التالية :-
(المدى ، الانحراف المتوسط والانحراف المعياري والتباين)

10, 77, 55, 40, 35, 30, 25, 30, 22

ثانياً // أحسب مقاييس التشتت التالية (المدى، الانحراف المعياري، الانحراف المتوسط، التباين) لسبعة مباريات قام بها احد لاعبي كرة السلة مع فريق المنتخب الوطني .
5, 6, 3, 7, 8, 4, 2,

ثالثاً // لاعب كرة طائرة احرز النقاط التالية لخمس مباريات :
16, 19, 21, 15, 14
أوجد مقاييس التشتت (الانحراف المعياري والتباين)

رابعاً // احرز لاعب كرة قدم في خمس مباريات في الدوري الاهداف التالية

5, 1, 2, 3, 4

أوجد (الانحراف المتوسط والمدى والتباين)

خامساً // في الجدول التالي يوضح بيانات لتوزيع 20 طالبة حسب درجاتهم في مادة الإحصاء المطلوب حساب مقاييس التشتت (المدى ، الانحراف المعياري والانحراف المتوسط ، التباين)

الدرجات (الفئات)	10 -	20 -	30 -	40 - 50
عدد الطالبات f	2	6	8	4

سادساً // احسب مقاييس التشتت للتوزيع التكراري

الفئات	5 -	15 -	25 -	35 - 45
التكرارات	3	5	8	4

سابعاً // للبيانات التالية أوجد التباين والانحراف المعياري

الفئات	التكرارات	مركز الفئة
3 - 5	5	4
6 - 8	12	7
9 - 11	22	10
12 - 14	7	13
15 - 17	4	16
	50	

ثانياً // مقاييس التشتت النسبي

كانت مقاييس التشتت المطلق تقيس درجة تباعد او اقتراب القيم بوحدات قياس هذه القيم. لكن في الاحيان كثيرة يتطلب مقارنة تشتت مجموعتين من البيانات المختلفة في وحدات قياسها او مقارنة صفتين متطلبين بنفس المجموعة كالطول والوزن لعدد من طلاب كلية التربية البدنية وعلوم الرياضة.

وعلى هذا الاساس لا يجوز استخدام مقاييس التشتت المطلقة وانما نحتاج الى مقاييس تشتت اخرى تزيل اثر الاختلاف في وحدات القياس بين مجموعتين او الصفتين. اي تحويل المقياس المطلق الى مقياس نسبي وذلك عن طريق تنسيب مقياس التشتت الى احد مقاييس النزعة المركزية والذي يتلائم مع مقياس التشتت المستعمل.

1. معامل الاختلاف C.V

Coffìcient of variation

ويعرفا على انه مقياس تشتت نسبي لمعرفة التشتت داخل المجموعة الواحدة او بين المجموعات، ومايميزه انه لا يتطلب توحيد وحدات القياس لانه يعطي التشتت نسبياً.

* يستخدم عندما تختلف المتوسطات الحسابية اي بمعنى لو كانت هناك مجموعة من البيانات لها وسط حسابي يساوي 8 ومجموعة اخرى لها وسط حسابي فيساوي 6.
وفي حالة المتوسطات الحسابية المتساوية يمكننا المقارنة بالاعتماد على الانحراف المعياري. دون اللجوء الى استخراج معامل الاختلاف.

والصيغة الرياضية لمعامل الاختلاف تكتب بالشكل الآتي :-

$$C.V = \left(\frac{S}{\bar{X}}\right) 100\%$$

يمثل كل من :-
S :- الانحراف المعياري .
$\bar{X}$:- الوسط الحسابي .

ملاحظة // قيمة معامل الاختلاف (%)، وكلما اقترب معامل الاختلاف من (1%) يعد التجانس عاليا" (اقل تشتتا") في حين اذا زاد عن (30%) تعد العينة غير متجانسة)

مثال // متوسط اعمار طلاب كلية التربية البدنية وعلوم الرياضة (20) سنة وانحرافهم المعياري (4) ومتوسط أعمار طلاب كلية القانون (20) سنة وانحرافهم المعياري (3).
المطلوب // اي من المجموعتين اكثر تجانسا" (أقل تشتتا")؟

- الحل // كون المتوسطات الحسابية متساوية عليه نعتمد على الانحراف المعياري دون الحاجة لتطبيق معامل الاختلاف وهنا اعمار طلاب كلية القانون (المجموعة 2) اكثر تجانسا" (أقل تشتتا") من المجموعة الاولى . وذلك لكون كلية القانون تمتلك انحرافا" معياريا" أقل من كلية التربية البدنية وعلوم الرياضة .

مثال // عينة استخدمت في احد البحوث كان متوسطها الحسابي (22) سنة وانحرافها المعياري (5) المطلوب . ايجاد معامل الاختلاف لمعرفة اذا كانت العينة متجانسة ام غير متجانسة ؟
الحل //

$$C.V = \frac{S}{\bar{X}} . 100\%$$

$$= \frac{5}{22} . 100\% = 0.227 * 100\% = 22.7\%$$

∴ نعم العينة متجانسة
لكون معامل الاختلاف اقل من 30% .

مثال // ثلاث مجموعات أوساطها الحسابية على التوالي (50 , 46 , 40) وانحرافها المعياري على التوالي (4.4 , 4.2 , 4) المطلوب // ايجاد معامل الاختلاف للمجموعات الثلاثة مع بيان من المجاميع الاكثر تجانسها.

الحل // نجد معامل الاختلاف لكل مجموعة على حدة ونقرر أيهم أكثر تجانسا (أقل تشتتا).

1) المجموعة الاولى $C.V = \frac{S}{\bar{X}} . 100\% = \frac{4.4}{50} . 100\%$

$C.V = 0.088 * 100\% = 8.8\%$

2) $C.V = \frac{S}{\bar{X}} . 100\% = \frac{4.2}{46} . 100\% = 9.13\%$

3) $C.V = \frac{S}{\bar{X}} . 100\% = \frac{4}{40} . 100\% = 10\%$

- المجموعات الثلاث متجانسة اذ ان المجموعة الاولى اقل تشتتا واكثر تجانسا تليها المجموعة الثانية ثم الثالثة.

مثال // الوسط الحسابي للمجموعة الاولى (70 كغم) وانحرافها المعياري (10) والوسط الحسابي للمجموعة الثانية (154.7) باون وانحرافها المعياري (22.1)
المطلوب // اي المجموعتين اكثر تجانسا اقل تشتتا ؟

* عندما يطلب اي المجموعتين اكثر تجانسا او اي المجموعتين اقل تشتتا يعني المطلوب هو معامل الاختلاف. حتى ولو لم يذكر ذلك بالسؤال.

1) $C.V = \frac{S}{\bar{X}} . 100\% = \frac{10}{70} \times 100\% = 0.142 \times 100\%$
$= \boxed{14.2\%}$

2) $C.V = \frac{S}{\bar{x}} . 100\% = \frac{22.1}{154.7} . 100\% = 0.142\%$
$\boxed{= 14.2\%}$

* المجموعتين متجانستين ولهم نفس التشتت وذلك لانهما يمتلكان نفس معامل الاختلاف .

س// احسب معامل الاختلاف للبيانات الآتية :- H.W
(1, 4, 10, 8, 7, 5, 2, 0, 6, 3) وهل هناك تشتت ؟

س// احسب معامل الاختلاف للبيانات الآتية :- H.W
وهل المجموعة متجانسة ؟
(2, 5, 8, 16, 10, 6, 3)

مثال// ادناه درجات مجموعة من لاعبي كرة السلة وأراد مدرب الفريق التعرف على نسبة التشتت بمهارة التهديف لدى الفريق . (اي هل هناك تجانس للعينة)

$X_i =$ 5, 4, 4, 2, 1, 3, 4, 5, 4, 3

خطوات الحل //

1. ايجاد الوسط الحسابي (هنا البيانات غير مبوبة)

$$\bar{X} = \frac{\sum_{i=1} x_i}{n} = \frac{5+4+4+\dots+3}{10} = \frac{35}{10}$$

$$\therefore \bar{X} = 3.5$$

2) إيجاد الانحراف المعياري (في حالة البيانات غير مبوبة)

$\bar{x} = 3.5 \quad n = 10$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
5	5−3.5=1.5	2.25
4	0.5	0.25
4	0.5	0.25
2	−1.5	2.25
1	−2.5	6.25
3	−0.5	0.25
4	0.5	0.25
5	1.5	2.25
4	0.5	0.25
3	0.5	0.25
		14.5

$\therefore S = \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n}}$

$\therefore S = \sqrt{\dfrac{14.5}{10}}$

$S = 1.2$

بما أن تم استخراج الوسط الحسابي والانحراف المعياري نستطيع الآن استخراج معامل الاختلاف (C.V)

$C.V = \dfrac{S}{\bar{x}} \times 100\% = \dfrac{1.2}{3.5} \cdot 100\% = 0.343 \times 100\%$

معامل الاختلاف $C.V = 34.3\%$

عند مقارنة ناتج معامل الاختلاف بـ 30% وهي المعيار أو النسبة المقارن بها نجد أن قيمة C.V المحسوبة أكبر من 30% وعليه نستطيع ان نقول هنا العينة غير متجانسة.

2. الدرجة المعيارية standard score

في كثير من الاحيان قد نحتاج الى مقارنة قيمتين مختلفتين تنتمي كل منهما الى مجموعة معينة، ولما كانت طبيعة البيانات في اي مجموعة احصائية يتم وصفها عن طريق معرفة كل من متوسطها وتشتتها كان من المناسب الاستعانة بهذين البعدين في عملية المقارنة. اي يجب الا تتم المقارنة بين القيم بشكل مطلق وانما من خلال تحويل القيم المراد مقارنتها الى وحدات معيارية خالية التأثير من الوسط الحسابي والتشتت ومن وحدات قياس المتغير الاصلي وعلى هذا الاساس تعتبر الدرجة المعيارية من أحد مقاييس التشتت النسبي والتي تعرف على النحو الآتي :-

$$Z = \frac{X_i - \bar{X}}{S}$$

حيث إن :-

$\bar{X}$: الوسط الحسابي.

S : الانحراف المعياري.

X_i : القيم (مفردات العينة).

مثال// اذا كانت تقديرات أحد العاملين في مصنع ما على مقياسين أولهما لمستوى الانتاج وثانيهما لمستوى الرضا عن العمل. علماً بأن المقياسين طبقا على مجموعة من العاملين في ذلك المصنع. ووجد من ذلك أن الوسط الحسابي والانحراف المعياري في كل حالة كانت على النحو الآتي :-

	مقياس الانتاج	مقياس الرضا عن العمل
تقديرات العامل X_i	42	80
متوسط التقديرات $\bar{X}$	35	60
الانحراف المعياري S	5	12

ففي أي حالة كان وضع هذا العامل أفضل ؟

الحل // يتم احتساب الدرجة المعيارية لهذا العامل لمقياسين الانتاج والرضا عن العمل وذلك من خلال تطبيق الصيغة الرياضية للدرجة المعيارية وكما يلي :-

(1) الدرجة المعيارية لتقدير العامل على مقياس الانتاج هي :-

$$Z_1 = \frac{X_i - \bar{X}}{S} = \frac{42 - 35}{5} = 1.2$$

(2) الدرجة المعيارية لتقدير العامل على مقياس الرضا عن العمل فهي :-

$$Z_2 = \frac{X_i - \bar{X}}{S} = \frac{80 - 60}{12} = 1.67$$

ومن مقارنة قيمتي الدرجات المعيارية للمقياسين يتضح أن Z_2 اكبر من Z_1 وهذا يعني ان وضع هذا

العامل على مقياس الرضا عن العمل أفضل حالاً من وضعه على مقياس الانتاج بالنسبة لجميع العاملين.

* تفسير النتائج مهم جداً جداً.

مثال // اذا كانت أهداف أحد اللاعبين في فريق كرة القدم كانت مقدرة على مقياسين أولهما مستوى التدريب والثاني مستوى اللياقة البدنية لدى اللاعب. علماً ان المقياسين طبقا على بقية اللاعبين في الفريق. ووجد في ذلك ان الوسط الحسابي والانحراف المعياري في كل حالة كانت على النحو التالي :-

	مستوى التدريب	اللياقة البدنية
تقديرات اللاعب	50	98
الوسط الحسابي للاعب	22	25
الانحراف المعياري للاعب	9	13

* في اي حالة كان وضع اللاعب أفضل ؟ يعني ما العامل المؤثر ؟

الحل // يتم احتساب الدرجة المعيارية للاعب طبقاً لمستوى التدريب وكذلك مستوى اللياقة البدنية .

(1) الدرجة المعيارية لتقدير اللاعب على مقياس التدريب هي

$$Z_1 = \frac{X_i - \bar{X}}{S} = \frac{50-22}{9} = \underline{\underline{3.111}}$$

(2) الدرجة المعيارية للاعب لمقياس مستوى اللياقة البدنية هي :-

$$Z_2 = \frac{X_i - \bar{X}}{S} = \frac{98-25}{13} = \underline{\underline{5.615}}$$

ويظهر من مقارنة قيمتي الدرجة المعيارية للمقياسين أن Z_2 أكبر من Z_1 وهذا يعني بأن وضع

هذا اللاعب على مقياس اللياقة البدنية أفضل مما لا هو وضعه على مقياس مستوى التدريب .

سؤال // احسب الدرجة المعيارية لقيم الدرجات التي حصل عليها 8 لاعبين في مباراة كرة القدم . H.W

الدرجات : 66 92 60 48 68 90 81 56

الفصل الرابع: الارتباط Correlation

ويقسم الى نوعين :-
١. الارتباط الخطي (علاقة خطية).
٢. الارتباط غير الخطي (علاقة غير خطية).

أولاً // الارتباط الخطي Linear Correlation

- الارتباط يوضح مامدى العلاقة الموجودة بين المتغيرين X و Y ويرمز له بالرمز (r_{xy}) مثلاً هناك علاقة بين القلق النفسي والاداء لدى اللاعبين في المباراة. أو هل هناك علاقة بين طول ووزن الشخص او هل هناك علاقة بين مستوى ذكاء الطالب وتحصيله الدراسي ؟

- الارتباط بين ظاهرتين معناه وجود علاقة بينهما. بحيث اذا تغير احدهما في اتجاه معين مال الآخر الى التغير في نفس الاتجاه او في الاتجاه المعاكس. اذ يسمى الارتباط في الحالة الاولى طردياً او موجباً مثال على ذلك العلاقة بين زيادة ساعات التدريب فتساعد على زيادة نسبة الفوز او تحسن اداء اللاعب. اما الاتجاه الثاني (الارتباط المعاكس) مثلاً زيادة استهلاك الطاقة الكهربائية للفرد تقلل من فرصة حصول الافراد الآخرين على الطاقة الكهربائية.

صفات معامل الارتباط :-

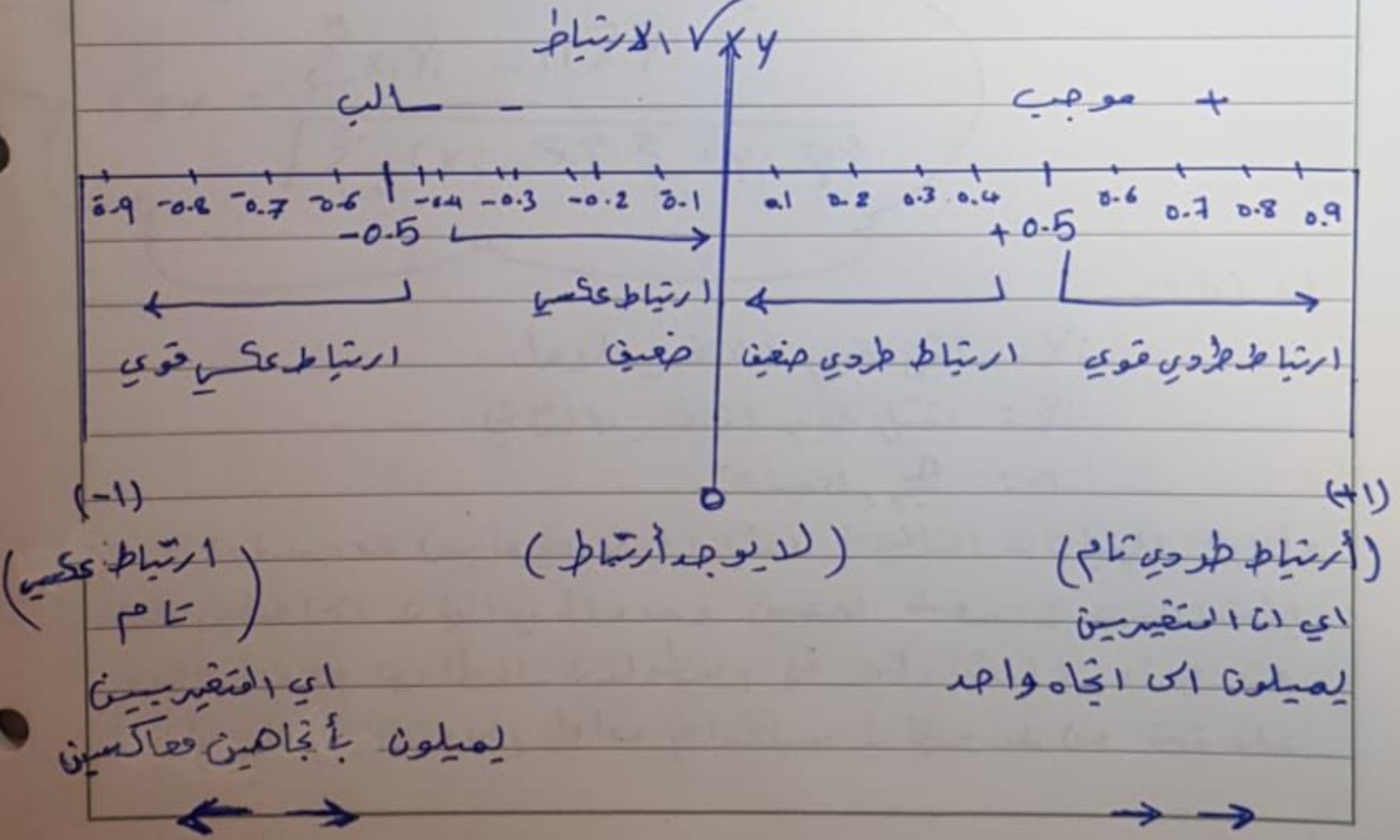
١. قيمة معامل الارتباط لا تزيد على الواحد ولا تقل عن سالب واحد.

يعني ⇐ $-1 \leq r_{xy} \leq +1$

2) كلما اقتربت قيمة معامل الارتباط من (+1) او (-1) كلما كان الارتباط قوياً وبالعكس كلما اقتربت قيمة معامل الارتباط من الصفر كلما انعدم الارتباط.

3) اذا كانت قيمة معامل الارتباط موجب (+) يعني هذا وجود علاقة ارتباط موجبة اي (طردية) اي ان المتغيرين X و Y يسيران في اتجاه واحد.
وبالعكس اي اذا كانت قيمة معامل الارتباط سالبة (-) ذلك يدل على وجود علاقة ارتباط سالبة (عكسية) اي ان المتغيرين X , Y يسيران في اتجاهين متعاكسين.

4) اذا كان معامل الارتباط مساوي الى (+1) فان هناك علاقة دالة طردية تامة بين المتغيرين وكذلك اذا كان الارتباط (-1) فالعلاقة دالة عكسية بين المتغيرين. اما اذا كان معامل الارتباط يساوي (صفر) فمعنى ذلك انعدام الارتباط. اي لا يوجد علاقة بين المتغيرين X و Y.



هناك انواع للارتباط بحسب عدد المتغيرات

1. الارتباط الخطي البسيط Simple correlation
2. الارتباط الجزئي partial Correlation
3. الارتباط المتعدد Multiple Correlation

* سندرس في هذا الفصل النوع الاول للارتباط فقط *

1. الارتباط البسيط Simple Correlation

أولاً :-

معامل ارتباط بيرسون Pearson Correlation Coefficient

يعرف معامل الارتباط الخطي البسيط بانه القيمة العددية للعلاقة الخطية بين متغيرين ، هذا النوع من الارتباط يستعمل عندما تكون المتغيرات كمية. وكما في الصيغة الآتية :-

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

إذ ان :-

x_i : تمثل قيم المتغير الاول.

y_i : تمثل قيم المتغير الثاني.

n : حجم العينة.

مثال // البيانات التالية تمثل العلاقة ما بين عدد ساعات التدريب لتسعة لاعبين وعدد الميداليات الحاصلين عليها من جراء المشاركة في البطولات. المطلوب معرفة وجود علاقة من عدمها باستخدام معامل الارتباط المناسب.

اي هل يوجد ارتباط ام لا ؟

عدد ساعات التدريب X_i : 3 5 7 8 9 11 6 8 6

عدد المداليات Y_i : 2 2 5 4 5 6 3 5 4

الحل //

∵ البيانات كمية ∴ نستخدم معامل ارتباط بيرسون

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

يجب استخراج قيمة الوسط الحسابي للمتغيرين $\bar{x}$, $\bar{y}$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{3+5+7+8+9+\cdots+6}{9} = \frac{63}{9} = 7$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{2+2+5+\cdots+4}{9} = \frac{36}{9} = 4$$

x_i	y_i	$x_i y_i$	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$
3	2	6	-4	16	-2	4
5	2	10	-2	4	-2	4
7	5	35	0	0	1	1
8	4	32	1	1	0	0
9	5	45	2	4	1	1
11	6	66	4	16	2	4
6	3	18	-1	1	-1	1
8	5	40	1	1	1	1
6	4	24	-1	1	0	0
		276		44		16

نعوض في الصيغة الرياضية :-

$$r_{xy} = \frac{276 - 9(7)(4)}{\sqrt{(44)(16)}} = \frac{276 - 252}{\sqrt{704}} = \frac{24}{26.53}$$

$r_{xy} = 0.905$ أي ان الارتباط قوي

أو نقول هناك علاقة طردية قوية بين عدد ساعات التدريب
و عدد الميداليات.

مثال // أراد أحد المدربين في كرة القدم معرفة وجود
علاقة ارتباط بين الوزن والطول لمجموعة من اللاعبين
وعددهم (5) لاعبين ، المطلوب أستخدام معامل
الارتباط المناسب لايجاد العلاقة لمعرفة هل يوجد ارتباط
ام لا.

الطول X_i : 183 , 175 , 156 , 166 , 167

الوزن Y_i : 80 , 70 , 65 , 64 , 63

الحل // نحتاج الى ايجاد الوسط الحسابي $\bar{X}$, $\bar{Y}$

لان البيانات كمية ∴ نستخدم معامل الارتباط بيرسون.
والتي تكتب صيغته الرياضية بالشكل الآتي :-

$$r_{xy} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{X}\bar{Y}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{X})^2 \sum_{i=1}^{n}(y_i - \bar{Y})^2}}$$

$$\bar{X} = \frac{\sum x_i}{n} = \frac{183 + 175 + \cdots + 167}{5} = \frac{847}{5} = 169.4$$

$$\bar{Y} = \frac{\sum y_i}{n} = \frac{80 + 70 + 65 + 64 + 63}{5} = \frac{342}{5} = 68.4$$

X_i	Y_i	X_iY_i	$X_i-\bar{X}$	$(X_i-\bar{X})^2$	$Y_i-\bar{Y}$	$(Y_i-\bar{Y})^2$
183	80	14640	13.6	184.96	11.6	134.56
175	70	12250	5.6	31.36	1.6	2.56
156	65	10140	-13.4	179.56	-3.4	11.56
166	64	10624	-3.4	11.56	-4.4	19.36
167	63	10521	-2.4	5.76	-5.4	29.16
		58175		413.2		197.2

$$\therefore r_{xy} = \frac{58175 - 5(169.4)(68.4)}{\sqrt{(413.2)(197.2)}}$$

$$r_{xy} = \frac{58175 - 57934.8}{\sqrt{81483.04}} = \frac{240.2}{285.45} = 0.841$$

- اي ان هنالك ارتباط طردي قوي بين الوزن والطول.

H.W سؤال// ادناه درجات اختبار اللياقة البدنية لمجموعتين من الرياضيين المبتدئين و المتمرسين كالاتي :-

المبتدئين X_i : 44 39 36 35 33 29 22

المتمرسين Y_i :- 58 42 18 22 31 38 38

المطلوب ايجاد معامل الارتباط بين المجموعتين باستخدام معامل الارتباط المناسب.

(63)

ثانياً // معامل ارتباط الرتب لسبيرمان

Spearman's Rank Correlation Coefficient

يستخدم هذا النوع من الارتباط عندما تكون المتغيرات وصفية وليست كمية وحسب الصيغة الرياضية التالية:-

$$r_{xy} = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

إذ ان :-

d_i ماذا يمثل ؟ يمثل الفرق بين رتبتي المفردة نفسها في المتغيرين اي ان :-

$$d_i = X_i - Y_i$$

اي ان

$$d_i = \text{Rank } X - \text{Rank } Y$$

Rank : رتبة

حيث يكون

$$\boxed{\sum_{i=1}^{n} d_i = 0}$$

n : تمثل عدد المفردات (عدد الازواج المرتبة)

مثال // تقديرات ستة طلاب في مادتي الاحصاء والرياضيات كانت :-

درجة الاحصاء X_i : جيد ، متوسط ، ضعيف ، مقبول ، جيد جداً ، ممتاز

درجة الرياضيات Y_i : متوسط ، جيد ، مقبول ، ضعيف ، ممتاز ، جيد جداً

المطلوب // جد معامل الارتباط بين تقدير الطالب في امتحان الاحصاء وتقديره في امتحان الرياضيات ، استخدم معامل الارتباط المناسب لايجاد هذه العلاقة.

الحل //

1) بما ان المتغيرات وصفية ∴ تستخدم معامل الارتباط سبيرمان.

2) نبدأ بترتيب التقديرات وفق ترتيب تصاعدي ، ثم نخصص رتبة تمثل كل تقدير.

ضعيف	مقبول	متوسط	جيد	جيد جداً	ممتاز
1	2	3	4	5	6

ثم نعود لتخصيص هذه الرتب للتقديرات الاصلية كل حسب موقعها وكما موضح بالجدول الآتي :-

درجة الاحصاء X_i	درجة الرياضيات Y_i	رتبة x_i	رتب y_i	رتب $d_i = x_i - y_i$	d_i^2
جيد	متوسط	4	3	1	1
متوسط	جيد	3	4	-1	1
ضعيف	مقبول	1	2	-1	1
مقبول	ضعيف	2	1	1	1
جيد جداً	ممتاز	5	6	-1	1
ممتاز	جيد جداً	6	5	1	1
					6

$$\therefore \; r_{xy} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6(6)}{6((6)^2-1)}$$

$$r_{xy} = 1 - \frac{6}{35} = 0.829$$

اي ان الارتباط طردي قوي
اي ان هناك علاقة قوية
بين مادة الاحصاء ومادة الرياضيات.

مثال // في أحد الابحاث الحديثة أراد الباحث لمعرفة فيما لو كان هناك علاقة بين الحالة المادية للطالب وتقديرات الطالب في الامتحانات لمجموعة من الطلاب عددهم (5) ما المطلوب // ما هي الطريقة المناسبة لمعرفة وجود او عدم وجود علاقة بينهما ؟

الحالة المادية X_i : جيد ، ممتاز ، متوسط ، جيد جداً ، ضعيف

تقديرات الطالب Y_i : متوسط ، جيد جداً ، ممتاز ، ضعيف ، جيد

الحل // :- البيانات وصفية :. نستخدم معامل الارتباط سبيرمان

$$r_{xy} = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$$

1) نرتب التقديرات المعطاة بالسؤال ترتيباً تصاعدياً (2) ثم نخصص رتباً لكل تقدير.

ضعيف	متوسط	جيد	جيد جداً	ممتاز
5	4	3	2	1

الحالة المادية X_i	تقدير الطالب Y_i	رتب X_i	رتب Y_i	$d_i = X_i - Y_i$	d_i^2
جيد	متوسط	3	2	1	1
ممتاز	جيد جداً	5	4	1	1
متوسط	ممتاز	2	5	-3	9
جيد جداً	ضعيف	4	1	3	9
ضعيف	جيد	1	3	-2	4
					24

$$r_{xy} = 1 - \frac{6(24)}{5((5)^2-1)} = 1 - \frac{144}{120} = 1 - 0.85 = 0.15$$

وهذا يعني وجود ارتباط طردي ضعيف بين الحالة المادية وتقديرات الطالب.

H.w

سؤال // قيم مدربان خمسة لاعبين من لاعبي كرة السلة فأعطى كل منهما تقييماً للاعبين الخمسة وكانت نتائج التقييم للمدربين مبينة كالآتي :-

المدرب الاول x_i : مقبول ، ممتاز ، ~~ممتاز~~ جيد جداً ، جيد ، متوسط

المدرب الثاني y_i : متوسط ، جيد جداً ، جيد ، ممتاز ، مقبول

هل توجد علاقة بين تقييم المدرب الاول والمدرب الثاني للاعبين ؟ استخدم معامل الارتباط المناسب لايجاد العلاقة .

الارتباط

بيانات كمية (ارقام)

⇓

نستخدم

معامل الارتباط (بيرسون)

$$r_{xy} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum (x_i-\bar{x})^2 (y_i-\bar{y})^2}}$$

x_i : قيم عددية (ارقام)

y_i : قيم عددية (ارقام)

بيانات وصفية (نوعية)

⇓

نستخدم

معامل الارتباط سبيرمان

$$r_{xy} = 1 - \frac{6\sum_{i=1} d_i^2}{n(n^2-1)}$$

x_i : متغيرات وصفية .

y_i : وصفية .

بعض الملاحظات حول معامل الارتباط :-

1. يتأثر معامل الارتباط بحجم العينة فكلما زاد حجم العينة كان إحتمال الحصول على معامل إرتباط قوي وكلما قل حجم العينة كلما كان إحتمال الحصول على معامل إرتباط ضعيف.
2. من أهم فوائد معامل الارتباط هو التنبؤ وتعتمد دقة التنبؤ على درجة معامل الارتباط بين المتغيرين اي كلما كانت العلاقة (الارتباط) قوية بين المتغيرين زادت دقة التنبؤ والعكس صحيح.

مثال // توفرت المعلومات في أدناه والتي تخص درجة تقييم الاداء (X) ومستوى الحالة الصحية (Y) لسبعة من العاملين في إحدى الشركات الصناعية :-

العامل	الاول	الثاني	الثالث	الرابع	الخامس	السادس	السابع
تقييم الاداء X_i	93	57	88	75	40	62	80
الحالة الصحية Y_i	ممتازة	وسط	جيدة	جيدة جداً	رديئة جداً	دون الوسط	فوق الوسط

المطلوب أحسب معامل الارتباط المناسب لمعرفة فيما لو كان هناك علاقة بين درجة تقييم الاداء والحالة الصحية للعاملين في الشركة.

الجواب //

تقييم الاداء X_i	الحالة الصحية Y_i	رتب X_i	رتب Y_i	$d_i = X_i - Y_i$	d_i^2
93	ممتازة	6	6	0	0
57	وسط	2	3	-1	1
88	جيدة	5	4	1	1
75	جيدة جداً	4	5	-1	1
40	رديئة جداً	1	1	0	0
62	دون الوسط	3	2	-1	1
80	فوق الوسط	5	4	-1	1
					5

تم ترتيب الجدول حسب التعريف لكل تقدير بما يساويها

اي ان :-

ضعيف ، مقبول ، متوسط ، جيد ، جيد جداً ، ممتازة

1 ، 2 ، 3 ، 4 ، 5 ، 6

بما ان السؤال يتضمن تفاصيل للتقدير المشروحة

هـ.

	الرتب Y	
ممتازة	6	
وسط	3	
جيدة	4	
جيدة جداً	5	
رديئة جداً	1	(يعني ضعيف)
دون الوسط	2	(يعني مقبول)
فوق الوسط	4	(يعني جيدة)

		الرتب X
93	ممتازة	6
57	مقبول	2
88	جيدة جداً	5
75	جيدة	4
40	ضعيف	1
62	متوسط	3
80	جيد جداً	5

$$\therefore \quad r_{xy} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6(5)}{7((7)^2-1)} = 1 - \frac{30}{7(48)}$$

$$r_{xy} = 1 - \frac{30}{336} = 0.911$$

اي هناك علاقة قوية وطردية

بين الحالة الصحية والاداء للعاملين في الشركة.

Test's of Hypotheses
اختبار الفرضيات

ما المقصود بالفرضية ؟ :-
وهي تخمين أو إستنتاج ذكي مبني على أسباب معقولة أو منطقية ولكنه ليس مبنياً على حسابات دقيقة خاصة بالمجتمع لاننا لانفترض انه لايمكن دراسة المجتمع بالكامل عن طريق الحصر الشامل بل نحاول الإستنتاج أو الاستدلال على مقاييس المجتمع باستعمال بيانات ونتائج العينة .

مثال على ذلك //
قد يفترض الباحث ان متوسط عدد الميداليات الحاصل عليها فريق العاب القوى في دولة ما هو 10 ، ويحتاج الى إختبار علمي (إحصائي) لمعرفة مدى صحة هذه الفرضية .
وذلك بوصول الباحث الى اتخاذ القرار المناسب عن طريق قبول الفرضية او رفض الفرضية وذلك باحتمال معين .

- أهم المفاهيم الاساسية لهذا الفصل هي :-

١. فرضية العدم (او الفرضية الصفرية) Null Hypotheses

وهي فرضية محايدة ويرمز لها بالرمز (H_0) وتعرف على انها تلك الفرضية التي يتم اختبار إمكانية رفضها بفرض أنها صحيحة وعادة يتم صياغتها بشكل يبين فيه بأن التقديرات للمعلمات التي يتم الحصول عليها من العينة مثلاً تكون مطابقة للمجتمع المسحوبة منه تلك العينة .

أي اذا كانت فرضية العدم المراد اختبارها هي ان متوسط عدد ميداليات الحاصل عليها فريق العاب القوى في دولة ما هو 10 ميداليات فان هذا الفرض يكتب بالصيغة التالية : $H_0 : \mu = 10$

2. الفرضية البديلة Alternative Hypotheses

وهي فرضية أخرى محتملة لفرضية العدم ويرمز لها بالرمز (H_1) وعادةً يتم صياغتها بشكل يختلف عن فرضية العدم ويكون مكمل له وتبين فيه بأن تقديرات المعلمات التي يتم الحصول عليها من العينة مثلاً لا تكون مطابقة للمجتمع (قد يكون عدم مطابقة فقط او أصغر او أكبر منه).

او بمعنى آخر هي الفرضية التي ستقبل في حالة رفض فرضية العدم اذ يمكن صياغتها على ثلاث حالات حسب متطلب السؤال فعلى سبيل المثال //
لو أخذنا المثال السابق

1. الاختبار من جانبين ($\neq$) وتكون صياغة الفرضية البديلة بالشكل الآتي :-

$$H_1: \mu \neq 10$$

2. الاختبار من جانب واحد ($>$) الكبر من الجانب الايمن.

$$H_1: \cancel{\mu} > 10$$

3. الاختبار من جانب واحد ($<$) اصغر من الجانب الايسر

$$H_1: \mu < 10$$

∴ ان أهمية الفرضية البديلة تبرز في تحديدها بالشكل الصحيح وذلك لان تحديد مستوى المعنوية الذي سيعتمد في الاختبار يعتمد "ايضاً" على نوع الفرضية البديلة.

3- مستوى المعنوية (الدلالة الإحصائية)
Level of Significance

هي قيمة احتمال رفض فرضية العدم H_0 عندما تكون صحيحة ويرمز عادةً لمستوى المعنوية بالرمز (α).

وتحدد قيمة (α) مسبقاً وقبل البدء بالاختبار وعادةً ما يتم اختيار (α) لأن تكون مساوية الى 0.05 أو 0.01 علماً ان كلما قلت قيمة α المختارة فذلك يعني ان احتمال اتخاذ قرار خاطئ يقل وان احتمال اتخاذ قرار صحيح يكبر.

فضلاً عن ذلك مستوى المعنوية يتأثر بنوع الاختبار (المطلوب في السؤال). اي اذا كان الاختبار من جانبين ($\neq$) فان قيمة (α) تبقى كما هي، اما اذا كان الاختبار من جانبين ($\neq$) فان قيمة مستوى المعنوية (α) تقسم على (2) اي تصبح ($\frac{\alpha}{2}$)

4. درجات الحرية Degrees of Freedom

وتمثل عدد مفردات العينة (حجم العينة) مطروحاً منه عدد القيود المستقلة التي فرضت على تلك العينة.

أي :- $d.f = n_1 + n_2 - 2$

خطوات الاختبار الاحصائي

① نحدد معطيات السؤال.

② صياغة فرضية العدم H_0 وكذلك الفرضية البديلة H_1 وبالشكل التالي :-

عادةً "تأخذ شكل المساواة" ⟸ $H_0: \mu_1 - \mu_2 =$ (رقم معين) العدم

وتأخذ واحدة من الاشكال الثلاثة حسب معطيات السؤال:

البديلة
- إما $H_1: \mu_1 - \mu_2 \neq$ (رقم معين)
- أو $H_1: \mu_1 - \mu_2 >$ (رقم معين)
- أو $H_1: \mu_1 - \mu_2 <$ (رقم معين)

③ احصاءة الاختبار (معيار الاختبار) Statistic Test

نقوم بحساب معيار الاختبار (t Test t) مثلاً: أي t المحسوبة

④ نستخرج قيمة معيار الاختبار (t الجدولية) من جداول احصائية عند مستوى معنوية محددة (معطاة بالسؤال) ودرجة حرية معينة.

⑤ المقارنة والقرار: بمعنى نقارن قيمة (t المحسوبة) مع t الجدولية

① اذا كانت (t المحسوبة) أصغر من (t الجدولية) نقبل فرضية العدم H_0 ونرفض الفرضية البديلة H_1.

② اذا كانت قيمة (t المحسوبة) اكبر من (t الجدولية) نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1.

بالمختصر المفيد:
- اذا t المحسوبة $>$ t الجدولية
  القرار // نرفض H_0 نقبل H_1
- اذا t المحسوبة $<$ t الجدولية
  القرار // نقبل H_0 نرفض H_1

إختبار الفرق بين وسطين حسابيين في لعينتين مستقلتين

عندما نفترض ان تباين المجتمعين متساويين علماً بأنهما غير معلومتين ، اي نفترض انهما متجانسين فان معيار الاختبار الملائم لهذه الحالة هو :-

$$t_{\text{المحسوبة}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

اذ إن :-

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

واذ نقارن قيمة معيار الاختبار (t المحسوبة) مع قيمة معيار الاختبار الجدولية (t الجدولية) عند مستوى معنوية محددة (α او $\frac{\alpha}{2}$) ودرجة حرية $(n_1 + n_2 - 2)$

مثال // لدراسة معينة لبيان الفروق بين متوسطي أوزان الذكور والاناث في احد المراحل الدراسية في كلية التربية البدنية وعلوم الرياضة تم اختيار (11) فرد من الذكور عشوائياً و (18) فرد من الاناث وتم الحصول على البيانات التالية :-

$$\bar{X}_1 = 82 \quad , \quad S_1^2 = 10$$
$$\bar{X}_2 = 73 \quad , \quad S_2 = 7$$

هل تتفق هذه النتائج مع الفرض القائل ان متوسطي أوزان الذكور والاناث في هذه المرحلة الدراسية مختلفة عند مستوى معنوية (0.05) علماً ان القيمة

$t(\alpha/2, n_1+n_2-2) = t(0.025, 27) = (2.052)$ الجدولية

$$H_0: \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 - \mu_2 \neq 0$$

خطوات الحل //

① نكتب معطيات السؤال

$n_1 = 11 \quad n_2 = 18 \quad \bar{X}_1 = 82 \quad S_1^2 = 10$

$\bar{X}_2 = 73 \quad S_2^2 = 7$

$\left(t_{\text{الجدولية}}\right) = 2.052$

② كتابة فرضية العدم H_0 والفرضية البديلة H_1

$$H_0: M_1 - M_2 = 0$$
$$H_1: M_1 - M_2 \neq 0$$

③ حساب معيار الاختبار (t المحسوبة)

$$(t_{\text{المحسوبة}}) = \frac{(\bar{X}_1 - \bar{X}_2) - (M_1 - M_2)}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$Sp = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} = \sqrt{\frac{(11-1)(10)+(18-1)(7)}{11+18-2}}$$

$$= \sqrt{\frac{219}{27}} = \sqrt{8.11} = 2.848$$

$$\therefore \underset{\text{المحسوبة}}{t} = \frac{(82-73)-0}{2.848\sqrt{\frac{1}{11}+\frac{1}{18}}} = \frac{9}{1.0899} = 8.258$$

④ استخراج (t الجدولية) معطاة بالسؤال وتساوي 2.052

⑤ المقارنة والقرار

نقارن قيمة $t_{\text{المحسوبة}} = 8.258$ مع قيمة $t_{\text{الجدولية}} = 2.052$

نلاحظ ان قيمة (t المحسوبة) اكبر من (t الجدولية)

∴ نرفض فرضية H_0 ونقبل الفرضية البديلة H_1

اي ان متوسط أوزان الذكور تختلف عن متوسط أوزان الاناث في مجتمع الدراسة.

* التفسير مهم جداً".

مثال // البيانات التالية تمثل نتائج عينتين عشوائيتين مستقلتين مسحوبتين من منتخبين لكرة القدم عن اعمار اللاعبين.

$n_1 = 10$, $n_2 = 10$, $\bar{X}_1 = 28$, $\bar{X}_2 = 26$

$S_1^2 = 50$, $S_2^2 = 30$

أختبر الفرضية التالية عند مستوى معنوية (0.05) علماً" ان القيمة الجدولية

$$t\left(\frac{\alpha}{2}, n_1+n_2-2\right) = t(0.025, 18) = 2.101$$

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

خطوات الحل :

① كتابة معطيات السؤال

$n_1 = 10 \quad , n_2 = 10 \qquad \bar{X}_1 = 28 \ , \ \bar{X}_2 = 26$

$S_1^2 = 50 \ , \ S_2^2 = 30$

$t_{\text{الجدولية}} = 2.101$

② صياغة فرضية العدم H_0 والفرضية البديلة

$H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 \neq 0$

③ حساب معيار الاختبار (t المحسوبة)

$$t_{\text{المحسوبة}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

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

$$S_p = \sqrt{\frac{(10-1)(50) + (10-1)(30)}{10 + 10 - 2}} = \sqrt{\frac{9(50) + 9(30)}{18}}$$

$$= \sqrt{\frac{450 + 270}{18}} = \sqrt{\frac{720}{18}} = \sqrt{40} = 6.325$$

$$t_{\text{المحسوبة}} = \frac{(28 - 26) - 0}{6.325 \sqrt{\frac{1}{10} + \frac{1}{10}}} = \frac{2}{6.325(0.447)} = 0.707$$

④ نستخرج قيمة (t الجدولية)

$t_{\text{الجدولية}} = 2.101$

⑤ المقارنة والقرار :-

نقارن قيمة $t_{\text{المحسوبة}} = 0.707$ مع $t_{\text{الجدولية}} = 2.101$

وقد تبين ان قيمة (t المحسوبة) أقل من (t الجدولية)

∴ نقبل فرضية العدم H_0 ونرفض الفرضية البديلة H_1

- أي ان متوسط اعمار اللاعبين في المنتخب الاول يساوي متوسط اعمار اللاعبين في المنتخب الثاني.

H.W مثال // البيانات التالية تمثل عينتين عشوائيتين مستقلتين مسحوبتين من منتخبين رياضيين لبلدين مختلفين

$S_1^2 = 12 \quad \bar{X}_1 = 95 \quad n_1 = 35$

$S_2^2 = 22 \quad \bar{X}_2 = 63 \quad n_2 = 23$

اختبر الفرضية القائلة بان متوسط حجم العينة الاولى اكبر من متوسط حجم العينة الثانية وعند مستوى معنوية $\alpha = 0.05$ علماً ان t الجدولية

$t_{\text{الجدولية}} (0.05, n_1 + n_2 - 2) = 3.456$

$t_{\text{الجدولية}} (0.05, 56)$

$H_0 : M_1 > M_2$

$H_1 : M_1 < M_2$