مبادئ علم الاحصاء

إعداد
نوار الاسدي

الأحصاء Statistics

المقدمة :

يحتل علم الأحصاء مكاناً مرموقاً بين العلوم لما له من استعمالات واسعة كأداة أو وسيلة للوصول الى قرارات صائبة لوصف أو تفسير الظواهر المختلفة في جميع العلوم إذ من خلال علم الأحصاء يمكن حل كثير من المشاكل منها الأدارية والأقتصادية والحياتية والطبية والهندسية وعلوم الزراعة والأجتماع وغيرها فهو تارة يستخدم للأشارة الى البيانات العددية عندما نتحدث عن الأحصاء السكاني أو الأحصاء الحيوي وتارة اخرى يستخدم للأشارة الى كافة الأختصاصات المتعلقة بالعلم نفسه فهو يهتم بدراسة نواتج التجارب العشوائية وهو يستعمل من قبل الأفراد والجماعات والدول على حد سواء وفي الحقيقة ان الأنتصار العظيم في نزول الأنسان على القمر ما كان ليحدث لولا مساعدة علم الأحصاء .

ان كلمة الأحصاء (statistics) مشتقة من عدة كلمات قديمة فهي مشتقة من الكلمة اللاتينية (status) والألمانية (statistik) والأيطالية (stato) وكلها تعني سياسة الدولة أو شؤون الدولة .

ان كلمة (statistics) استعملت لأول مرة في كتاب (Baran J.F. Von Bielfeld) وقد ضم هذا الكتاب فصلاً كاملاً عنوانه (statistics) وقد عرفه بأنه العلم الذي يدرس شؤون ونظم سياسة الدولة .

مفهوم الأحصاء The concept of statistics

هو العلم الذي يبحث في طريقة جمع وترتيب وتلخيص وعرض وتفسير وتحليل البيانات الخاصة بالظواهر المختلفة الخاضعة للقياس العددي والتي كثيراً ما تكون بشكل مشاهدات مختلفة يتسهل علم الاحصاء معرفة اتجاهاتها والعلاقات التي تربط بعضها ببعض ومعرفة القوانين والأنظمة التي تسيطر عليها وتتحكم فيها .

أصل علم الأحصاء :

ان أصل علم الأحصاء وتطوره ينبع من مصدرين رئيسيين وهما :

أ/ السجلات الحكومية (Governmental Records)

ب/ علم الرياضيات (Mathematics)

وظائف علم الأحصاء :

١- الوضوحية (Definiteness) : اي عرض الحقائق والبيانات بصورة واضحة ومحددة .

٢- التكثيف (Condensation) : اي تلخيص البيانات الكثيرة بقيم قليلة ذات معنى .

٣- المقارنة (Comparison) : اي وضع الأسس السليمة لمقارنة العوامل العائدة لنفس الظاهرة .

٤- صياغة واختبار الفرضيات (Formulating and testing of hypotheses)

وذلك لأن الطرق الأحصائية ذات فائدة عظيمة في صياغة واختبار الفرضيات وتطوير نظريات جديدة .

الأحصاء Statistics

٥- التنبؤ أو التكهن (Prediction) : حيث يساعد علم الأحصاء على التنبؤ أو التكهن بأتجاه وقيمة ظاهرة ما خلال فترة زمنية مستقبلية

٦- يساعد علم الأحصاء على وضع الخطط واتخاذ القرارات المناسبة من قبل مؤسسات الدولة لوضع السياسة المناسبة لقطاعاتهم المختلفة وذلك لأنه يوفر البيانات اللازمة للتخطيط ويحدد أتجاه وحجم التغير فيها.

مصطلحات أحصائية :-

الظاهرة : هي عبارة عن حادث عام طبيعي (حدث ما من الطبيعة) كالعلاقة بين المرض والتدخين.

العشوائية : هي نوع من الفوضى وعدم الأنتظام.

المجتمع (Population) : هو مجموعة البيانات التي تصف أو تميز ظاهرة معينة.

العينة (Sample) : هو مجموعة من البيانات التي يتم اختيارها من المجتمع ذات العلاقة لذا فالعينة مجموعة جزئية من المجتمع.

اشكال الأحصاء :

اولاً : الأحصاء الوصفي : وهو عبارة عن جمع معلومات معينة وتصنيفها في جداول خاصة وهذه المعلومات قد تكون نوعية وقد تكون كمية.

ثانياً : الأحصاء الأستدلالي : ويهتم بالوصول الى توابيت المجتمع الأحصائي.

* للاحصاء نوعان هما: الأحصاء الوسيطي والأحصاء غير الوسيطي

عرض البيانات :

هناك ثلاث طرق أساسية لعرض البيانات وهي :

١- طريقة الجداول ٢- طريقة العرض البياني ٣- طريقة حساب المقاييس الأحصائية المختلفة.

طريقة العرض الجدولي Tabular presentation

وفي هذه الطريقة تعرض البيانات الأحصائية بجداول ذات تقسيم واحد أو تقسيمات متقاطعة تعطي خلاصة البيانات وفق تلك التقسيمات وبشكل يسمح للقارئ ادراك ما تتضمنه البيانات من معاني وأبعاد واستنتاجات بسهولة ويسر.

مثال (١) : يبين الجدول الآتي عرضاً ملخصاً لدرجة أصابة (١٠٠) شجرة بآفة زراعية معينة

درجة الأصابة	عدد الأشجار (التكرار)
خفيفة	70
متوسطة	20
شديدة	10

من الجدول المجاور يستطيع القارئ أن يدرك بسهولة ان درجة الأصابة هي خفيفة في أكثر الأشجار وذلك لمجرد ألقاء نظرة على ملخص البيانات المعروضة فيه أما مصطلح : التكرار (Frequency) فأنه يعني عدد المرات التي تكررت فيها الصفة أو المجموعة قيد الدرس وعليه فأن الجدول المجاور يعتبر توزيع تكراري للأشجار المائة حسب درجة الأصابة.

مثال (٢): يبين الجدول الآتي ملخصاً للفئات العمرية لكادر فني تقني يعمل في منطقة إدارية معينة.

فئة العمر (بالسنوات)	عدد افراد الكادر الفني (التكرار)	النسبة المئوية %
30 – 20	100	10
40 – 30	150	15
60 – 40	700	70
60 فأكثر	50	5
المجموع	1000	100

ويبدو واضحاً من هذا الجدول ان اكثر العاملين هم في فئة الأعمار المتقدمة وهذا قد يعني اعراض الشباب عن اختيار هذا المضمار من العمل كمهني ومصدر للعيش الأمر الذي يترتب عليه وضع خطط فعالة لتشجيع الشباب على العمل في هذا الميدان الحيوي وتجعل العديد منهم يختاره دون سواه.

التوزيع التكراري: هو تجميع قيم المتغير بعدد من الفئات المتساوية بالطول غالباً.

وهنالك اعتبارات يجب ملاحظتها عند عرض البيانات جدولياً على شكل فئات وتكراراتها واهم هذه الأعتبارات مايلي:

أ/ طول الفئة (h): ويقصد بطول الفئة المسافة (أو الفترة) بين القيمة الصغرى للفئة والقيمة الكبرى لها معبراً عنها بوحدات القياس الخاصة بالبيانات المراد ترتيبها بفئات وعليه فأن طول الفئة (20 – 30) سنة هو 10 سنوات في الجدول اعلاه ولا يشترط ان تكون اطوال جميع الفئات متساوية خاصة اذا كانت هناك أسباب ومبررات ذات علاقة بالهدف من جمع البيانات تملي على عارض البيانات ترتيبها حسب فئات معينة وذات اطوال مختلفة كما هو الحال بالنسبة للفئة 60 فأكثر في الجدول اعلاه فهي فئة ذات طول غير محدد ويطلق عليها (بالفئة المفتوحة) او فئة مفتوحة الطرف.

ب/ المدى (Range): وهو الفرق بين أكبر واصغر قيمة من المجموعة حيث:

المدى (R) = أعلى قيمة – أقل قيمة

جـ/ عدد الفئات (m): ويقصد بها عدد الفئات المراد توزيع البيانات عليها ويعتمد عدد الفئات على عدد البيانات الأحصائية المتوفرة أذ يزداد بزيادتها إلا أن عدد الفئات يجب ان يبقى محدوداً كي يبقى العرض تلخيصياً وفعالاً وعملياً للبيانات.

ويحسب طول الفئة (L) على النحو الآتي: $\boxed{L = \frac{R}{m}}$

ج/ توزيع البيانات الى فئات متساوية الأطوال:

اذا اراد الباحث توزيع بياناته الأحصائية الى فئات متساوية الطول لعدم وجود مبرر لتوزيعها الى فئات غير متساوية الطول فلابد أن يحدد عدد الفئات وطول الفئة الواحدة وقد يحسب عدد الفئات (m) وفق المعادلة الآتية:

$\boxed{m = 1 + 3.3 \log n}$ أو من المعادلة $\boxed{m = 2.5 \sqrt[4]{n}}$

حيث log يرمز الى اللوغاريتم الأعتيادي (للاساس 10)

n يمثل عدد القيم الأحصائية (أي عدد البيانات)

ولابد من تقريب عدد الفئات الى العدد الصحيح الأعلى اللاحق في حالة الحصول على عدد صحيح وكسر عشري كما هو متوقع في اكثر الحالات بسبب طبيعة المعادلة اعلاه.

مثال توضيحي: لو كانت البيانات التالية تمثل أوزان اللحم الصافي بالكغم لخمسة عشر (15) رأساً من الغنم

12 ، 8 ، 8 ، 7 ، 5 ، 6 ، 8.5 ، 10 ، 8 ، 9 ، 11 ، 19 ، 7.5 ، 20 ، 14

المطلوب:

عرض البيانات اعلاه بجدول تكراري ذي فئات متساوية الطول

خطوات الحل هي: 1// نجد عدد الفئات m

$m = 1 + 3.3 \log n \Rightarrow m = 1 + 3.3 \log 15 \Rightarrow$

$m = 1 + (3.3)(1.176) \Rightarrow m = 1 + 3.8808 \Rightarrow$

$m = 4.8808$

ملاحظة: $\log 3 = 0.476$ ، $\log 2 = 0.3$

وعليه فأن عدد الفئات ((m = 5)) [بعد التقريب الى العدد الصحيح الأعلى]

2// نجد المدى R

$R = 20 - 5 = 15$

3// نجد طول الفئة L

$L = \frac{R}{m} = \frac{15}{5} = 3$

4// الجدول التكراري المطلوب سيكون

الفئة (الوزن بالكغم)	التكرار (عدد رؤوس الغنم)
8 – 5	4
11 – 8	6
14 – 11	2
17 – 14	1
20 – 17	2
المجموع /	15

ملاحظة: يفضل في بداية الحل ترتيب الاوزان تصاعدياً أو تنازلياً لتسهيل دراستها وكما يلي:

5 و 6 و 7 و 7.5 و 8 و 8 و 8 و 8.5 و 9 و 10 و 11 و 12 و 14 و 19 و 20

ثم نبدأ بخطوات الحل الأربعة

تلميح: يلاحظ من الجدول في الخطوة الرابعة من الحل للمثال السابق ان كل فئة تشمل على القيم التي تساوي بالضبط حدها الأدنى وتقل عن حدها الأعلى باستثناء الفئة الأخيرة التي تشمل حدها الأعلى لأن اعطائها فئة خاصة بها سيجعل عدد الفئات ستة بدلاً من خمسة وهذا مخالف لعدد الفئات الذي تم تحديده من سياق الحل.

طريقة العرض البياني Graphic presentation

تعرض البيانات احياناً بأشكال مختلفة كالدوائر المجزأة والأعمدة والخطوط المنكسرة وغيرها بحيث يتمكن القارئ من معرفة الأفكار والأتجاهات التي تتضمنها البيانات وذلك لمجرد القاء نظرة سريعة على الشكل البياني الذي يمثلها. أما مسألة اختيار هذا أو ذاك الشكل البياني فأنها تعتمد على الهدف من العرض والأمكانيات الفنية المتوفرة وعادة ما يجد القارئ اشكالاً بيانية مختلفة في النشرات الخاصة بهذه أو تلك المؤسسة أو في المجلات والكتب العلمية. وسوف نركز على ثلاثة أنواع من العرض البياني بسبب شيوع استخدامها في البحوث العلمية والنشرات واللافتات الدعائية وهذه الأنواع هي:

أ- طريقة الأعمدة ب- طريقة المدرج التكراري ج- طريقة المضلع التكراري

ولتوضيح اوجه الأختلاف بين الطرق الثلاثة للعرض البياني نأخذ المثال الآتي وفيه الجدول التكراري لغلة 20 سلالة من سلالات محصول الحنطة.

الغلة للدونم (كغم)	عدد السلالات (التكرار)
400 - 200	4
700 - 400	13
800 - 700	3

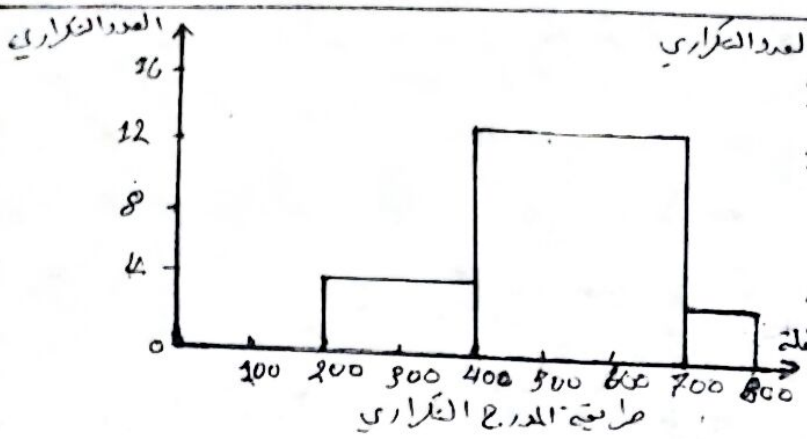
أما كيفية عرض هذه البيانات بالأشكال البيانية الثلاثة نبدأ الحل بأيجاد مركز كل فئة والذي يحسب على النحو الآتي:

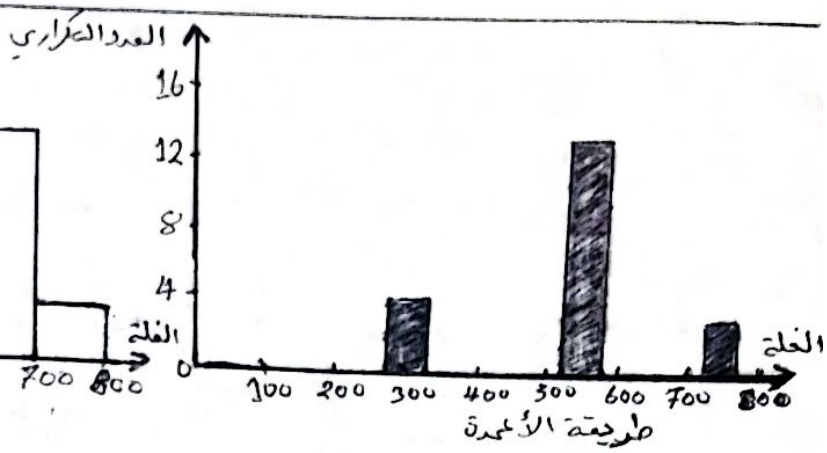
$$\text{مركز الفئة} = \frac{\text{قيمة الحد الأعلى} + \text{قيمة الحد الأدنى}}{2}$$

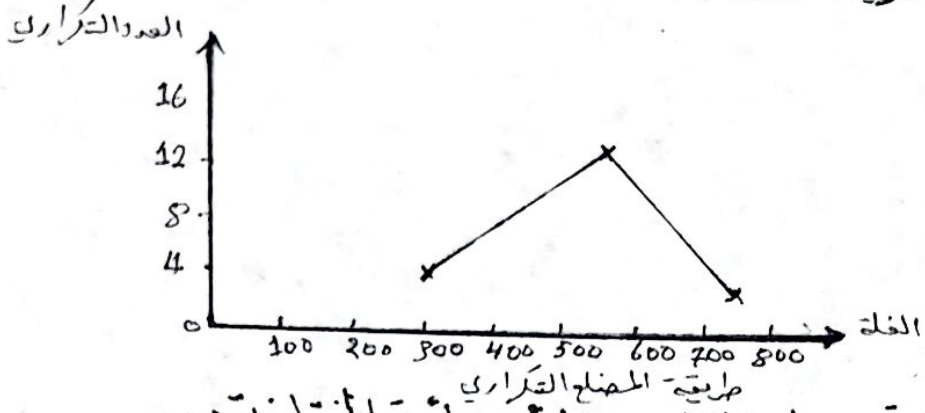
وبذلك سنحصل على جدول تكراري يتضمن الفئة ومركز الفئة والتكرار وكما هو مبين أدناه:

الفئة	مركز الفئة	التكرار
400 - 200	300	4
700 - 400	550	13
800 - 700	750	3

الأحصاء Statistics







طريقة حساب المقاييس الأحصائية المختلفة :-

وهي من الطرق الأولى لعرض وتلخيص البيانات والتي تأخذ بنظر الأعتبار خصوصية تلك البيانات بما في ذلك تركيزها حول قيمة معينة ومدى تشتتها عن بعضها البعض وهذه الطرق عادة ما تعطي قيمة واحدة قادرة على اعطاء القارئ تصوراً مفيداً لطبيعة توزيع قيم المتغير المدروس وهذه القيم تمثل مقاييس احصائية مختلفة بعضها يتعلق بخاصية تركيز البيانات والبعض الآخر بخاصية انتشارها أو اختلافها وتبين الفقرات التالية المقاييس الأحصائية الأكثر تداولاً وفائدة .

مقاييس النزعة المركزية (Measures of central tendency)

عند اخذ عدة قياسات لمتغير معين كالوزن والطول وضغط الدم وعدد الكريات الحمر نلاحظ ان هذه القياسات (القيم) ككل يمكن ان تمثل بقيمة معينة ويطلق على هذه الظاهرة مصطلح ((النزعة المركزية)) ومن اهم مقاييس النزعة المركزية مايلي :

أولاً : الوسيط (Median) : هو القيمة التي تحتل المرتبة الوسطى عندما ترتب القيم قيد الدرس تصاعدياً أو تنازلياً .

اذا كان عدد القيم (ن) عدداً فردياً فأن مرتبة الوسيط = $\frac{ن+1}{2}$

اذا كان عدد القيم (ن) عدداً زوجياً فأن قيمة الوسيط تقع بين منتصف المسافة بين قيمتين هما : المرتبة الوسطى الأولى = $\frac{ن}{2}$

المرتبة الوسطى الثانية = $\frac{ن}{2} + 1$

وبذلك تكون قيمة الوسيط = $\frac{\text{القيمة التي تحتل المرتبة الوسطى الاولى} + \text{القيمة التي تحتل المرتبة الوسطى الثانية}}{2}$

الأحصاء Statistics

مثال (1) : ما هو الوسيط لمجموعة القيم 9 و 16 و 12 و 11 و 8 و 9 و 10

الحل : نرتب القيم تصاعدياً فتصبح 8 و 9 و 9 و 10 و 11 و 12 و 16

مرتبة الوسيط = $\frac{1+7}{2}$ = $\frac{8}{2}$ = 4 ⇐ قيمة الوسيط = 10

مثال (2) : ما هو الوسيط لمجموعة القيم 12 و 22 و 7 و 20 و 5 و 7 و 14 و 10

الحل : ترتب القيم تصاعدياً فتصبح 5 و 7 و 7 و 10 و 12 و 14 و 20 و 22

المرتبة الوسطى الأولى = $\frac{8}{2}$ = 4 ⇐ قيمة المرتبة الوسطى الأولى = 10

المرتبة الوسطى الثانية = $\frac{8}{2}$ + 1 = 5 ⇐ قيمة المرتبة الوسطى الثانية = 12

قيمة الوسيط = $\frac{12+10}{2}$ = $\frac{22}{2}$ = 11

مثال (3) : اوجد قيمة الوسيط لمجموعة البيانات التالية :

الوزن (بالكغم) (القيمة)	عدد الحيوانات (التكرار)
5	3
6.5	7
8	5

الحل : هناك طريقتان لمعالجة هذه المسألة وهما :

أ/ أعادة القيم البالغ عددها (15) الى وضعها الأصلي أي :

5 و 5 و 5 و 6.5 و 6.5 و 6.5 و 6.5 و 6.5 و 6.5 و 6.5 و 8 و 8 و 8 و 8 و 8

مرتبة الوسيط = $\frac{1+15}{2}$ = $\frac{16}{2}$ = 8 ⇐ قيمة الوسيط = 6.5

ب/ نعمل جدول التكرار التجميعي التصاعدي كما هو مبيناً في الجدول الآتي :

الوزن بالكغم	عدد الحيوانات (التكرار)	التكرار التجميعي التصاعدي (آخر مرتبة احتلتها القيمة المناظرة)
5	3	3
6.5	7	10
8	5	15
	المجموع 15	

مرتبة الوسيط = $\frac{1+15}{2}$ = 8

فتكون قيمة الوسيط هي القيمة المناظرة لأول تكرار تجميعي تصاعدي قيمته أكبر أو تساوي مرتبة الوسيط لذلك نجد قيمة الوسيط = 6.5 تناظر التكرار التجميعي التصاعدي 10 وهو أكبر من 8

مثال (4) : جد قيمة الوسيط للبيانات التالية التي تمثل غلة 18 شجرة

الغلة بالكغم (القيمة)	عدد الأشجار (التكرار)
5 - 1	2
20 - 5	6
25 - 20	10

الحل/

بما ان عدد الاشجار ن = 18 وهو عدد زوجي فأن

المرتبة الوسطى الأولى = $\frac{18}{2}$ = 9 و المرتبة الوسطى الثانية = $\frac{18}{2}$ + 1 = 10

اما جدول التكرار التجميعي التصاعدي هو

فئة الغلة (بالكغم)	التكرار	التكرار التجميعي التصاعدي
5 – 1	2	2
20 – 5	6	8
25 – 20	10	18

نلاحظ من الجدول اعلاه ان أول تكرار تجميعي تصاعدي اكبر أو يساوي مرتبة الوسيط هو 18 لذا فأن الفئة الوسيطية المناظرة هي (20 – 25) وتكرارها 10 وقيمة الوسيط نجدها بتطبيق المعادلة الآتية :

$$\text{قيمة الوسيط} = \text{أ} + \frac{\text{ع} - \text{أ}}{\text{ك}} \times (\text{ر} - \text{ت})$$

حيث ع – أ = طول الفئة الوسيطية

أ تمثل الحد الأدنى لفئة الوسيط (أ = 20) ، ع الحد الأعلى لفئة الوسيط (ع = 25)

ك تكرار فئة الوسيط (ك = 10) ، ر المرتبة الوسط وهي قيمتان ر = 9 ، ر = 10

ت تمثل التكرار التجميعي التصاعدي للفئة السابقة لفئة الوسيط وهي $\boxed{\text{ت} = 8}$

$$\text{قيمة المرتبة الوسط الأولى} = 20 + \frac{20-25}{10} \times (9-8) = 20.4$$

$$\text{قيمة المرتبة الوسط الثانية} = 20 + \frac{20-25}{10} (10-8) = 20.8$$

$$\text{قيمة الوسيط} = \frac{20.8 + 20.4}{2} = 20.6 \text{ Kg.}$$

ثانياً : المنوال (الشائع) Mode : هو القيمة (أو القيم) الأكثر شيوعاً أو وجوداً بين مجموعة القيم قيد الدرس .

مثال (1) : جد قيمة المنوال للقيم التالية التي تمثل اطوال (15) أشجار مقاسة بالمتر :

13 ، 14 ، 16 ، 14 ، 12 ، 12 ، 14 ، 15 ، 14 ، 15

الحل : نعرض هذه البيانات بجدول تكراري ومنه نشخص القيمة (أو القيم) الأكثر شيوعاً التي تقابل (تناظر) اكبر تكراراً وكما يلي :

القيمة (الطول بالمتر)	التكرار (عدد الأشجار)
13	1
14	4
16	1
12	2
15	2

نلاحظ من الجدول المجاور ان الطول (14m) هو الأكثر شيوعاً وعليه فأن قيمة المنوال = 14 م

مثال (2): جد قيمة المنوال للقيم التالية التي تمثل أوزان 20 سمكة:

الوزن بالكغم	عدد الأسماك
2.0	3
2.5	7
4.0	7
4.5	3

الحل: بما أن القيمتين (2.5 و 4.0) كغم متساويتان في التكرار (أي الشيوع) فأن كل منهما تمثل قيمة مستقلة للمنوال وعليه فأن قيمتي المنوال هما (2.5 و 4.0) كغم.

ملاحظة: هناك حالات تتطلب تحديد قيمة المنوال لبيانات مصنفة حسب فئات معينة وتكراراتها المناظرة ولمعالجة هذه الحالة فأننا نجد قيمة المنوال حسب الخطوات التالية:

أ- تشخيص الفئة المنوالية (أي الفئة الأكثر تكراراً)

ب- تحدد قيمة المنوال من تطبيق معادلة (كارل بيرسون) الآتية:

$$\text{قيمة المنوال} = \text{أ} + \frac{\text{ك}_{\text{م}} - \text{ك}_{\text{ق}}}{(\text{ك}_{\text{م}} - \text{ك}_{\text{ق}}) + (\text{ك}_{\text{م}} - \text{ك}_{\text{ب}})} \times (\text{ع} - \text{أ})$$

حيث: أ تمثل الحد الأدنى للفئة المنوالية، ع تمثل الحد الأعلى للفئة المنوالية

$\text{ك}_{\text{م}}$ تمثل تكرار الفئة المنوالية، $\text{ك}_{\text{ق}}$ تمثل تكرار الفئة السابقة للفئة المنوالية

$\text{ك}_{\text{ب}}$ تمثل تكرار الفئة اللاحقة للفئة المنوالية.

مثال (3): جد قيمة المنوال للبيانات التالية:

الفئة	التكرار
20 – 16	5
24 – 20	10
28 – 24	40
32 – 28	20

الفئة المنوالية ←

التكرار السابق $\text{ك}_{\text{ق}}$ ←

التكرار المنوالي $\text{ك}_{\text{م}}$ ←

التكرار اللاحق $\text{ك}_{\text{ب}}$ ←

الحل / بما ان الفئة (24 – 28) هي الفئة المنوالية لأن تكرارها (40) يفوق جميع التكرارات الأخرى، وبتطبيق معادلة كارل بيرسون يؤدي الى ما يلي:

$$\text{قيمة المنوال} = 24 + \frac{10 - 40}{(20 - 40) + (10 - 40)} \times (24 - 28)$$

$$= 24 + \frac{30}{50} \times 4 = 26.4$$

ملاحظة: نلاحظ ان القيمة المحسوبة للمنوال لابد لها ان تقع داخل حدود الفئة المنوالية.

وتجدر الأشارة الى ضرورة حساب قيم المنوال المختلفة وفق معادلة كارل بيرسون في حالة وجود أكثر من فئة منوالية واحدة لنفس مجموعة البيانات.

ثالثاً: المتوسط الحسابي (Arithmetic Mean): وهو أكثر مقاييس النزعة المركزية شيوعاً واستعمالاً ويطلق عليه أحياناً بالمعدل الحسابي (Arithmetic average) أو الوسط الحسابي ويرمز لهذا المقياس برمزين مختلفين هما:

أ- الرمز (μ) ليمثل المتوسط الحسابي للمجتمع.

ب- الرمز ($\bar{X}$) ليمثل المتوسط الحسابي للعينة.

وتجدر الإشارة الى أن (μ) هي قيمة ثابتة لا تتغير ولهذا فأنها تعتبر من بين معالم أو ثوابت المجتمع أما $\bar{X}$ فأنها تتغير من عينة الى اخرى اعتماداً على العناصر التي تشملها كل عينة ولهذا السبب فأن $\bar{X}$ تعتبر من بين الأحصاءات (Statistics).

أ/ المتوسط الحسابي للمجتمع (μ_x):

تعتمد الطريقة المتبعة لتحديد قيمة المتوسط الحسابي للمجتمع (μ_x) على طبيعة البيانات المتوفرة والتي قد تكون على شكل سلسلة من القيم المنفردة أو فئات من القيم وتكراراتها المطلقة أو النسبية وعليه فأن المعادلة المستخدمة قد تأخذ أحد الأشكال الثلاثة التالية:

١) $$\mu_x = \frac{\sum_{i=1}^{N} X_i}{N}$$ حيث N تمثل حجم المجتمع، X_i تمثل قيمة المتغير (X) للمفرد أو العنصر i

٢) $$\mu_x = \frac{\sum_{i=1}^{N} k_i X_i}{N} = \frac{\sum_{i=1}^{N} k_i X_i}{\sum_{i=1}^{N} k_i}$$ حيث k_i تمثل تكرار القيمة X_i

٣) $$\mu_x = \sum_{i=1}^{N} f_i X_i$$ حيث f_i تمثل التكرار النسبي للقيمة X_i بالنسبة للمجموع الكلي للقيم وتحسب f_i على النحو التالي:

$$\boxed{f_i = \frac{k_i}{N}}$$ أي ان التكرار النسبي = $\frac{\text{تكرار الفئة}}{\text{التكرار الكلي}}$ وأن $\sum_{i=1}^{N} f_i = N$

ب/ المتوسط الحسابي للعينة ($\bar{X}$): ويحسب على النحو التالي:

١) $$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$ حيث n تمثل حجم العينة

٢) $$\bar{X} = \frac{\sum_{i=1}^{n} k_i X_i}{\sum_{i=1}^{n} k_i}$$

٣) $$\bar{X} = \sum_{i=1}^{n} f_i X_i$$ حيث أن $$\boxed{f_i = \frac{k_i}{n}}$$

خواص « $\sum$ »

1) $\sum_{i=1}^{n} A = nA$ 2) $\sum_{i=1}^{n} Ax_i = A\sum_{i=1}^{n} x_i$

3) $\sum_{i=1}^{n} (x_i + f_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} f_i$ 4) $\sum_{i=1}^{n} (x_i - \bar{X}) = 0$

برهان الخاصية الرابعة :

$$\sum_{i=1}^{n} (x_i - \bar{X}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{X}$$

حسب الخاصية (2) فأن :

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

وحسب الخاصية (1) فأن

$$\sum_{i=1}^{n} \bar{X} = n\bar{X}$$

$$\therefore \sum_{i=1}^{n} (x_i - \bar{X}) = n\bar{X} - n\bar{X} = 0$$

تمرين : برهن ما يأتي :

1) $\sum_{i=1}^{n} (x_i - \bar{X})^2 = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}$

2) $\sum_{i=1}^{N} (x_i - \mu_x)^2 = \sum_{i=1}^{N} x_i^2 - \frac{(\sum_{i=1}^{N} x_i)^2}{N}$

ملاحظة :

هنالك متوسط حسابي خاص يطلق عليه المتوسط الحسابي المرجح ويرمز له بالرمز ($\bar{X}_{wt}$) والناشئ عن المعاينة الطبقية التي تؤخذ بموجبها عينة من كل طبقة ويحدد المتوسط الحسابي للعينة المأخوذة من كل طبقة على انفراد ويطلق عليه بالرمز ($\bar{X}_j$) وبما أن الطبقات تختلف في احجامها فأن المتوسطات الحسابية للعينات المأخوذة من هذه الطبقات توزن حسب ما تمثله طبقاتها المناظرة في المجتمع الكلي بهدف الوصول الى متوسط حسابي مرجح يمثل العينة الكلية المأخوذة من المجتمع وعليه فأن المتوسط المرجح ($\bar{X}_j$) يحسب وفق المعادلة الآتية

$$\bar{X}_{wt} = \sum_{j=1}^{N} \bar{X}_j \cdot \frac{N_j}{N}$$

حيث $\bar{X}_j$ تمثل المتوسط الحسابي للعينة المأخوذة من الطبقة رقم ل

N_j تمثل حجم الطبقة رقم ل

N تمثل حجم المجتمع

مثال (1): اذا كانت البيانات التالية تمثل اطوال ثمان نخلات تشكل مجتمعاً قائماً بذاته:
11 و 16 و 16 و 15 و 11 و 16 و 11 و 16 (متراً) والمطلوب هو
المتوسط الحسابي لطول النخلة في هذا المجتمع

الحل:

$$\mu_x = \frac{\sum_{i=1}^{N} X_i}{N}$$

$$= \frac{11+16+16+15+11+16+11+16}{8} = \frac{112}{8} = 14 \text{ metre}$$

هنالك طريقة أخرى لحل المثال اعلاه عن طريق ترتيب البيانات المعطاة بجدول تكراري على النحو التالي:

القيمة (بالمتر)	التكرار
11	3
16	4
15	1
	8 المجموع

$$\mu_x = \frac{\sum_{i=1}^{N} k_i X_i}{\sum_{i=1}^{N} k_i}$$

$$= \frac{3(11)+4(16)+(1)15}{8} = \frac{112}{8} = 14 \text{ m}$$

وهنالك طريقة ثالثة لحل المثال السابق وكما يلي:
نجد التكرار النسبي لكل قيمة من قيم الجدول اعلاه وطبقاً للمعادلة $f_i = \frac{k_i}{N} = \frac{\text{تكرار الفئة}}{\text{التكرار الكلي}}$
ونعمل جدول للبيانات على النحو الآتي:

X_i	f_i
11	0.375 ($\frac{3}{8}$)
16	0.500 ($\frac{4}{8}$)
15	0.125 ($\frac{1}{8}$)
	1.0 المجموع

$$\mu_x = \sum_{i=1}^{N} f_i X_i$$

$$= (0.375)(11) + (0.500)(16) + (0.125)(15)$$

$$= 14 \text{ metre}$$

ملاحظة: لو فرضنا أن البيانات المعطاة في المثال السابق لأطوال 8 نخلات تمثل عينة مأخوذة بشكل عشوائي من مجتمع معين فأن المتوسط الحسابي لطول النخلة الواحدة (أي $\bar{X}$) في هذه العينة يمكن حسابه وفق المعادلة:

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

$$= \frac{11+16+16+15+11+16+11+16}{8} = \frac{112}{8} = 14 \text{ metre}$$

أو وفق المعادلة:

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

$$= (0.375)(11) + (0.500)(16) + (0.125)(15) = 14 \text{ metre}$$

مثال (2) : لو أخذت عينة من تيلات صنف معين من القطن وزنها (5 gm) وفرزت تلك التيلات الى ثلاث فئات لطول التيلة ووزنت تيلات كل فئة وكانت النتائج كما يلي:

فئة طول التيلة (بالسم)	الوزن (بالغم)
2 — 1	0.5
4 — 2	3.0
7 - 4	1.5
المجموع	5.0

والمطلوب تحديد قيمة المتوسط الحسابي لطول الفئة ($\bar{X}$) في العينة .

الحل :

يمكن الوصول الى ذلك عن طريق استخدام مركز الفئة وتكرارها النسبي أي :

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

مركز الفئة (بالسم)	التكرار النسبي	
X_i	f_i	
1.5	0.1	$\frac{0.5}{5.0}$
3.0	0.6	$\frac{3.0}{5.0}$
5.5	0.3	$\frac{1.5}{5.0}$
المجموع	1.0	

$$= (0.1)(1.5) + (0.6)(3.0) + (0.3)(5.5)$$

$$= 3.6 \text{ cm.}$$

مثال (3) : ما قيمة المتوسط الحسابي المرجح لانتاج الدونم الواحد لمحصول معين في محافظة ذي قار تختلف حقولها المزروعة بهذا المحصول بحيث تشكل هذه الحقول ثلاث مجاميع خصوبة (اي ثلاث طبقات) تختلف عن بعضها البعض مع وجود تقارب بالنسبة لخصوبتها على ان المتوسط الحسابي لعينة كل طبقة من الطبقات ($\bar{X}_j$) كان محسوباً وعلى فرض ان نتائج المعاينة كانت على النحو الآتي :

رقم الطبقة j	مساحة الطبقة المزروعة بالمحصول قيد الدرس (دونم) N_j	المتوسط الحسابي لانتاج كل دونم في العينة المأخوذة من الطبقة (كغم) $\bar{X}_j$
1	1000	650
2	6000	420
3	3000	530
المجموع	10000	

الحل/ المتوسط الحسابي المرجح لانتاج الدونم الواحد على نطاق محافظة ذي قار يحسب وفق المعادلة الآتية

$$\bar{X}_{wt} = \sum_{j=1}^{N} \bar{X}_j \frac{N_j}{N}$$

$$= (650)\left(\frac{1000}{10000}\right) + (420)\left(\frac{6000}{10000}\right) + (530)\left(\frac{3000}{10000}\right)$$

$$= (650)(0.1) + (420)(0.6) + (530)(0.3) = 476 \text{ Kg.}$$

ملاحظة: في المثال السابق مثال (3) لايجوز جمع المتوسطات الحسابية ($\bar{X}_j$) للعينات الثلاثة وتقسيمها على عددها (أي على 3) وذلك لأنها غير متساوية في الأهمية بسبب اختلاف المساحات المزروعة بهذا المحصول في كل طبقة (أي اختلاف أحجام الطبقات الثلاثة).

تمرين: جد قيمة كل من الوسيط والمنوال والمتوسط الحسابي للبيانات التالية التي تمثل الدخل السنوي لعدد من الحيازات الزراعية:

الدخل السنوي (دينار)	عدد الحيازات
أقل من 250	1000
250 — 1250	4500
1250 — 5000	1500
5000 فأكثر	700

الجواب/
قيمة الوسيط = 883.38
قيمة المنوال = 788.46
قيمة المتوسط الحسابي = 930.84

تلميح: بما أن الفئة الأخيرة مفتوحة الطرف من الأعلى (أي حدها الأخير غير محدد) فأن مركز هذه الفئة غير معروف وفي مثل هذه الحالات يمكن الاستعانة بالعلاقة التالية التي تربط مقاييس النزعة المركزية الثلاثة (المتوسط الحسابي، الوسيط، المنوال) ومنها نجد المتوسط الحسابي بشكل تقريبي.

$$\boxed{\text{المتوسط الحسابي} = \left(\frac{3}{2}\right)(\text{الوسيط}) - \left(\frac{1}{2}\right)(\text{المنوال})}$$

رابعاً: المتوسط الهندسي (Geometric Mean): هو القيمة العددية للجذر النوني لحاصل ضرب مجموعة من القيم

فلو رمزنا للمتوسط الهندسي بالرمز (G) وللقيم بالرموز ($X_1, X_2, \ldots, X_n$) فأن

$$G = \sqrt[n]{(X_1)(X_2)\cdots(X_n)} \Rightarrow G = \left[(X_1)(X_2)\cdots(X_n)\right]^{\frac{1}{n}}$$

وعادة ما يستعان باللوغاريتم الأعتيادي لأيجاد قيمة المتوسط الهندسي أي أن:

$$\log G = \log\left[(X_1)(X_2)\cdots(X_n)\right]^{\frac{1}{n}} \Rightarrow$$

$$\log G = \frac{1}{n}\left[\log X_1 + \log X_2 + \cdots + \log X_n\right]$$

وعليه سيكون قيمة المتوسط الهندسي هي

$$\log G = \frac{\sum_{i=1}^{n} \log X_i}{n}$$

وهذا يعني ان لوغاريتم المتوسط الهندسي ما هو الا المتوسط الحسابي للوغاريتمات القيم قيد الدرس

ملاحظة: 1) ان المتوسط الهندسي لمجموعة من القيم يكون دائماً أقل من المتوسط الحسابي لها

2) لا يمكن حساب المتوسط الهندسي حسب المعادلة الأولى اعلاه عندما تكون قيمة احد القيم = صفر وكذا الحال بالنسبة لوجود عدد فردي من القيم السالبة

وفي حالة الجداول التكرارية فأن المتوسط الهندسي يحسب وفق المعادلة التالية :

$$\log G = \frac{k_1 \log X_1 + k_2 \log X_2 + \cdots + k_n \log X_n}{n}$$

حيث أن :

$$\log G = \frac{\sum_{i=1}^{n} k_i \log X_i}{\sum_{i=1}^{n} k_i}$$

k_i يمثل تكرار القيمة X_i

X_i تمثل القيمة رقم i والتي قد تكون قيمة قائمة بذاتها أو ممثلة لمركز الفئة المناظرة لها

مثال (1) : جد المتوسط الهندسي للقيم (2 , 4 , 8)

الحل : قيمة المتوسط الهندسي هي

$$G = \sqrt[3]{(2)(4)(8)} = \sqrt[3]{64} = 4$$

وبأستخدام طريقة اللوغاريتم نحصل على

$$\log G = \frac{1}{3}[\log 2 + \log 4 + \log 8]$$

$$\log G = \frac{1}{3}[\log 2 + 2\log 2 + 3\log 2]$$

$$\log G = \frac{1}{3}[6\log 2] \Rightarrow \log G = 2\log 2 \Rightarrow$$

$$\log G = \log (2)^2 \Rightarrow G = 4$$

تمرين (1) : جد قيمة المتوسط الهندسي للقيم (1 , 5 , 125 , 625) بطريقتي الجذر واستخدام اللوغاريتم الأعتيادي . الجواب / 25

تمرين (2) : جد الوسط الحسابي للقيم (4000 , 6000) ثم جد الوسط الهندسي بطريقتي الجذر واستخدام اللوغاريتم الاعتيادي ثم قارن بين قيمتي الوسط الحسابي والوسط الهندسي وماذا تستنتج من تلك المقارنة .

الجواب / الحسابي 5000
الهندسي 4898

فائدة : $\sqrt{2} = 1.414$ ، $\sqrt{3} = 1.732$

تمرين (3) : في صفة اعمار احدى المدارس في قرية ما شارك (30) فرداً اعمارهم مبينة في الجدول الآتي :

الفئة (العمر)	10 – 12	13 – 15	16 – 18	19 – 21	22 – 24	25 – 27
التكرار (العدد)	3	5	7	8	6	1

والمطلوب ايجاد الوسط الهندسي لأعمارهم والوسط الحسابي وقارن بين الوسطين علماً أن : $\log 11 = 1.0413$ ؛ $\log 14 = 1.1461$ ؛ $\log 17 = 1.2304$ ؛ $\log 20 = 1.3010$ ؛ $\log 23 = 1.3617$ ؛ $\log 26 = 1.4149$ ؛ $\log 17.73 = 1.2487$

تلميح : استخدم قانون الجداول التكرارية اعلاه . الجواب / الهندسي 17.73 والحسابي 17.83

خامساً : المتوسط التوافقي (Harmonic Mean) : هو مقلوب المتوسط الحسابي لمقلوبات القيم ويرمز له بالرمز (H) .

فإذا كانت القيم هي (X_1 و X_2 و ---- و X_n) فإن المتوسط التوافقي لها هو :

$$H = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \cdots + \frac{1}{X_n}} \Rightarrow H = \frac{n}{\sum_{i=1} \frac{1}{X_i}}$$

مثال (1) : جد المتوسط التوافقي للأعداد : 10 و 20 و 30 و 40

الحل :

$$H = \frac{4}{\frac{1}{10} + \frac{1}{20} + \frac{1}{30} + \frac{1}{40}} = \frac{4}{\frac{12+6+4+3}{120}} = \frac{480}{25} = 19.2$$

تمرين (1) جد المتوسط التوافقي للأعداد : 2 و 3 و 4 و 8 و 9 و 27 ثم جد المتوسط الحسابي والهندسي .

الجواب : $H = 4.4$; $\bar{X} = 8.8$ و $G = 6$

العلاقة بين الأوساط الحسابية والهندسية والتوافقية :

هناك علاقتان هما : (1) $\bar{X} \geq G \geq H$ (2) $\bar{X} \cdot H = G^2$

تمرين (2) جد قيمة كل من المتوسط الحسابي والمتوسط التوافقي والمتوسط الهندسي والوسيط لأوزان ثلاثة أطفال مقاسة بالكغ : 27 ; 25 و 40 ثم تحقق من العلاقتين أعلاه .

الجواب : قيمة الوسيط = 27 و $\bar{X} = 30.6$; $H = 29.4$; $G = 30$

تمرين (3) ما هما العددان اللذان المتوسط الهندسي لهما يساوي (8) والمتوسط التوافقي لهما يساوي (6.4) .

الجواب : $X_1 = 4$; $X_2 = 16$ أو بالعكس

اذا كانت (X_1 و X_2 و ---- X_n) تمثل مراكز فئات من جدول تكراري وكانت (k_1 و k_2 و ---- k_n) تمثل التكرارات المناظرة لها وكان مجموع التكرار = n فإن المتوسط التوافقي سيكون على النحو الآتي :

$$H = \frac{n}{\sum_{i=1}^{n} \frac{k_i}{X_i}}$$

مثال (2) : جد المتوسط التوافقي لجدول التوزيع التكراري الآتي :

الفئات	6-4	9-7	12-10	15-13	18-16	21-19
التكرار	1	4	5	6	3	1

الحل : نضيف الى الجدول المعطى في السؤال عمود يمثل مراكز الفئات X_i ومن الجدول الناتج نلاحظ ان مجموع التكرار (n = 20)

الفئة	مركز الفئة X_i	التكرار k_i
6 - 4	5	1
9 - 7	8	4
12 - 10	11	5
15 - 13	14	6
18 - 16	17	3
21 - 19	20	1
المجموع		n = 20

قيمة المتوسط التوافقي نجده من العلاقة الآتية :

$$H = \frac{\sum_{i=1}^{n} K_i}{\sum_{i=1}^{n} \frac{K_i}{X_i}} \Rightarrow H = \frac{20}{\frac{1}{5} + \frac{4}{8} + \frac{5}{11} + \frac{6}{14} + \frac{3}{17} + \frac{1}{20}} \Rightarrow$$

$$H = \frac{20}{\frac{1}{5} + \frac{1}{2} + \frac{5}{11} + \frac{3}{7} + \frac{3}{17} + \frac{1}{20}} \Rightarrow$$

$$H = \frac{20}{\frac{5236 + 13090 + 11900 + 11220 + 4620 + 1309}{26180}} \Rightarrow$$

$$H = \frac{523600}{47375} \Rightarrow H = 11.05$$

تمرين (4) جد المتوسط التوافقي لجدول التوزيع التكراري التالي :

الجواب: 14.28

الفئات	7-5	10-8	13-11	16-14	19-17	22-20
التكرار	1	3	4	9	6	7

التشتت – مقاييس التشتت

التشتت في أي مجموعة من القيم هو التباعد بين مفرداتها أي التفاوت بينها ويكون التشتت صغيراً متى ما كانت القيم قريبة من بعضها اي اذا كان التفاوت بين المفردات قليلاً وبالطبع يزداد التشتت متى ما كان التباعد بين القيم كبيراً اي اذا كان التفاوت بين المفردات كبيراً وبهذا يكون لدينا مقياساً لمقدار تجانس المجموعات الإحصائية يسمى مقياس التشتت واهم مقاييس التشتت هي :

أولاً : المدى : هو الفرق بين اكبر واصغر قيمة للمتغير في حالة البيانات غير المجمعة أو هو الفرق بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى في حالة البيانات المجمعة

مثال (1) : جد المدى لأسعار ستة انواع من قماش الصوف (بالدينار) والذي طرح في الأسواق من قبل شركة المخازن العراقية كما يأتي : 8000 ؛ 7500 ؛ 15000 ؛ 4000 ؛ 3000 ؛ 11000

الحل : المدى = اكبر قيمة للمتغير – اصغر قيمة للمتغير

= 15000 – 3000 = 12000

مثال (2) : جد قيمة المدى لمجموعة القيم التالية : 45 ؛ 100 ؛ 45 ؛ 55 ؛ 95 ؛ 50

الحل : المدى = 100 – 45 = 55

هذا يعني ان قيم المجموعة لا تتشتت عن بعضها البعض بأكثر من (55) .

مثال (3): الجدول التكراري الآتي يمثل أوزان (30) طالباً في كلية العلوم مقاسة بالكغم والمطلوب ايجاد المدى لتلك الأوزان

الوزن (بالكغم)	45 - 49	50 - 54	55 - 59	60 - 64	65 - 69	70 - 74
العدد	2	5	9	7	4	3

الحل: المدى = الحد الأعلى للفئة الأخيرة - الحد الأدنى للفئة الأولى

$= 74 - 45 = 29$

ثانياً: الأنحراف المتوسط: هو متوسط الأنحراف المطلق لقيم مفردات التوزيع عن وسطه الحسابي ويرمز له بالرمز (AD) وهو مختصر للكلمتين (Average Deviation) ويحسب على النحو الآتي:

$$AD = \frac{\sum_{i=1}^{n} |X_i - \bar{X}|}{n}$$

حيث ان البسط يمثل مجموع القيم المطلقة لأنحرافات قيم (X_i) عن الوسط الحسابي $(\bar{X})$ اما (n) فتمثل عدد القيم في المجموعة.

مثال: جد قيمة الأنحراف المتوسط للبيانات التالية: 3 و 1.5 و 2.5 و 1 و 2

الحل: نجد قيمة المتوسط الحسابي $\bar{X}$ أولاً وكما يلي:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{3 + 1.5 + 2.5 + 1 + 2}{5} = \frac{10}{5} = 2$$

ثم نعمل جدولاً نبين فيه مدى اختلاف (أو تشتت) كل قيمة من القيم الخمسة عن قيمة المتوسط الحسابي ثانياً وكما يلي:

X_i	$X_i - \bar{X}$	$\lvert X_i - \bar{X} \rvert$
3	+1	1
1.5	-0.5	0.5
2.5	+0.5	0.5
1	-1	1
2	0	0
المجموع 10	0	3

$$AD = \frac{\sum_{i=1}^{n} |X_i - \bar{X}|}{n}$$

$$AD = \frac{3}{5}$$

$$AD = 0.6$$

وهذا يعني انحراف كل قيمة عن المتوسط الحسابي يساوي 0.6 في المتوسط

ملاحظة: نجد ان مجموع الفروق عن المتوسط الحسابي يساوي صفراً دائماً أي ان $\sum_{i=1}^{n}(X_i - \bar{X}) = 0$ كما هو مبين في الجدول أعلاه على سبيل المثال لا على سبيل الحصر.

تمرين: جد قيمة الأنحراف المتوسط للقيم التالية:

50 و 45 و 30 و 25 و 60 و 54

الجواب: $AD = 11$

ثالثاً: التباين (Variance): هو مقدار موجب يعبّر عن قيمة متوسط مربع التشتتات أو الأختلافات ويمكن التعبير عن ذلك رياضياً على النحو التالي لتباين المجتمع

$$\sigma_x^2 = \frac{\sum_{i=1}^{N} (x_i - \mu_x)^2}{N} \Longrightarrow \sigma_x^2 = \frac{\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}}{N}$$

حيث μ_x يمثل المتوسط الحسابي للمجتمع المتغير X

N يمثل حجم المجتمع ، σ_x^2 تمثل تباين المجتمع

كما ويمكن التعبير عن تباين العينة على النحو التالي :

$$S_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{n-1} \Longrightarrow S_x^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$$

حيث $\bar{X}$ يمثل المتوسط الحسابي للعينة

n يمثل حجم العينة ، S_x^2 تمثل تباين العينة

وتجدر الأشارة الى ان سبب استخدام $(n-1)$ بدلاً من (n) في مقام المعادلتين اعلاه يعود الى اعتبارات رياضية تتعلق بالتقديرات غير المتحيزة .

مثال (1): جد قيمة التباين للقيم التالية التي تمثل وزن اللحم الصافي لستة رؤوس من الغنم تمثل مجتمعاً قائماً بذاته : 35 ، 34 ، 40 ، 38 ، 37 ، 32 (كغم)

الحل: ان قيمة المتوسط الحسابي للمجتمع μ_x تساوي

$$\mu_x = \frac{35+34+40+38+37+32}{6} = \frac{216}{6} = 36 \text{ kg}$$

وعليه فأن قيمة التباين للمجتمع σ_x^2 هي:

$$\sigma_x^2 = \frac{[(35)^2+(34)^2+(40)^2+(38)^2+(37)^2+(32)^2] - \frac{(216)^2}{6}}{6}$$

$$\sigma_x^2 = \frac{7818 - \frac{46656}{6}}{6}$$

$$\sigma_x^2 = \frac{7818-7776}{6} = \frac{42}{6} = 7 \ (Kg.)^2$$

هذا يعني ان متوسط تباين كل قيمة عن المتوسط الحسابي للمجتمع هو (7) كيلوغرامات مربعة .

كما ويلاحظ بأن التباين يقيس التشتت بوحدات مربعة كالكيلوغرامات المربعة كما في المثال اعلاه وحسب طبيعة البيانات المعطاة .

مثال (2): جد قيمة التباين للقيم التالية التي تمثل وزن اللحم الصافي لستة رؤوس من الغنم تمثل عينة مأخوذة من مجتمع معين: 35 و 34 و 40 و 38 و 37 و 32 (كغم)

الحل: ان قيمة المتوسط الحسابي للعينة ($\bar{X}$) تساوي

$$\bar{X} = \frac{35+34+40+38+37+32}{6} = \frac{216}{6} = 36 \text{ kg}$$

وعليه فأن قيمة التباين للعينة S_x^2 هي:

$$S_x^2 = \frac{7818 - \frac{(216)^2}{6}}{6-1} = \frac{42}{5} = 8.4 \ (\text{kg.})^2$$

تمرين: جد قيمة التباين للقيم التالية: 8 و 10 و 14 و 16 اذا كانت هذه القيم تمثل
أ/ مجتمعاً قائماً بذاته ب/ عينة مأخوذة من مجتمع معين

الجواب: $10 (\text{unit})^2$ ، $13.33 (\text{unit})^2$

رابعاً: الانحراف المعياري (Stamdard Deviation): هو عبارة عن الجذر التربيعي للتباين وعليه فأن الانحراف المعياري للمجتمع (σ_x) يحسب من احدى المعادلتين الآتيتين:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu_x)^2}{N}} \Longrightarrow \sigma_x = \sqrt{\frac{\sum_{i=1}^{N} x_i^2 - \frac{(\sum_{i=1}^{N} x_i)^2}{N}}{N}}$$

وبأسلوب مماثل فأن قيمة الانحراف المعياري للعينة (S_x) يمكن ان يحسب من احدى المعادلتين الآتيتين:

$$S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}} \Longrightarrow S_x = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n-1}}$$

مثال: جد قيمة الانحراف المعياري للبيانات المعطاة في المثالين السابقين 1 و 2 أعلاه

الحل: ان قيمة الانحراف المعياري للمجتمع σ_x هي:

$$\sigma_x = \sqrt{\frac{7818 - \frac{(216)^2}{6}}{6}} = \sqrt{7} = 2.64 \text{ kg}$$

أما قيمة الانحراف المعياري للعينة S_x هي:

$$S_x = \sqrt{\frac{7818 - \frac{(216)^2}{6}}{6-1}} = \sqrt{8.4} = 2.9 \text{ kg}$$

ملاحظة: أن الانحراف المعياري يعتبر من أكثر مقاييس التشتت شيوعاً سواء بصيغته المباشرة أو عن طريق المقاييس المشتقة منه كمعامل الاختلاف (معامل التباين).

خامساً: معامل الاختلاف (معامل التباين) Coefficient of Variation

هو النسبة المئوية لحاصل قسمة الانحراف المعياري على المتوسط الحسابي ويرمز لمعامل الاختلاف بالرمز (C.V) ويحسب على النحو الآتي:

$$C.V._{\text{للمجتمع}} = \frac{\sigma_x}{\mu_x} \cdot \%100 \qquad C.V._{\text{للعينة}} = \frac{S_x}{\bar{X}} \cdot \%100$$

مثال: احسب معامل الارتباط للبيانات المعطاة في المثالين 1 و 2 صفحة (19) وصفحة (20).

$$C.V._{\text{للمجتمع}} = \frac{2.64}{36} \%100 = 7.33\%$$

$$C.V._{\text{للعينة}} = \frac{2.9}{36} \%100 = 8.05\%$$

تمارين عامة:

(1) البيانات التالية تمثل اوزان (10) أطفال مقدرة بالكغ مأخوذة من مدرسة ما

10 و 8 و 12 و 7 و 10 و 16 و 13 و 9 و 11 و 14

المطلوب:

أ/ حساب قيمة كل من الوسيط والمنوال والمتوسط الحسابي والمدى والتباين والانحراف المعياري ومعامل الاختلاف والانحراف المتوسط

الجواب: 10.5 و 10 و 11 و 9 و 7.77 و 2.78 و 25.3% و 2.2

ب/ وزع البيانات الى فئات متساوية الطول بجدول تكراري

(2) اذا كانت اوزان رؤوس الغنم في قطيع معين موزعة على النحو التالي لعينة مأخوذة من هذا القطيع

الوزن (بالكغم)	20 – 25	25 – 35	35 – 40	40 – 50
عدد الرؤوس	20	60	10	6

المطلوب:

أ/ عرض البيانات بشكل بياني وحسب طريقة: 1) الاعمدة 2) المدرج التكراري 3) المضلع التكراري

ب/ حساب قيمة كل من الوسيط والمنوال والمتوسط الحسابي والمتوسط التوافقي والمتوسط الهندسي والمدى والانحراف المعياري ومعامل الاختلاف.

الجواب: بالنسبة للفرع ب:

قيمة الوسيط = 29.7 و قيمة المنوال = 29.44 و المتوسط الحسابي = 30.16 و المتوسط التوافقي = 29.35

المتوسط الهندسي = 29.75 و المدى = 30 و الانحراف المعياري = 4.4 و معامل الاختلاف = 14.58%

تلميح: استعن بالعلاقة $G^2 = \bar{X} \cdot H$ لايجاد قيمة المتوسط الهندسي G

التوزيعات الأحتمالية : أن طبيعة أي توزيع احتمالي يعتمد على نوع المتغيّر العشوائي تحت الدراسة وهناك نوعان من التوزيعات الأحتمالية هما :

أ- التوزيعات الأحتمالية المستمرة (Continuous probability distribution)

وهي توزيعات تصف متغيراً عشوائياً مستمراً تنحصر قيمه بين حدين ودالته موجبة لجميع قيم المتغير X حيث $(-\infty < X < +\infty)$.

وبما أن احتمال اخذ المتغير العشوائي المستمر لأي قيمة معينة يكون صفراً لذلك لا يمكن تمثيل التوزيع الأحتمالي للمتغير بجدول ولكن نعبّر عنه بمعادلة رياضية ويكون الأحتمال عبارة عن مساحة تحت منحني اقتران تلك المعادلة $Y = f(x) \quad (-\infty < X < +\infty)$

وعليه فأن احتمال اتخاذ المتغير العشوائي المستمر قيماً تتراوح بين a و b يمثل المساحة المحصورة على النحو التالي :

$$P(a < x < b) = \int_a^b f(x)\, dx$$

ومن أهم التوزيعات المستمرة هي التوزيع الطبيعي وتوزيع f وتوزيع t وتوزيع Z^2 وغيرها .

ب- التوزيعات الأحتمالية المتقطعة (Discrete probability distribution)

وهي التوزيعات التي تصف متغيراً عشوائياً متقطعاً على شكل جدول لجميع القيم الممكنة للمتغير العشوائي المتقطع مع جميع الأحتمالات المناظرة لكل قيمة .

فلو كانت $P(X_i)$ تمثل احتمال اتخاذ المتغير المتقطع للقيمة X_i حيث ان : ($i = 1, 2, 3, \ldots, N$) فأن الدالة $P(X_i)$ يطلق عليها دالة توزيع متقطع اذا توافر الشرطان التاليان :

١- احتمال اتخاذ المتغير المتقطع للقيمة X_i يساوي أو يزيد على الصفر أي أن

$$P(X_i) \geq 0$$

٢- مجموع احتمالات القيم التي يمكن ان يأخذها المتغير المتقطع يساوي واحد أي أن

$$\sum_{i=1}^{N} P(X_i) = 1 \Longleftrightarrow P(X_1) + P(X_2) + \cdots + P(X_n) = 1$$

وعليه فأن احتمال أتخاذ المتغير المتقطع قيماً تتراوح بين a و b يحسب كالتالي

$$P(a \leq X \leq b) = \sum_{i=a}^{b} P(X_i)$$

ومن اهم التوزيعات المتقطعة هي توزيع برنولي وذو الحدين والبواسون والهندسي الزائدي وغيرها .

توزيع ذو الحدين : يعتبر توزيع ذو الحدين من التوزيعات المتقطعة المهمة ذات التطبيق الواسع في الأحصاء وهو من التوزيعات المركبة بتكرار التجربة. وعلى فرض ان عينة عشوائية ذات حجم (n) قد اخذت من توزيع برنولي ولتكن

$(X_1 , X_2 , X_3 , \ldots , X_n)$

وبما ان X تأخذ الواحد الصحيح ($X=1$) اذا وقع الحادث (عند نجاح الحادث) وتأخذ الصفر ($X=0$) عند فشله لذا يمكن حساب احتمال ظهور الحادث X من المرات في n من التجارب أو المحاولات بقانون توزيع ذو الحدين التالي :

$$\boxed{P(X) = \binom{n}{x} \cdot p^{x} \cdot q^{n-x}} \quad \text{when} \quad x = 0, 1, 2, \ldots, n \quad , \quad x \le n$$

حيث ان : p = احتمال النجاح ، q = احتمال الفشل

$\binom{n}{x}$ = عدد الحالات الممكنة لظهور الحادث X من المرات من بين n من المحاولات ويطلق عليه بالمصطلح الرياضي (بالتوافيق) Combinations

تلميح : $1! = 0! = 1$

خواص تجارب ذو الحدين :

١. نتيجة كل محاولة تصنف الى صنفين هما نجاح الحادث أو فشله .
٢. احتمال النجاح p يبقى ثابتاً في محاولة الى اخرى .
٣. ان المحاولات تتكرر بـ (n) من المرات
٤. ان تكرار هذه المحاولات مستقلة .

ان المتوسط الحسابي لتوزيع ذو الحدين يحسب من المعادلة $[\mu = E(x) = np]$

ان تباين توزيع ذو الحدين يحسب من المعادلة $[\sigma^2 = npq]$

مثال (١) : في احدى تجارب مندل الوراثية وجد بأنه عند تهجين نبات طويل نقي مع نبات قصير نقي من نباتات البزاليا فأنه ينتج نباتات طويلة وقصيرة في الجيل الثاني بنسبة ($\frac{3}{1}$) على التوالي فأذا تم اختيار (5) نباتات من الجيل الثاني فما هو : ١- احتمال الحصول على (3) نباتات طويلة ؟ ٢- المتوسط الحسابي والتباين ؟

الحل /

١) احتمال ظهور نبات طويل في الجيل الثاني هو p حيث $p = \frac{3}{4}$

احتمال ظهور نبات قصير في الجيل الثاني هو q حيث $q = \frac{1}{4}$

عدد المحاولات هو n حيث $n = 5$

عدد النباتات الطويلة هو X حيث $X = 3$

وعند استخدام المعادلة

$$P(X) = \binom{n}{x} \cdot p^{x} \cdot q^{n-x}$$

نجد ان :

$$P(3) = \binom{5}{3} \cdot \left(\frac{3}{4}\right)^3 \cdot \left(\frac{1}{4}\right)^2$$

$$P(3) = \frac{(5)(4)^2(3!)}{(3!)(2)(1)}\left(\frac{27}{64}\right)\left(\frac{1}{16}\right) = \frac{270}{1024} = 0.269$$

∴ احتمال الحصول على (3) نباتات طويل هو 0.269 أو 26.9%

ملاحظة: اذا كان المطلوب ايجاد احتمال الحصول على ثلاث نباتات طويلة على الأكثر

فأن الحل سيكون: $P(X \leq 3) = P(3) + P(2) + P(1) + P(0)$

أما اذا كان المطلوب ايجاد احتمال الحصول على ثلاث نباتات طويلة على الأقل

فأن الحل سيكون: $P(X \geq 3) = P(3) + P(4) + P(5)$

٢) $\mu = E(X) = np \Rightarrow \mu = E(3) = (5)\left(\frac{3}{4}\right) = \frac{15}{4} = 3.75$

٣) $\sigma^2 = npq = \left(\frac{15}{4}\right)\left(\frac{1}{4}\right) = \frac{15}{16} = 0.9375$

مثال (٢): في احدى بساتين البرتقال في منطقة ما كانت نسبة اصابة الثمار بمرض معين هي (15%) فأذا قطفت اربع برتقالات عشوائياً من هذا البستان فما هو احتمال:

١) أن تكون واحدة مصابة فقط؟ ٢) ان تكون هناك برتقالتان مصابتان على الأقل؟

الحل/

١) بما ان احتمال الأصابة هو (P) حيث $P = 15\% = 0.15$

∴ احتمال عدم الأصابة هو (q) حيث $q = 1 - P = 1 - 0.15 = 0.85$

$X = 1$ ، $n = 4$

وعند استخدام المعادلة

$$P(X) = \binom{n}{X} \cdot P^X \cdot q^{n-X}$$

نجد أن

$$P(1) = \binom{4}{1}(0.15)^1(0.85)^3$$

$$P(1) = \left(\frac{4}{1}\right)(0.15)(0.614125) = 0.3685$$

٢) $P(X \geq 2) = P(2) + P(3) + P(4)$

$$= \binom{4}{2}(0.15)^2(0.85)^2 + \binom{4}{3}(0.15)^3(0.85)^1 + \binom{4}{4}(0.15)^4(0.85)^0$$

$$= \frac{(4)(3)}{(2)(1)}(0.15)^2(0.85)^2 + \frac{(4)(3)(2)}{(3)(2)(1)}(0.15)^3(0.85)^1 + \frac{4!}{4!}(0.15)^4(1)$$

$$= 0.0975 + 0.0114 + 0.0005$$

$$= 0.1094$$

التوزيع المعتدل أو الطبيعي : يعتبر التوزيع المعتدل أو الطبيعي من اهم التوزيعات المستمرة على الأطلاق لما له من اهمية كبيرة في الأختبارات الأحصائية وترجع أهمية التوزيع المعتدل الى خاصية ما يسمى (بنظرية النهاية المركزية) والتي تنص على ان توزيع المعاينة للمتوسط الحسابي يقترب من التوزيع الطبيعي كلما كان حجم العينة كبيراً. لذا بالأمكان استخدام التوزيع المعتدل كتوزيع تقريبي جيد لعدة توزيعات غير مستمرة كتوزيع ذي الحدين. ويطلق احياناً على التوزيع المعتدل بتوزيع (كاوس) أو بتوزيع (كاوس - لابلاس).

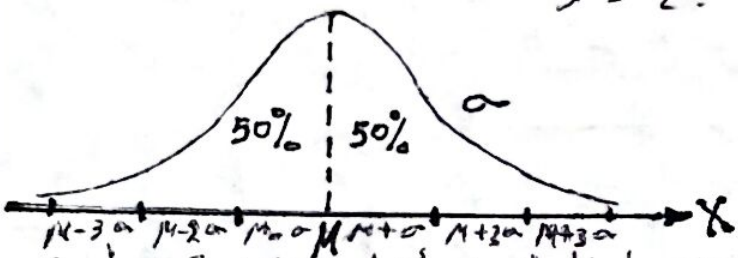
ويعبر عن التوزيع الأحتمالي المعتدل بالمعادلة التالية:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \quad , -\infty < x < \infty$$

علماً ان $\pi = 3.14159$ ، $e = 2.71828$

μ = المتوسط الحسابي للمجتمع ، σ = الأنحراف المعياري للمجتمع

ومن المعادلة اعلاه يتضح بأن شكل المنحني يختلف من مجتمع لآخر لان قيمة μ ، σ يحددان موقع وشكل المنحني وانهما يختلفان من مجتمع لآخر.



الصفات العامة للتوزيع المعتدل :

١. أن شكل المنحني هو على هيئة ناقوس.
٢. ان المساحة الكلية تحت المنحني المعتدل تساوي الواحد الصحيح.
٣. ان أعلى ارتفاع للمنحني المعتدل هو عند قيمة المتوسط الحسابي له وأن لهذا المنحني قمة واحدة.
٤. يتميز المنحني المعتدل بتماثله حول متوسط الحسابي.
٥. ان المتوسط الحسابي والوسيط والمنوال للمنحني المعتدل متساوية.
٦. ان المساحة الواقعة على يمين المتوسط الحسابي للمنحني المعتدل تساوي 50% من المساحة الكلية وهذا يعني ان القيم التي تزيد على المتوسط الحسابي تشكل نصف المجموع الكلي للقيم وينطبق القول نفسه على القيم التي تقل عن المتوسط الحسابي الى يساره.
٧. ان المساحة بين $(\mu-\sigma)$ و $(\mu+\sigma)$ تساوي 68.27% أي القيم التي تقع بين الحدين المذكورين تشكل نسبة 0.6827 من المجموع الكلي للقيم.
٨. ان المساحة بين $(\mu-2\sigma)$ و $(\mu+2\sigma)$ تساوي 95.45%.
٩. ان المساحة بين $(\mu-3\sigma)$ و $(\mu+3\sigma)$ تساوي 99.73%

مثال : لو علمت ان الوزن الصافي للبزاليا داخل علبة من حجم معين لأنتاج معمل معين متوزع توزيعاً معتدلاً بالثوابت التالية:

المتوسط الحسابي $\mu = 250$ غم ، الأنحراف المعياري $\sigma = 10$ غم

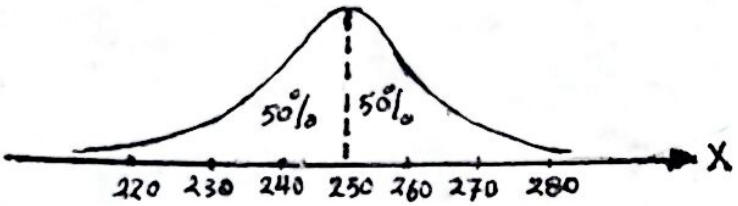
والمطلوب هو : ١- استخدام معلومات المثال لتوضيح مميزات التوزيع المعتدل.

٢- ما هو احتمال كون الوزن الصافي للبزاليا في علبة مختارة عشوائياً اكثر من 250 غم

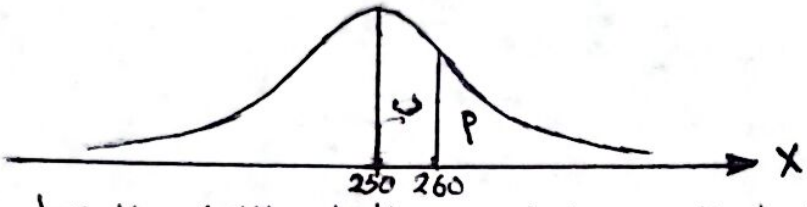
٣- ما هو احتمال كون الوزن الصافي للبزاليا في علبة مختارة عشوائياً يتراوح بين 230 و 260 غم ؟

٤- ما هو احتمال كون الوزن الصافي للبزاليا في علبة مختارة عشوائياً يتراوح بين 250 و 255 غم ؟

الحل/ ١) اذا رمزنا لمتغير الوزن الصافي بالحرف X فأن أهم صفات التوزيع المعتدل ممثلة في الشكل التالي مع مراعاة $\mu = 250$ ، $\sigma = 10$



٢) ان المساحة (أ) في الشكل التالي تمثل الأحتمال المطلوب



∵ الصفة المذكورة في البند (٦) للصفات العامة للتوزيع المعتدل التي تشير الى أن المساحة الواقعة على يمين المتوسط الحسابي للمنحني المعتدل هي 50% من المساحة الكلية فهذا يعني أن :

المساحة (أ) = 0.50 − المساحة ب

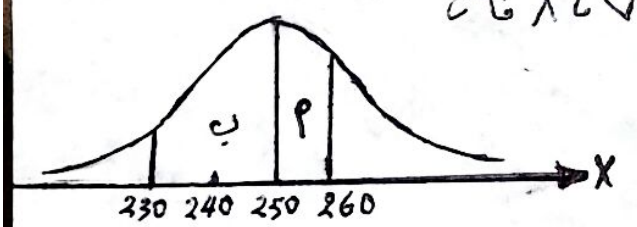
ومن الصفة المذكورة في البند (٧) المساحة الواقعة بين $\mu + \sigma$ وهي 260 وبين $\mu - \sigma$ وهي 240 تساوي 0.6827 من المجموع الكلي

∴ مساحة (ب) = $(\frac{1}{2})(0.6827) = 0.3414$

وعليه فأن الأحتمال المطلوب هو مساحة (أ) = $0.50 - 0.3414$

$= 0.1586$

٣) ان مجموع المساحتين أ و ب من الشكل التالي يمثل الأحتمال المطلوب وبالاستعانة من الصفتين المذكورتين في البندين ٧ ، ٨ فأن



المساحة (أ) = $(\frac{1}{2})(0.6827) = 0.3414$

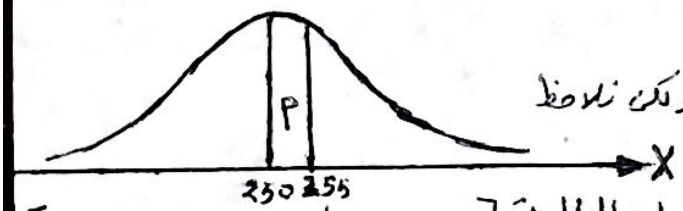
المساحة (ب) = $(\frac{1}{2})(0.9545) = 0.4773$

∴ الاحتمال المطلوب = مساحة أ + مساحة ب

$0.4773 + 0.3414 =$

$= 0.8187$

٤) المساحة (أ) من الشكل التالي تمثل الأحتمال المطلوب ولكن نلاحظ ان نجد 255 عن 250 لا يساوي الأنحراف المعياري 10 لذا نستخدم التكامل التالي لأيجاد المساحة المطلوبة



$$\left[P(a<x<b) = \int_a^b f(x)\,dx\right]$$