بسم الله الرحمن الرحيم

١. مقدمة في علم الإحصاء والاحتمالات

(١-١) علم الإحصاء: Statistics

علم الإحصاء هو ذلك الفرع من المعرفة المكون من مجموعة من الطرق الرياضية التي تستخدم لجمع وتبويب وتنظيم وعرض وتلخيص المعلومات والبيانات ومن ثم تحليلها وفق طرق وأساليب علمية للحصول على استدلالات وقرارات سليمة وذلك بالإجابة على أسئلة بحثية أو التحقق من فروض مسبقة متعلقة بموضوع البحث والدراسة. ومن فروع علم الإحصاء نذكر ما يلي:

أولاً: الإحصاء الوصفي: Descriptive Statistics

الإحصاء الوصفي هو الفرع المتعلق بطرق جمع وتبويب وتنظيم وعرض وتلخيص المعلومات والبيانات ووصف توزيع البيانات وذلك باستخدام جداول تكرارية أو رسوم بيانية وكذلك إيجاد بعض المقاييس العددية أو الوصفية التي تصف توزيع البيانات.

ثانياً: الإحصاء الاستدلالي: Statistical Inference

الإحصاء الاستدلالي هو الفرع المتعلق بطرق اتخاذ القرارات حول المجتمع تحت الدراسة وذلك عن طريق دراسة العينة. وهذا يهدف إلى إيجاد تقديرات لمعالم مجهولة أو الإجابة عن بعض الأسئلة البحثية أو التحقق من بعض الفروض المسبقة حول هذه المعالم المجهولة.

(١-٢) المجتمع والعينة:

المجتمع: Population

المجتمع هو المجموعة الكلية من الأشياء (تسمى عناصر المجتمع) ذات خصائص مشتركة ويكون الباحث مهتمًا بها حيث سيتم عمل بعض الاستدلالات والنتائج حولها. والمجتمع يجب أن يكون محددًا تحديدًا دقيقًا لدى الباحث. ويسمى عدد أفراد المجتمع بحجم المجتمع ويرمز له عادة بالرمز N. والمجتمع إما أن يكون محدودًا أو غير محدود:

١. **المجتمع المحدود:** عدد عناصر هذا المجتمع يكون محدودًا أو منتهيًا مثل مجتمع طلاب كلية العلوم بجامعة الملك سعود.

٢. **المجتمع غير المحدود:** عدد عناصر هذا المجتمع يكون غير محدود أو غير منته مثل المجتمع المكون من النجوم في السماء.

<u>العينة: Sample</u>

العينة هي مجموعة مكونة من عدد من أفراد المجتمع يتم اختيارهم بطريقة مناسبة بحيث تمثل المجتمع تمثيلاً جيدًا وذلك لدراسة صفات المجتمع إذ يستدل بصفات العينة على صفات المجتمع. ويتم جمع معلومات وبيانات الدراسة من خلال العينة. ولكي تكون العينة ممثلة للمجتمع تمثيلا جيدًا فإنه لابد أن تختار العينة وفق أساليب المعاينة الإحصائية.

<u>حجم العينة: Sample Size</u>

إن عدد أفراد العينة يسمى حجم العينة ويرمز له عادة بالرمز n.

<u>(١-٣) المتغير: Variable</u>

المتغير هو مقدار كمي أو وصفي ويستخدم لقياس خاصية (أو مميزة) معينة لعناصر المجتمع أو العينة. يتم قياس المتغير على أفراد المجتمع أو العينة ويتغير من فرد إلى آخر. ومن أمثلة المتغيرات:

- طول الشخص
- عدد أولاد الشخص
- فصيلة الدم للشخص
- المستوى التعليمي للشخص

<u>أنواع المتغيرات: Types of Variables</u>

يمكن أن تصنف المتغيرات إلى فئتين رئيسيتين على النحو التالي:

١. متغير كمي Quantitative Variable

إن قيم المتغير الكمي عبارة عن أرقام أو قيم عددية تدل على كميات أو أعداد أشياء معينة لأفراد المجتمع. ومن أمثلة المتغيرات الكمية ما يلي:

- طول الشخص
- عدد أولاد الشخص

٢. متغير وصفي (أو نوعي) Qualitative Variable

إن قيم المتغير الوصفي (أو النوعي) عبارة عن صفات أو كلمات تدل فقط على انتماء أفراد المجتمع إلى فئات أو أصناف معينة. ومن أمثلة المتغيرات الوصفية ما يلي:

- فصيلة الدم للشخص

- المستوى التعليمي للشخص

(١-٤) البيانات: Data

البيانات هي مجموعة المشاهدات أو الملاحظات المأخوذة من الدراسة الإحصائية. وتنقسم البيانات إلى نوعين تبعًا لنوع المتغير:

١. بيانات كمية (مثل: بيانات أطوال مجموعة من الأشخاص)

٢. بيانات وصفية (مثل: بيانات المستوى التعليمي لمجموعة من الأشخاص)

(١-٥) المعلمة (Parameter) والإحصاءة (Statistic):

المعلمة:

هي مقدار (أو خاصية) يتميز بها المجتمع (مثل متوسط أعمار طلاب كلية العلوم)

الإحصاءة:

هي مقدار (أو خاصية) تتميز بها العينة (مثل متوسط أعمار عشرة طلاب من طلاب كلية العلوم).

- المعلمة تكون عادة مجهولة وأما الإحصاءة فتكون معلومة حيث يمكن حسابها ومعرفتها من العينة. ولذلك فإن الإحصاءة المعلومة تستخدم كقيمة تقريبية (تقديرية) للمعلمة المجهولة.

٢. تنظيم البيانات وعرضها

(٢-١) تنظيم البيانات وتلخيصها وعرضها جدوليًا:

في هذه الفقرة سنتطرق إلى عدة مفاهيم هي:

- الجداول التكرارية للبيانات الوصفية (النوعية)
- الجداول التكرارية للبيانات الكمية
- الجداول التكرارية النسبية والجداول التكرارية المئوية
- مراكز الفترات (الفئات)
- الجداول التكرارية المتجمعة الصاعدة

لتصميم جدول التوزيع التكراري للبيانات فإنه يلزمنا معرفة أن البيانات الإحصائية تنقسم إلى نوعين هما:

بيانات وصفية مثل: لون الشعر – فصيلة الدم – الجنس – المستوى التعليمي وغيرها

بيانات كمية مثل: الطول – الوزن – العمر – عدد الاولاد وغيرها

* والبيانات الإحصائية سواءً كانت وصفية أم كمية فهي تنظم وتلخص في جداول تسمى الجداول التكرارية (أو جداول التوزيع التكرارية).

* والجدول التكراري عبارة عن جدول يلخص البيانات فيوزعها على فئات (أو طوائف أو فترات) ويحدد عدد البيانات التي تنتمي لكل فئة.

* عدد البيانات التي تنتمي إلى فئة معينة يسمى بتكرار تلك الفئة.

(٢-١-١) الجدول التكراري للبيانات الوصفية:

لإنشاء الجدول التكراري للبيانات الوصفية فإننا نقوم بعمل الخطوات التالية:

- نحصر جميع الفئات أو الطوائف وهي الصفات المختلفة في البيانات
- ننشئ جدول تفريغ البيانات وهو عبارة عن جدول مكون من ثلاثة أعمدة هي: الفئة (أو الصفة) – العلامات – التكرار
- نوجد الجدول التكراري من جدول تفريغ البيانات وذلك بحذف عمود العلامات

مثال (٢-١):

نريد إيجاد الجدول التكراري لتقديرات عينة من ستين طالبًا معطاة فيما يلي:

D	B	E	C	D	B	D	C	E	A	D	B	C	C	C
B	E	C	D	B	D	D	A	E	C	A	D	E	C	C
C	D	A	C	E	D	C	C	D	B	B	C	D	E	D
D	E	D	D	A	D	D	C	D	C	D	D	B	D	A

الحل:

- المتغير = تقدير الطالب (متغير وصفي / نوعي)
- حجم العينة = عدد البيانات = n = 60
- البيانات عبارة عن تقديرات الطلاب وهي بيانات وصفية
- القيم أو الصفات المختلفة للبيانات هي: A, B, C, D, E

نقوم بتلخيص البيانات السابقة في توزيع تكراري وفق الخطوات التالية:

١. جدول تفريغ البيانات:

الفئة أو الصفة (التقدير)	العلامات	التكرار (f) (عدد الطلاب)
A	~~IIII~~ I	6
B	~~IIII~~ III	8
C	~~IIII~~ ~~IIII~~ ~~IIII~~ I	16
D	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ II	22
E	~~IIII~~ III	8

٢. جدول التوزيع التكراري لتقديرات الستين طالبًا:

الفئة أو الصفة (التقدير)	التكرار (f) (عدد الطلاب)
A	6
B	8
C	16
D	22
E	8
المجموع	n = 60

* ملاحظة: مجموع التكرارات = عدد البيانات = عدد الطلاب = n = 60.

الجدول التكراري النسبي والجدول التكراري المئوي:

نعرف التكرار النسبي والتكرار المئوي للفئة كما يلي:

- التكرار النسبي للفئة = تكرار الفئة ÷ عدد البيانات

= تكرار الفئة ÷ مجموع التكرارات

- التكرار المئوي للفترة = التكرار النسبي × ١٠٠%

مثال:

نوجد الجدول التكراري النسبي والجدول التكراري المئوي في مثال (٢-١) كما يلي:

الفئة أو الصفة (التقدير)	التكرار (f)	التكرار النسبي= f /n	التكرار المئوي= (f /n) * 100%
A	6	6/60 = 0.100	10.0%
B	8	8/60 = 0.133	13.3%
C	16	16/60 = 0.267	26.7%
D	22	22/60 = 0.367	36.7%
E	8	8/60 = 0.133	13.3%
المجموع	n = 60	1.000	100%

ملاحظة:

- مجموع التكرارات = عدد البيانات = n
- مجموع التكرارات النسبية = 1.00
- مجموع التكرارات المئوية = 100%

(٢-١-٢) الجدول التكراري للبيانات الكمية:

لإنشاء الجدول التكراري للبيانات الكمية فإننا

- نختار بطريقة مناسبة الفترات (الفئات) المختلفة لقيم المتغير وهي عبارة عن فترات على خط الأعداد الحقيقية. وينبغي ملاحظة ما يلي عند تحديد الفترات:
  ١. الفترات غير متداخلة: لا يمكن لمشاهدة أن تصنف في أكثر من فترة
  ٢. الفترات شاملة للبيانات: كل مشاهدة لابد أن تصنف في إحدى الفترات
  ٣. أصغر قيمة لابد أن تقع في الفترة الأولى (الفترة الدنيا) وأكبر قيمة لابد أن تقع في الفترة الأخيرة (الفترة العليا).
  ٤. لايوجد طريقة واحدة لاختيار الفترات إذ أن اختيار الفترات يحدده الباحث بطريقة مناسبة لبحثه.
  ٥. بداية الفترة يسمى بالحد الأدنى للفترة ونهاية الفترة يسمى بالحد الأعلى للفترة
- نقوم بتوزيع البيانات على هذه الفترات ثم نقوم بحصر أو تحديد عدد البيانات الواقعة في كل فترة من الفترات

- ننشئ جدول تفريغ البيانات وهو عبارة عن جدول مكون من ثلاثة أعمدة هي: الفترات – العلامات – التكرار
- نوجد الجدول التكراري من جدول تفريغ البيانات وذلك بحذف عمود العلامات

مثال (٢-٢):

في أحد البحوث التي أجريت لدراسة مستوى الهيموجلوبين قام الباحث باختيار عينة مكونة من خمسين شخصا فحصل على البيانات التالية:

17.0	17.7	15.9	16.2	16.2	17.1	15.7	17.3	**13.5**	16.3
14.6	15.8	15.3	16.4	13.7	16.2	16.4	16.1	17.0	15.9
14.0	16.2	16.4	14.9	17.8	16.1	15.5	**18.3**	15.8	16.7
15.9	15.3	13.9	16.8	15.9	16.3	17.4	15.0	17.5	15.1
14.2	16.1	15.7	15.1	17.4	16.5	14.4	16.3	17.3	15.8

المطلوب هو تكوين التوزيع التكراري لبيانات مستوى الهيموجلوبين لهؤلاء الأشخاص باستخدام الفترات التالية:

12.95 – 13.95   13.9 5– 14.95   14. 95– 15.95   15. 95– 16.95   16. 95– 17.95   17. 95– 18.95

الحل:

- المتغير = مستوى الهيموجلوبين (متغير كمي)
- حجم العينة = عدد البيانات = n = 50
- البيانات عبارة عن قيم مستوى الهيموجلوبين وهي بيانات كمية
- أكبر قيمة= 18.3 وأصغر قيمة = 13.5

نقوم بتلخيص البيانات السابقة في توزيع تكراري وفق الخطوات التالية:

١. جدول تفريغ البيانات:

مستوى الهيموجلوبين (فترة الفئة)	العلامات	التكرار (f) (عدد الأشخاص)
12.95 – 13.95	|||	3
13.9 5– 14.95	~~||||~~	5
14. 95– 15.95	~~||||~~ ~~||||~~ ~~||||~~	15
15. 95– 16.95	~~||||~~ ~~||||~~ ~~||||~~ |	16
16. 95– 17.95	~~||||~~ ~~||||~~	10
17. 95– 18.95	|	1

٢. جدول التكراري لمستوى الهيموجلوبين:

مستوى الهيموجلوبين (فترة الفئة)	التكرار (f) (عدد الأشخاص)
12.95 – 13.95	3
13.9 5– 14.95	5
14. 95– 15.95	15
15. 95– 16.95	16
16. 95– 17.95	10
17. 95– 18.95	1
المجموع	n = 50

(٢-١-٣) الجدول التكراري النسبي والجدول التكراري المئوي:

نعرف التكرار النسبي والتكرار المئوي للفترة كما يلي:

- التكرار النسبي للفترة = تكرار الفترة ÷ عدد البيانات

  = تكرار الفترة ÷ مجموع التكرارات

- التكرار المئوي للفترة = التكرار النسبي × ١٠٠%

مثال:

نوجد الجدول التكراري النسبي والجدول التكراري المئوي في مثال (٢-٢) كما يلي:

مستوى الهيموجلوبين (فترة الفئة)	التكرار (f)	التكرار النسبي= f /n	التكرار المئوي= (f /n) * 100%
12.95 – 13.95	3	3/50= 0.06	6%
13.9 5– 14.95	5	5/50= 0.10	10%
14. 95– 15.95	15	15/50= 0.30	30%
15. 95– 16.95	16	16/50= 0.32	32%
16. 95– 17.95	10	10/50= 0.20	20%
17. 95– 18.95	1	1/50= 0.02	2%
المجموع	n = 50	1.00	100%

(٢-١-٤) مركز الفترة وطول الفترة:

نعرف مركز الفترة وطول الفترة كما يلي:

- مركز الفترة = منتصف الفترة. أي أن:

مركز الفترة = الحد الأعلى + الحد الأدنى

2

- طول الفترة (L) = الفرق بين الحد الأعلى والحد الأدنى للفترة
  = الحد الأعلى للفترة مطروحًا منه الحد الأدنى للفترة

مثال:

فيما يلي التوزيع التكراري متضمنا مراكز الفترات لمستوى الهيموجلوبين في مثال (٢-٢):

مستوى الهيموجلوبين (فترة الفئة)	مركز الفترة (m)	التكرار (f)
12.95 – 13.95	13.45	3
13.9 5– 14.95	14.45	5
14. 95– 15.95	15.45	15
15. 95– 16.95	16.45	16
16. 95– 17.95	17.45	10
17. 95– 18.95	18.45	1

طول الفترة لهذا الجدول التكراري هو:

$$L = 13.95 - 12.95 = 1.00$$

ملاحظات:

- مركز الفترة = مركز الفترة السابقة + طول الفترة
- إن عدد البيانات (التكرار) الواقعة في كل فترة يكون معروفا للتوزيع التكراري. ونظرًا لأن البيانات الأصلية في كل فترة غير معروفة فإن مركز الفترة يستخدم كقيمة تقريبية لجميع البيانات الواقعة في تلك الفترة. أي أن مركز الفترة يعتبر قيمة مثالية أو نموذجية لجميع البيانات الواقعة في تلك الفترة.

(٢-١-٥) الجدول التكراري المتجمع الصاعد:

نعرف التكرار المتجمع الصاعد للفترة كما يلي:

- التكرار المتجمع الصاعد للفترة = عدد البيانات التي تقل عن أو تساوي الحد الأعلى للفترة
  = تكرار الفترة + مجموع تكرارات الفترات السابقة لها.

يستخدم الجدول التكراري المتجمع الصاعد لإيجاد عدد البيانات التي تكون قيمتها أصغر من أو يساوي قيمة معينة. كما يستخدم لإيجاد عدد البيانات التي تكون قيمتها محصورة بين قيمتين معينتين.

مثال:

على سبيل المثال في مثال (٢-٢):

عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي 12.95 = 0

عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي 13.95 = 3

عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي 14.95 =3+5= 8

عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي 15.95 =3+5+15= 23

عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي 16.95 =3+5+15+16= 39

عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي 17.95 = 3+5+15+16+10 = 49

عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي18.95=3+5+15+16+10+1= 50

إذا وضعنا هذه المعلومات في جدول فإننا نحصل على الجدول التكراري المتجمع الصاعد كما يلي:

التكرار المتجمع الصاعد	مستوى الهيموجلوبين
0	أقل من 12.95
3	أقل من 13.9 5
8	أقل من 14. 95
23	أقل من 15. 95
39	أقل من 16. 95
49	أقل من 17. 95
50 = n	أقل من 18.95

من هذا الجدول نستطيع استخراج كثير من المعلومات، فعلى سبيل المثال:

١. عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي 16.95 هو 39 شخصًا.

٢. عدد الأشخاص الذين مستوى الهيموجلوبين لهم يتراوح بين 14.95 و 16.95 هو 31 شخصًا حيث يحسب كما يلي: 31 = 8 – 39.

(٢-٢) العرض البياني للجداول التكرارية:

في هذه الفقرة سوف نستعرض التمثيل أو العرض البياني للجدول التكراري والجدول التكـــراري المتجمع الصاعد.

(٢-٢-١) عرض وتمثيل الجدول التكراري بيانيًا:

لعرض وتمثل الجدول التكراري أو الجدول التكراري النسبي أو الجدول التكراري المئوي بيانيًـــا فإننا نستخدم الرسوم التالية:

- **المدرج التكراري:**

  وهو عبارة عن رسم بياني مكون من محورين. المحور الأفقـــي يمثـــل الفترات للمتغير والمحور العمودي يمثل التكرار. يـــتم رســـم المـــدرج التكراري برسم مستطيلات متلاصقة قواعـــدها عبـــارة عـــن الفتـــرات وارتفاعاتها عبارة عن التكرارات المقابلة.

- **المضلع التكراري:**

  وهو عبارة عن رسم بياني مكون من محورين. المحور الأفقـــي يمثـــل مراكز الفترات والمحور العمودي يمثل التكـــرار. يـــتم رســـم المـــضلع التكراري برسم نقاط فوق كل مركز فترة. ارتفاع النقطـــة عـــن مركـــز الفترة عبارة عن تكرار تلك الفترة. وبعد رسم جميع النقاط يتم توصـــيلها بخطوط مستقيمة ومن ثم غلق المضلع التكراري لكي تكـــون المـــساحة تحت المنحنى التكراري مساوية لمساحات المـــستطيلات فـــي المـــدرج التكراري.

(٢-٢-٢) عرض وتمثيل الجدول التكراري المتجمع الصاعد بيانيًا:

لعرض وتمثل التوزيع التكراري المتجمع الصاعد بيانيًا فإننا نستخدم الرسم التالي:

- **المضلع التكراري المتجمع الصاعد:**

  وهو عبارة عن رسم بياني مكون من محورين. المحور الأفقي يمثل الفترات والمحور العمودي يمثل التكرار المتجمع الصاعد. يتم رسم المضلع التكراري المتجمع برسم نقاط فوق حدود الفترات. ارتفاع النقطة عن حد الفترة عبـــارة عن التكرار المتجمع الصاعد عند ذلك الحد. وبعد رسم جميـــع النقـــاط يـــتم توصيلها بخطوط مستقيمة.

مثال:

فيما يلي الجدول التكراري والجدول التكراري المتجمع الصاعد لمستوى الهيموجلوبين لخمسين شخصًا في مثال (٢-٢).

١. مثل الجدول التكراري بيانيا باستخدام المدرج التكراري و المضلع التكراري.

٢. مثل الجدول التكراري المتجمع الصاعد بيانيا باستخدام المضلع التكراري المتجمع الصاعد.

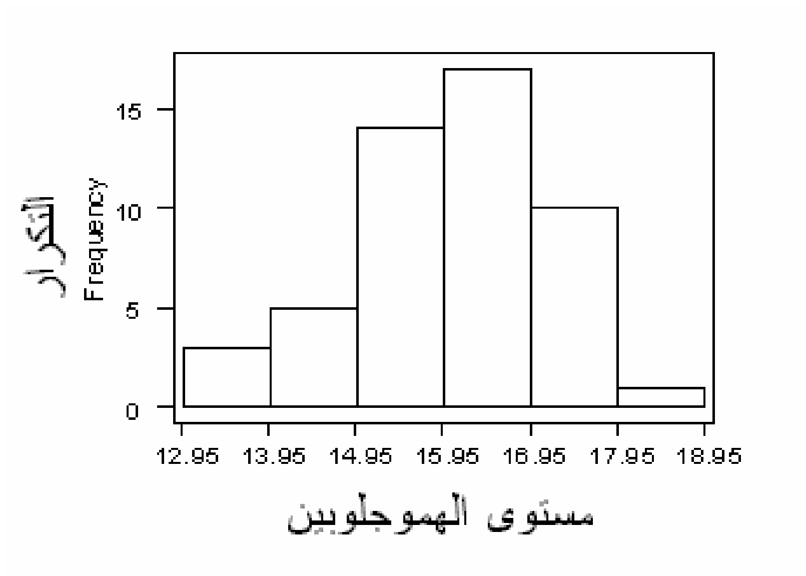
المطلوب هو تمثيل الجدول التكراري والجدول التكراري المتجمع الصاعد بيانيًا.

مستوى الهيموجلوبين	مركز الفترة	التكرار
12.95 – 13.95	13.45	3
13.95– 14.95	14.45	5
14. 95– 15.95	15.45	15
15. 95– 16.95	16.45	16
16. 95– 17.95	17.45	10
17. 95– 18.95	18.45	1

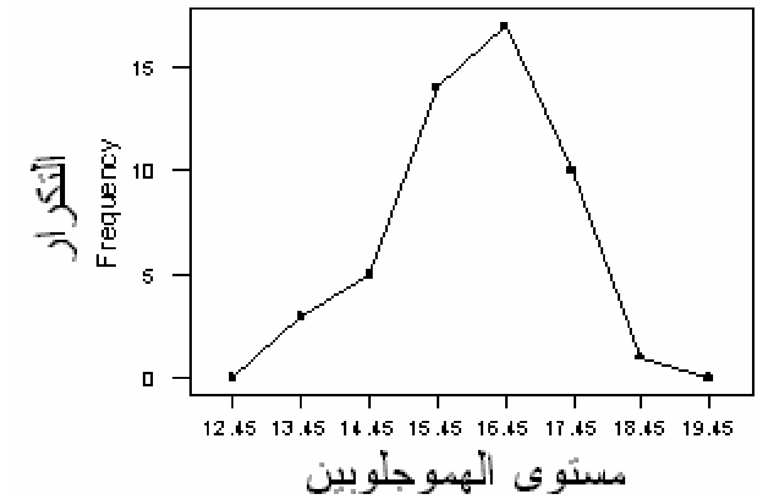
مستوى الهيموجلوبين	التكرار المتجمع الصاعد
أقل من 12.95	0
أقل من 13.9 5	3
أقل من 14. 95	8
أقل من 15. 95	23
أقل من 16. 95	39
أقل من 17. 95	49
أقل من 18.95	50

الحل:

(١) تمثيل الجدول التكراري بيانيا باستخدام المدرج التكراري و المضلع التكراري:

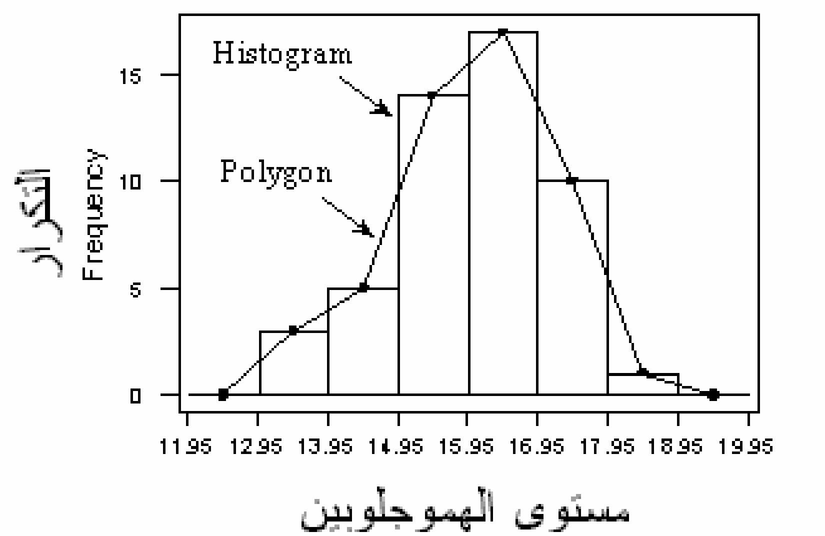


المدرج التكراري

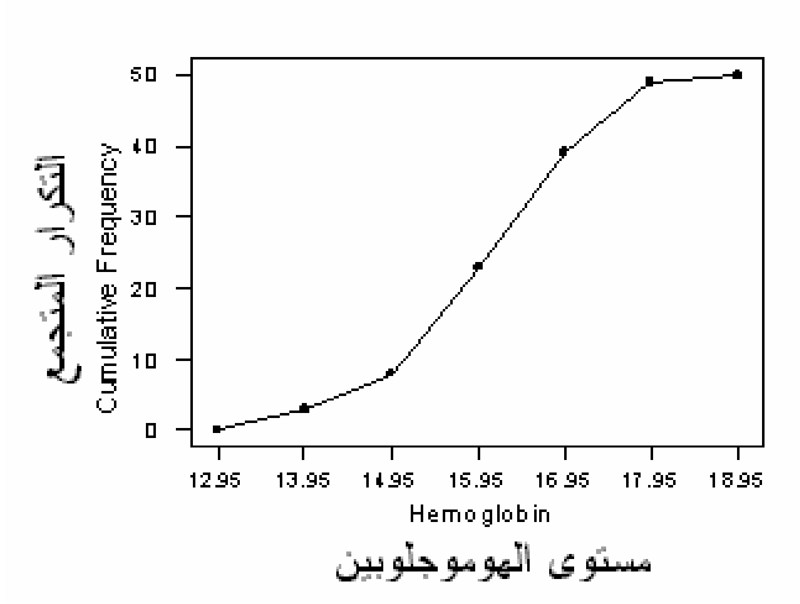


المضلع التكراري

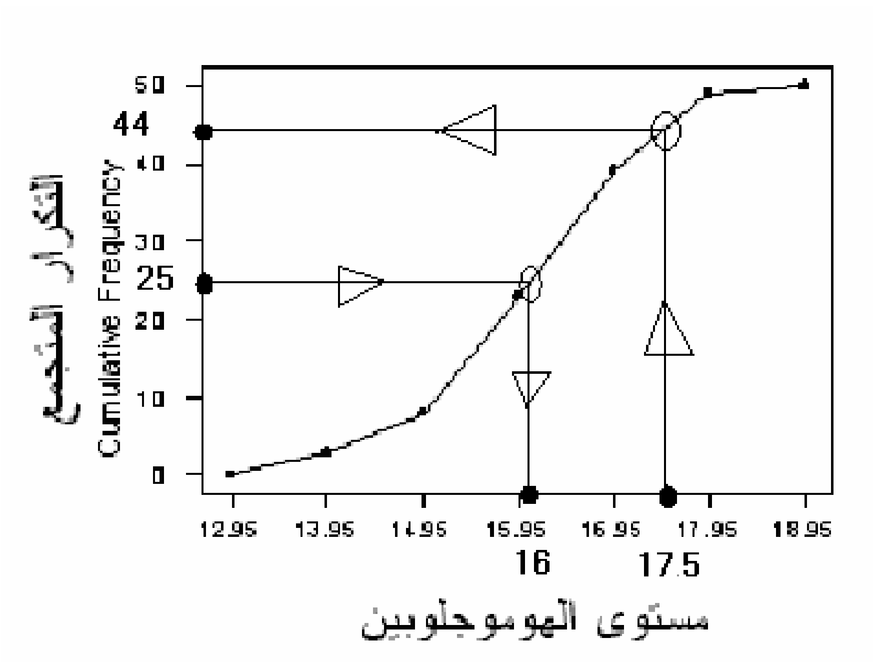
(٢) تمثيل الجدول التكراري المتجمع الصاعد بيانيا باستخدام المضلع التكراري المتجمع الصاعد:



المدرج التكراري والمضلع التكراري معًا



المضلع التكراري المتجمع الصاعد



- من المضلع التكراري المتجمع الصاعد نجد أن عدد الأشخاص الذين مستوى الهيموجلوبين لهم يقل عن أو يساوي 17.5 هو 44 شخصًا.
- مستوى هيموجلوبين نصف الأشخاص (25 شخصًا) يقل عن أو يساوي 16.

٣. مقاييس النزعة المركزية (مقاييس الموضع)
Measures of Central Tendency (Location Measures)

(٣-١) مقدمة:

مقاييس النزعة المركزية هي مقاييس عددية تستخدم لقياس موضع تركز أو تجمع البيانات. إذ أن بيانات أي ظاهر تنزع في الغالب إلى التركز والتجمع حول قيم معينة. هذه القيم هي ما يسمى بمقاييس النزعة المركزية. ومقاييس النزعة المركزية تستخدم لتلخيص البيانات عدديًا إذ أنها تعتبر قيم نموذجية أو مثالية للبيانات. كما أن هذه المقاييس تستخدم لوصف مجموعة البيانات وكذلك لمقارنة مجموعات البيانات المختلفة. ومن أهم هذه المقاييس نذكر: الوسط الحسابي (أو المتوسط)، الوسط الموزون (أو المرجح)، الوسيط، والمنوال.

تعريف رمز التجميع:

إذا كان عدد البيانات هو n وكانت البيانات هي $x_1, x_2, \ldots, x_n$ فإن مجموع هذه البيانات هو:

$$\boxed{\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n}$$

ويتمتع التجميع بالخواص التالية:

- $\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$
- $\sum_{i=1}^{n} c\,x_i = c \sum_{i=1}^{n} x_i$
- $\sum_{i=1}^{n} c = n\,c$

(٣-٢) الوسط الحسابي (المتوسط): Arithmetic Mean (Mean)

يعتبر المتوسط من أهم وأفضل مقاييس النزعة المركزية ومن أكثرها شيوعًا واستخدامًا في التحليل الإحصائي وذلك لما يتمتع به من خصائص وصفات إحصائية جيدة. ولإيجاد المتوسط للبيانات فإننا لابد أن نفرق بين البيانات المفردة (غير مبوبة في جدول تكراري) والبيانات المبوبة (الملخصة في جدول تكراري).

أولاً: المتوسط للبيانات المفردة (غير المبوبة):

إذا كان عدد البيانات (حجم العينة) هو n وكانت قيم أو مشاهدات العينة هـــي $x_1, x_2, \ldots, x_n$ فإن المتوسط (الوسط الحسابي) يرمز له بالرمز $\bar{x}$ ويعرف بالصيغة التالية:

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

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

مثال (٣-١):

أوجد المتوسط (الوسط الحسابي) للمشاهدات التالية والتي هي عبارة عـــن أوزان (بـــالكيلوجرام) مجموعة مكونة من سبعة أشخاص: 25, 30, 40, 45, 35, 55, 50

الحل:

$x_1=25,\ x_2=30,\ x_3=40,\ x_4=45,\ x_5=35,\ x_6=55,\ x_7=50$

$n = 7$

المتوسط هو:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_7}{7}$$

$$= \frac{25+30+40+45+35+55+50}{7} = \frac{280}{7} = 40 \quad \text{(كيلوجرامًا)}$$

ثانياً: المتوسط للبيانات المبوبة:

ينبغي علينا ملاحظة ما يلي في حالة البيانات الملخصة في توزيع تكراري مبوب:

- البيانات الأصلية غير معروفة.
- عدد البيانات في كل فترة (تكرار الفترة) معروف.
- يستخدم مركز الفترة كقيمة تقريبية لجميع البيانات في الفترة.

إذا كان لدينا بيانات عددها n وكانت هذه البيانات ملخصة في جدول تكراري بحيث أن:

- عدد الفترات هو k

- مراكز الفترات هي $x_1, x_2, \ldots, x_k$
- تكرارات الفترات هي $f_1, f_2, \ldots, f_k$

أي أن البيانات قد تم تلخيصها في التوزيع التكراري المبوب التالي:

التكرار f	مركز الفترة x	الفترة
f_1	x_1	الفترة رقم 1
f_2	x_2	الفترة رقم 2
$\vdots$	$\vdots$	$\vdots$
f_k	x_k	الفترة رقم k
$n = \sum f$		المجموع

لحساب المتوسط بالطريقة الحسابية فإنه يلزمنا فقط معرفة ما يلي:

- حجم العينة = عدد البيانات = $n = \sum f$
- مجموع البيانات = $\sum x f$

لذلك فإن المتوسط للتوزيع التكراري المبوب يمكن حسابه بشكل تقريبي بالصيغة التالية:

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum xf}{n} = = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_k f_k}{f_1 + f_2 + \cdots + f_k}$$

ويكمن تلخيص عملية إيجاد المتوسط باستخدام الجدول التالي:

$x f$	التكرار f	مركز الفترة x	الفترة
$x_1 f_1$	f_1	x_1	الفترة رقم 1
$x_2 f_2$	f_2	x_2	الفترة رقم 2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_k f_k$	f_k	x_k	الفترة رقم k
$\sum x f$	$\sum f = n$		المجموع

مثال (٣-٢):

أوجد المتوسط لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (٢-٢).

الحل:

مستوى الهيموجلوبين	مركز الفترة x	التكرار f	xf
12.95 – 13.95	13.45	3	40.35
13.9 5– 14.95	14.45	5	72.25
14. 95– 15.95	15.45	15	231.75
15. 95– 16.95	16.45	16	263.20
16. 95– 17.95	17.45	10	174.50
17. 95– 18.95	18.45	1	18.45
المجموع		$n = \sum f = 50$	$\sum x f = 800.5$

المتوسط (الوسط الحسابي) هو:

$$\overline{x} = \frac{\sum x f}{\sum f} = \frac{800.5}{50} = 16.01$$

بعض خصائص الوسط الحسابي (المتوسط):

١. المجموع الجبري لانحرافات القيم عن الوسط الحسابي $\overline{x}$ يساوي الصفر. أي أن:

$$\sum_{i=1}^{n}(x_i - \overline{x}) = (x_1 - \overline{x}) + (x_2 - \overline{x}) + \ldots + (x_n - \overline{x}) = 0$$

ملاحظة: $(x_i - \overline{x})$ = انحراف القيمة x_i عن وسطها الحسابي $\overline{x}$.

٢. الوسط الحسابي (المتوسط) يخضع للعمليات الجبرية بسهولة كما يلي:

الوسط الحسابي (المتوسط)	المشاهدات
$\overline{x}$	$x_1, x_2, \ldots, x_n$
$\overline{(x \pm b)} = \overline{x} \pm b$	$x_1 \pm b , x_2 \pm b , \ldots, x_n \pm b$
$\overline{ax} = a\,\overline{x}$	$ax_1 , ax_2 , \ldots, ax_n$
$\overline{(ax \pm b)} = a\,\overline{x} \pm b$	$ax_1 \pm b , ax_2 \pm b , \ldots, ax_n \pm b$

- مثال:

الوسط الحسابي (المتوسط)	المشاهدات	
$\bar{x} = 4$	2, 6, 4, 3, 5	: x
$\bar{x} + 5 = 9$	7, 11, 9, 8, 10	: x+5
$3\bar{x} = 12$	6, 18, 12, 9, 15	:3x
$3\bar{x} + 5 = 17$	11, 23, 17, 14, 20	: 3x +5

- مثال:

إذا كان الوسط الحسابي للمشاهدات $x_1, x_2, \ldots, x_n$ هو 15 فإن الوسط الحسابي للمشاهدات $\frac{x_1-10}{2}, \frac{x_2-10}{2}, \ldots, \frac{x_n-10}{2}$ هو $\frac{\bar{x}-10}{2} = \frac{15-10}{2} = 2.5$.

٣. إذا كان لدينا مجموعتان من البيانات بحيث أن عدد بيانات المجموعـــة الأولـــى هـــو n_1 ومتوسطها هو $\bar{x}_1$ وكان عدد بيانات المجموعة الثانية هو n_2 ومتوسطها هو $\bar{x}_2$ فـــإن متوسط المجموعة الكلية المكونة من دمج هاتين المجموعتين يمكن حسابه بالصيغة التالية:

$$\text{متوسط المجموعة الكلية} \quad \bar{X} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

- مثال:

إذا كان لدينا مجموعتان من البيانات بحيث أن عدد بيانـــات المجموعـــة الأولـــى هـــو 10 ومتوسطها هو 5 وكان عدد بيانات المجموعة الثانية هو 20 ومتوسطها هو 2 فإن متوســـط المجموعة الكلية المكونة من دمج هاتين المجموعتين هو:

$$\bar{X} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \frac{10\times 5 + 20\times 2}{10+20} = \frac{90}{30} = 3$$

بعض مميزات وعيوب الوسط الحسابي (المتوسط):

- **مميزات المتوسط:** إن المتوسط يعتبر من أفضل مقاييس النزعة المركزية ومـــن أكثرهـــا شيوعًا وذلك لما يتمتع به من صفات جيدة. ومن مميزات المتوسط نذكر ما يلي:
  ١. المتوسط سهل التعريف والحساب ويخضع للعمليات الجبرية بسهولة.
  ٢. المتوسط وحيد لمجموعة البيانات الواحدة.
  ٣. يأخذ المتوسط في الاعتبار جميع البيانات.

- **عيوب المتوسط:** بالرغم من أن المتوسط يعتبر من أفضل مقاييس النزعة المركزية إلا أن له بعض العيوب نذكر منها ما يلي:

١. يتأثر المتوسط بالقيم الشاذة أو المتطرفة.

٢. المتوسط غير معرف للبيانات الوصفية (النوعية) إذ يمكن حسابه للبيانات الكمية فقط.

ملاحظة:

وحدة المتوسط هي نفس وحدة البيانات الأصلية. فإذا كانت وحدة البيانات هي الكيلوجرام فإن وحدة المتوسط هي الكيلوجرام.

(٣-٣) الوسط المرجح (الموزون): Weighted Mean

في بعض الأحيان تكون المشاهدات $x_1, x_2, \ldots, x_n$ مقرونة بالأوزان $w_1, w_2, \ldots, w_n$ على التوالي. وفي هذه الحالة نعرف الوسط المرجح كما يلي:

$$\bar{x}_w = \frac{\sum x w}{\sum w} = \; = \; = \frac{x_1 w_1 + x_2 w_2 + \cdots + x_n w_n}{w_1 + w_2 + \cdots + w_n}$$

مثال (٣-٣):

أوجد الوسط المرجح لدرجات الطلاب باعتبار أن الوزن هو عدد الساعات للمقرر فيما يلي:

المقرر	عدد الساعات (w)	الدرجة (x)
إحص	2	40
فيز	4	65
ريض	3	70

الحل:

x	w	wx
40	2	80
65	4	260
70	3	210
	$\sum w = 9$	$\sum x w = 550$

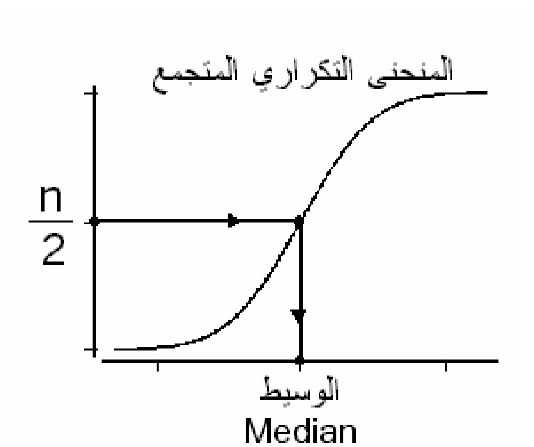
$$\bar{x}_w = \frac{\sum x w}{\sum w} = \; = \frac{550}{9} = 61.11$$ (درجة)

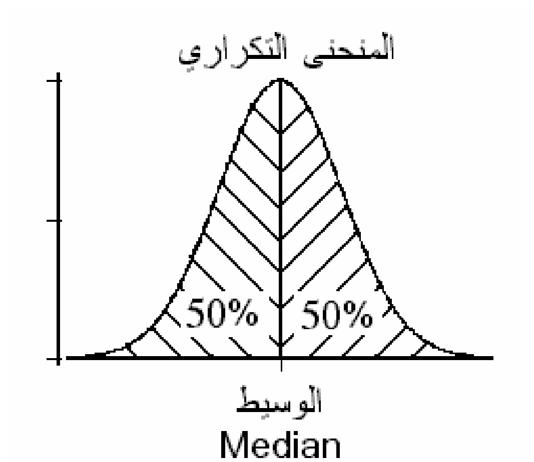
ملاحظات:

١. المعدل التراكمي والمعدل الفصلي للطالب في جامعة الملك سعود هما وسطان مرجحـــان للنقاط باعتبار أن أعداد الساعات هي الأوزان.
٢. الوسط الحسابي $\bar{x}$ هو وسط مرجح جميع أوزانه متساوية (أو مساوية للواحد).
٣. الوسط الحسابي للبيانات المبوبة هو وسط مرجح لمراكز الفتــرات باعتبــار أن تكــرار الفترات هي الأوزان.

(٣-٤) الوسيط: Median

الوسيط هو أحد مقاييس النزعة المركزية المشهورة. ويعرف الوسيط لمجموعة من البيانات على أنه تلك القيمة التي تتوسط البيانات عند ترتيبها تصاعديًا (أو تنازليًا) أي أنه تلك القيمة التي تقسم البيانات بعد ترتيبها إلى جزأين متساويين فتكون البيانات في الجزء الأول تقل عــن أو تــساوى الوسيط والبيانات في الجزء الثاني تزيد عن أو تساوى الوسيط. أي أن 50% من البيانات تساوي أو تقل عن الوسيط و 50% من البيانات تساوي أو تزيد عن الوسيط. يرمــز للوســيط بــالرمز (Med).





أولاً: الوسيط للبيانات المفردة (غير مبوبة):

إذا كانت قيم العينة هي $x_1, x_2, \ldots, x_n$ وحجم العينة هو n فإن الوسيط يعرف كما يلي:

- **أولاً: إذا كان حجم العينة n عددًا فرديًا:**

الوسيط = القيمة التي في منتصف البيانات بعد ترتيبها وهي القيمة المرتبـــة ذات الترتيب $\frac{n+1}{2}$.

البيانات مرتبة	$X_{(1)}$	$X_{(2)}$	…	$X_{(\frac{n+1}{2})}$	…	$X_{(n)}$
الترتيب	1	2	…	$\frac{n+1}{2}$	…	n.

القيمة في المنتصف = $X_{(\frac{n+1}{2})}$ = الوسيط

- **ثانيًا: إذا كان حجم العينة n عددًا زوجيًا:**

الوسيط = متوسط القيمتين في منتصف البيانات بعد ترتيبها وهما القيمتان المرتبتان ذاتا الترتيب $\frac{n}{2}$ و $\frac{n}{2}+1$.

البيانات مرتبة	$X_{(1)}$	$X_{(2)}$	…	$X_{(\frac{n}{2})}$	$X_{(\frac{n}{2}+1)}$	…	$X_{(n)}$
الترتيب	1	2	…	$\frac{n}{2}$	$\frac{n}{2}+1$	…	n.

القيمتان في المنتصف هما: $X_{(\frac{n}{2})}$ و $X_{(\frac{n}{2}+1)}$ لذلك فإن الوسيط =

$$\frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}$$

مثال (٣-٤):

أوجد الوسيط لمجموعة الأوزان (بالكيلوجرام) التالية: 8.3, 5.4, 2.5, 2.5, 7.1.

الحل:

بما أن n=5 عدد فردي فإن الوسيط هو القيمة التي في المنتصف بعد ترتيب البيانات وهي القيمة ذات الترتيب $3=\frac{5+1}{2}=\frac{n+1}{2}$.

البيانات مرتبة	2.5	2.5	5.4	7.1	8.3
الترتيب	1	2	3	4	5

الوسيط هو القيمة ذات الترتيب 3 لذلك فإن: الوسيط = 5.4 كيلوجراما

مثال (٣-٥):

أوجد الوسيط لمجموعة الأوزان (بالكيلوجرام) التالية: 8.3, 9.2, 5.4, 2.5, 2.5, 7.1.

الحل:

بما أن n=6 عدد زوجي فإن الوسيط هو متوسط القيمتين اللتين في المنتصف بعد ترتيب البيانات وهما القيمتان ذاتا الترتيب $3=\frac{6}{2}=\frac{n}{2}$ و $\frac{n}{2}+1=4$.

البيانات مرتبة	2.5	2.5	5.4	7.1	8.3	9.2
الترتيب	1	2	3	4	5	6

القيمتان في المنتصف هما 5.4 و 7.1 ولذلك فإن: الوسيط = $\frac{5.4+7.1}{2}$ = 6.25 كيلوجراما.

ثانيًا: الوسيط للبيانات المبوبة:

يمكن حساب الوسيط للبيانات الملخصة في جدول تكراري بطرقتين هما: طريقة حسابية وطريقة بيانية. ويستخدم الجدول التكراري المتجمع الصاعد لإيجاد الوسيط حسابيًا بينما يستخدم المضلع التكراري المتجمع الصاعد لإيجاد الوسيط بيانيًا. وفي حالة البيانات المبوبة فإننا نعرف ما يلي:

- رتبة (أو ترتيب) الوسيط = $\frac{n}{2}$ (سواءً كان عدد البيانات n زوجيًا أم فرديًا).
- الفترة الوسيطية = الفترة التي يقع فيها الوسيط

= أول فترة يزيد تكرارها المتجمع الصاعد عن $\frac{n}{2}$ أو يساويه

(أ) إيجاد الوسيط حسابيًا:

لإيجاد الوسيط حسابيًا نقوم بالخطوات التالية:

١. نكون الجدول التكراري المتجمع الصاعد

٢. نحدد رتبة (أو ترتيب) الوسيط = $\frac{n}{2}$ (التكرار المتجمع الوسيطي)

٣. بعد تحديد رتبة الوسيط نستخدم الجدول التكراري المتجمع الصاعد لتحديد ما يلي:

A = بداية الفترة الوسيطية   L = طول الفترة

F_1 = التكرار المتجمع الصاعد **السابق** للتكرار المتجمع الصاعد الوسيطي $\frac{n}{2}$

F_2 = التكرار المتجمع الصاعد **اللاحق** للتكرار المتجمع الصاعد الوسيطي $\frac{n}{2}$

٤. نوجد الوسيط بالعلاقة التالية:

$$Med = A + \left(\frac{\frac{n}{2} - F_1}{F_2 - F_1} \right) \times L = A + \left(\frac{(\text{رتبة الوسيط}) - F_1}{F_2 - F_1} \right) \times L$$

مثال (٣-٦):

أوجد قيمة الوسيط حسابيًا لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (٢-٢).

الحل:

التكرار المتجمع الصاعد	مستوى الهيموجلوبين
0	أقل من 12.95
3	أقل من 13.9 5
8	أقل من 14. 95
23 = F_1	أقل من **15. 95** = A
39 = F_2	أقل من 16. 95
49	أقل من 17. 95
50	أقل من 18.95

$\leftarrow \frac{n}{2} = 25$

Med $\Rightarrow$

$\frac{n}{2} = \frac{50}{2} = 25$

الفترة الوسيطية هي:

$15.95 - 16.95$

$A=15.95$

$L=16.95-15.95=1.0$

$F_1 = 23$

$F_2 = 39$

$$Med = A + \left(\frac{\frac{n}{2} - F_1}{F_2 - F_1} \right) \times L$$

$$Med = 15.95 + \left(\frac{25 - 23}{39 - 23} \right) \times 1.0 = 15.95 + \left(\frac{2}{16} \right) \times 1.0$$

$$= 15.95 + 0.125 \times 1 = 16.075$$

في هذا المثال نستطيع القول بأن 50% من الأشخاص (نصف الأشخاص) يقل مستوى الهيموجلوبين لهم عن Med=16.075.

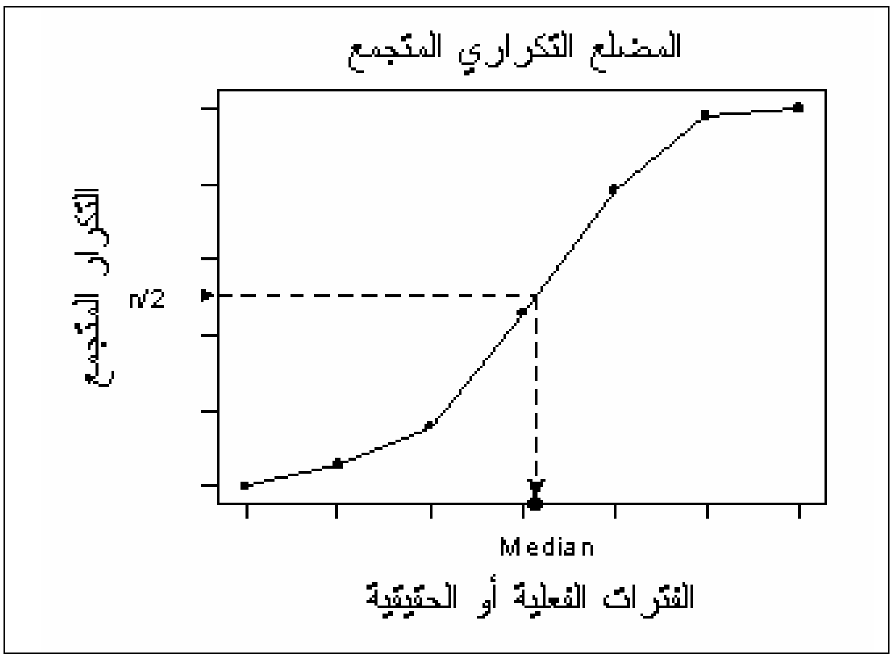
(ب) إيجاد الوسيط بيانيًا:

لإيجاد الوسيط بيانيًا نقوم بالخطوات التالية:

١. نرسم المضلع التكراري المتجمع الصاعد

٢. نحدد رتبة (أو ترتيب) الوسيط = $\frac{n}{2}$ (سواءً كان عدد البيانات فرديًا أو زوجيًا)

٣. في المضلع التكراري المتجمع الصاعد نحدد موقع رتبة الوسيط $\frac{n}{2}$ على المحور الرأسي (محور التكرار المتجمع) ومن ذلك الموقع نرسم خطًا أفقيًا يلتقي مع المضلع في نقطة. وعند نقطة الالتقاء نرسم عمودًا يتقاطع مع محور الفترات (المحور الأفقي) في نقطة. هذه النقطة هي قيمة الوسيط التقريبية. والشكل التالي يبين طريقة حساب الوسيط.



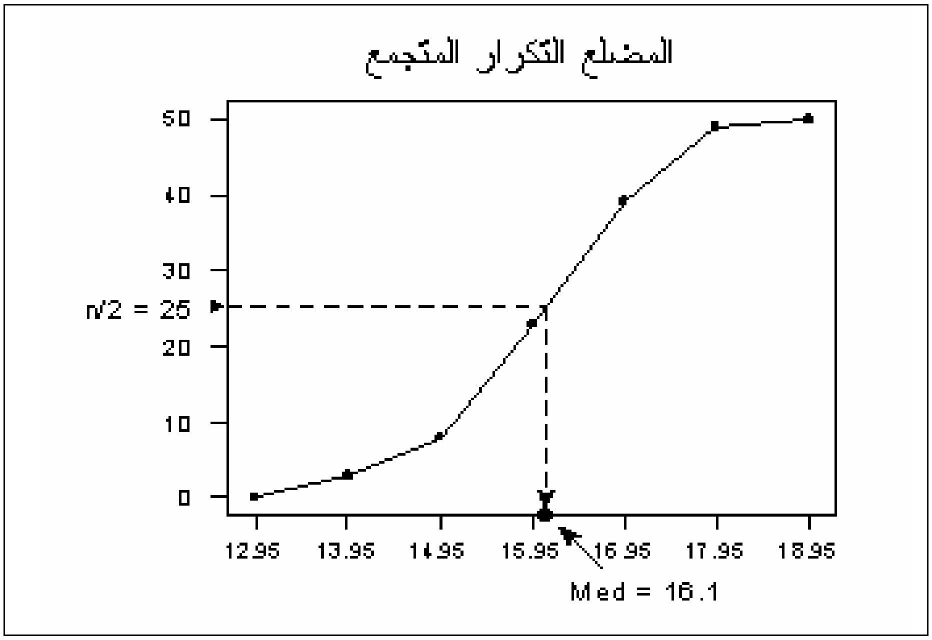
<u>مثال (٣-٧):</u>

أوجد قيمة الوسيط بيانيًا لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (٢-٢).

<u>الحل:</u>

نرسم المضلع التكراري المتجمع الصاعد كما مر معنا سابقًا ثم نطبق طريقة حساب الوسيط بيانيًا. مع ملاحظة أن رتبة الوسيط = $\frac{n}{2}=\frac{50}{2}=25$. باستخدام الشكل أدناه نجد أن القيمة التقريبية للوسيط هي:

الوسيط = 16.1.



بعض مميزات وعيوب الوسيط:

- **مميزات الوسيط:** إن الوسيط يعتبر من مقاييس النزعة المركزية الشائعة وذلك لما يتمتع به من بعض الصفات الجيدة. ومن مميزات الوسيط نذكر ما يلي:
  ١. الوسيط سهل التعريف والحساب.
  ٢. الوسيط وحيد لمجموعة البيانات الواحدة.
  ٣. الوسيط أقل تأثرًا من المتوسط بالقيم الشاذة أو المتطرفة.
- **عيوب الوسيط:** بالرغم من أن الوسيط يعتبر من مقاييس النزعة المركزية الجيدة إلا أن له بعض العيوب نذكر منها ما يلي:
  ١. لا يأخذ الوسيط في الاعتبار جميع البيانات إذا أنه يعتمد فقط على القيم التي في المنتصف وعلى ترتيب البيانات بغض النظر عن قيمها.
  ٢. لا يمكن بشكل عام حساب الوسيط للبيانات الوصفية (النوعية).

ملاحظة:

وحدة الوسيط هي نفس وحدة البيانات الأصلية. فإذا كانت وحدة البيانات هي الكيلوجرام فإن وحدة الوسيط هي الكيلوجرام.

(٣-٥) المنوال: Mode:

المنوال هو أحد مقاييس النزعة المركزية شائعة الاستخدام ولاسيما في حالة البيانات الوصفية (النوعية). ويعرف المنوال لمجموعة من البيانات على أنه تلك القيمة التي تتكرر أكثر من غيرها أي أنها القيمة ذات التكرار الأكبر (إن وجدت). يرمز للمنوال بالرمز (Mod). ومن تعريف المنوال تتضح لنا عدة أنواع من البيانات:

١. بيانات ليس لها منوال وتسمى عديمة المنوال.
٢. بيانات لها منوال واحد وتسمى وحيدة المنوال.
٣. بيانات لها أكثر من منوال وتسمى متعددة المنوال.

أولاً: المنوال للبيانات المفردة (غير مبوبة):

المنوال = المشاهدة الأكثر تكرارًا (إن وجدت).

مثال:

الجدول أدناه يتضمن بيانات إحدى الدراسات التي طبقت على خمسة أشخاص لقياس العمر (بالسنة) والوزن (بالكيلوجرام) والطول (بالسم) وفصيلة الدم. أوجد منوال للبيانات المختلفة.

رقم الشخص	1	2	3	4	5
العمر	25	20	25	30	35
الوزن	70	55	65	70	65
الطول	164	162	155	165	158
فصيلة الدم	O	A	B	A	AB

الحل:

البيانات	المنوال	نوع البيانات بالنسبة للمنوال
العمر	25	وحيدة المنوال
الوزن	المنوال الأول = 65 المنوال الثاني = 70	متعددة المنوال (ثنائية المنوال)
الطول	لا يوجد	عديمة المنوال

البيانات	المنوال	نوع البيانات بالنسبة للمنوال
فصيلة الدم	A	وحيدة المنوال

ثانياً: المنوال للبيانات المبوبة:

نعرف الفترة المنوالية بأنها الفترة ذات التكرار الأكبر وهي الفترة التي يقع فيها منوال. وفي الجدال التكراري قد يكون هناك فترة منوالية واحدة أو عدة فترات منوالية أو قد لا يوجد فترة منوالية. ويمكن حساب المنوال للبيانات الملخصة في جدول تكراري بطرقتين هما: طريقة حسابية وطريقة بيانية. ويستخدم الجدول التكراري لإيجاد المنوال حسابيًا بينما يستخدم المدرج التكراري لإيجاد المنوال بيانيًا.

(أ) إيجاد المنوال حسابيًا:

الطريقة التالية هي طريقة تقريبية لإيجاد المنوال حسابيًا:

المنوال = مركز الفترة المنوالية

مثال (٣-٨):

أوجد قيمة المنوال حسابيًا لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (٢-٢).

الحل:

التكرار	مركز الفترة	مستوى الهيموجلوبين
3	13.45	12.95 – 13.95
5	14.45	13.95– 14.95
15	15.45	14. 95– 15.95
16	**16.45**	**15. 95– 16.95**
10	17.45	16. 95– 17.95
1	18.45	17. 95– 18.95

أكبر تكرار = 16

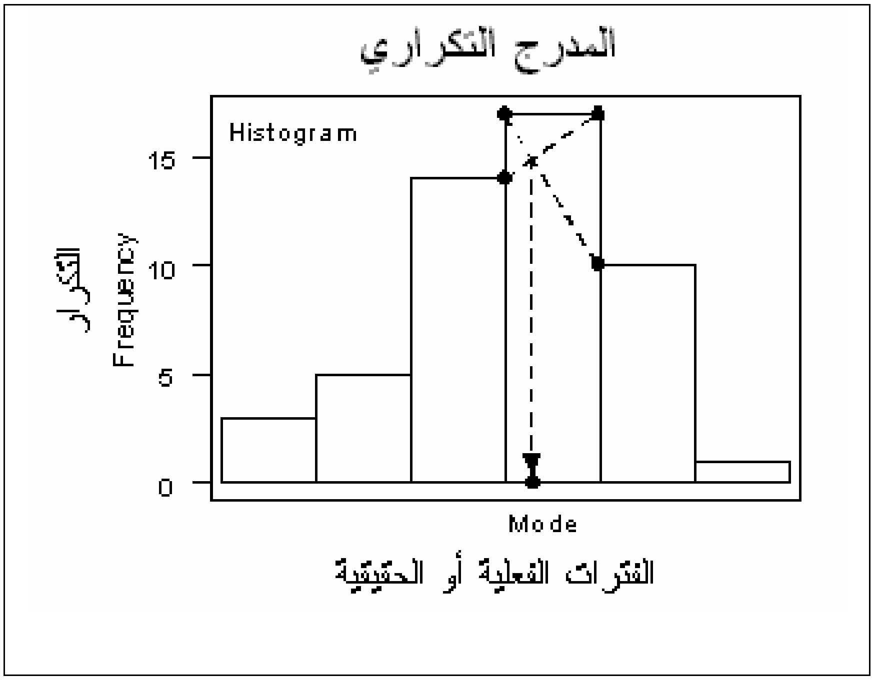
الفترة المنوالية هي: 15.95 – 16.95

المنوال = مركز الفترة المنوالية

= 16.45

(ب) إيجاد المنوال بيانيًا:

نستخدم المدرج التكراري لحساب المنوال. ففي المدرج التكراري نحدد الفترة المنوالية وهي الفترة ذات التكرار الأكبر (المستطيل الأطول). بعد تحديد الفترة المنوالية نحدد الفترتين السابقة واللاحقة للفترة المنوالية. بعد ذلك نرسم خط مستقيم يصل القمة اليمنى لمستطيل الفترة المنوالية بالقمة اليمنى لمستطيل الفترة السابقة ونرسم خط مستقيم يصل القمة اليسرى لمستطيل الفترة المنوالية بالقمة اليسرى لمستطيل الفترة اللاحقة. وعند نقطة تقاطع الخطين نرسم عمود

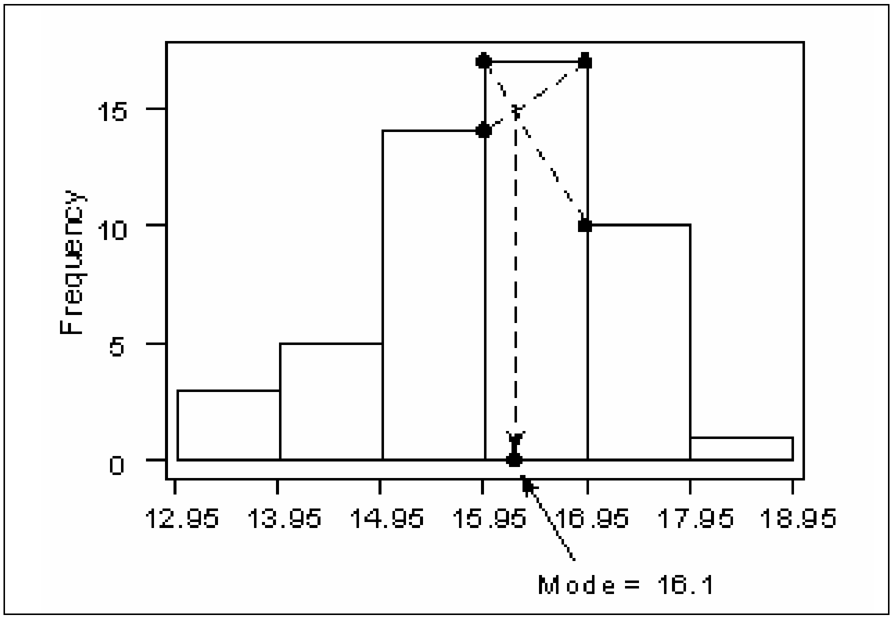


مثال (٣-٩):

أوجد قيمة المنوال بيانيًا لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (٢-٢).

الحل:

نرسم المدرج التكراري كما مر معنا سابقًا ثم نطبق طريقة حساب المنوال بيانيًا. باستخدام الشكل أدناه نجد أن القيمة التقريبية للمنوال هي: المنوال = 16.1.



بعض مميزات وعيوب المنوال:

- **مميزات المنوال:** يعتبر المنوال من مقاييس النزعة المركزية الشائعة ومن مميزاته نذكر ما يلي:
  1. المنوال سهل التعريف والحساب.
  2. المنوال أقل تأثرًا من المتوسط بالقيم الشاذة أو المتطرفة.
  3. يمكن حساب المنوال للبيانات الكمية والوصفية (النوعية).
- **عيوب المنوال:** بالرغم من أن المنوال يعتبر من مقاييس النزعة المركزية الشائعة إلا أن له بعض العيوب نذكر منها ما يلي:
  1. لا يأخذ المنوال في الاعتبار جميع البيانات إذا أنه يعتمد فقط على البيانات ذات التكرار الأكثر.
  2. قد لا يوجد منوال لمجموعة من البيانات أو قد يكون هناك أكثر من منوال.

ملاحظة:

وحدة المنوال هي نفس وحدة البيانات الأصلية. فإذا كانت وحدة البيانات هي الكيلوجرام فإن وحدة المنوال هي الكيلوجرام.

(٣-٦) الربيعات والعشيرات والمئينات:

رأينا سابقًا أن الوسيط هو القيمة التي تجزئ البيانات بعد ترتيبها إلى جزأين متساويين.

½ n 50% ½ n 50%

-------------------------|-----------------------------

Med

والآن لو جزأنا البيانات بعد ترتيبها إلى أربعة أجزاء متساوية:

¼ n =25% ¼ n=25% ¼ n=25% ¼ n= 25%

-----------------|--------------------|--------------------|----------------

Q_1 **Q_2** **Q_3**

فإن نقاط التقسيم هي:

Q_1 = الربيع الأول = القيمة التي يسبقها $\frac{1}{4}$ البيانات أو 25% من البيانات وتكون رتبتها = $\frac{n}{4}$

Q_2 = الربيع الثاني = القيمة التي يسبقها $\frac{2}{4}$ البيانات أو 50% من البيانات وتكون رتبتها = $\frac{2n}{4}$

Q_3 = الربيع الثالث = القيمة التي يسبقها $\frac{3}{4}$ البيانات أو 75% من البيانات وتكون رتبتها = $\frac{3n}{4}$

أما إذا جزأنا البيانات بعد ترتيبها إلى عشرة أجزاء متساوية:

$\frac{n}{10}$ $\frac{n}{10}$ $\frac{n}{10}$ $\frac{n}{10}$ $\frac{n}{10}$ $\frac{n}{10}$ $\frac{n}{10}$ $\frac{n}{10}$ $\frac{n}{10}$ $\frac{n}{10}$

------|---------|---------|---------|---------|----------|----------|----------|---------|-----

D_1 **D_2** **D_3** **D_4** **D_5** **D_6** **D_7** **D_8** **D_9**

فإن نقاط التقسيم هي:

D_1 = العشير الأول = القيمة التي يسبقها $\frac{1}{10}$ البيانات أو 10% من البيانات ويكون ترتيبها = $\frac{n}{10}$

D_2 = العشير الثاني = القيمة التي يسبقها $\frac{2}{10}$ البيانات أو 20% من البيانات ويكون ترتيبها = $\frac{2n}{10}$

D_3 = العشير الثالث = القيمة التي يسبقها $\frac{3}{10}$ البيانات أو 30% من البيانات ويكون ترتيبها = $\frac{3n}{10}$

وهكذا ··· حتى

D_9 = العشير التاسع = القيمة التي يسبقها $\frac{9}{10}$ البيانات أو 90% من البيانات ويكون ترتيبها = $\frac{9n}{10}$

أما إذا جزأنا البيانات بعد ترتيبها إلى مائة جزء متساوي:

$\frac{n}{100}$ $\frac{n}{100}$ $\frac{n}{100}$ $\frac{n}{100}$ $\frac{n}{100}$

------|---------|---------|---|---------|-----

$P_1 \quad P_2 \quad P_3 \quad \ldots \quad P_{98} \quad P_{99}$

فإن نقاط التقسيم هي:

P_1 = المئين الأول = القيمة التي يسبقها $\frac{1}{100}$ البيانات أو 1% من البيانات ويكون ترتيبها = $\frac{n}{100}$

P_2 = المئين الثاني = القيمة التي يسبقها $\frac{2}{100}$ البيانات أو 2% من البيانات ويكون ترتيبها = $\frac{2n}{100}$

P_3 = المئين الثالث = القيمة التي يسبقها $\frac{3}{100}$ البيانات أو 3% من البيانات ويكون ترتيبها = $\frac{3n}{100}$

وهكذا ... حتى

P_{99} =المئين التاسع والتسعون=القيمة التي يسبقها $\frac{99}{100}$ البيانات أو 99% من البيانات ويكون ترتيبها= $\frac{99n}{100}$

ملاحظات:

١. $Med = \mathbf{Q_2} = \mathbf{D_5} = \mathbf{P_{50}}$

٢. $\mathbf{Q_1} = \mathbf{P_{25}}\ ,\quad \mathbf{Q_3} = \mathbf{P_{75}}$

٣. $\mathbf{D_1} = \mathbf{P_{10}}\ ,\ \mathbf{D_2} = \mathbf{P_{20}}\ ,\ \mathbf{D_3} = \mathbf{P_{30}}\ ,\ \ldots,\ \mathbf{D_9} = \mathbf{P_{90}}$

٤.

المقياس	الرمز	الرتبة (R)
الربيع رقم k	$\mathbf{Q_k}$	$\frac{k\,n}{4}$
العشير رقم k	$\mathbf{D_k}$	$\frac{k\,n}{10}$
المئين رقم k	$\mathbf{P_k}$	$\frac{k\,n}{100}$

إيجاد الربيعات والعشيرات والمئينات للبيانات المبوبة:

إن طريقة حساب الربيعات والعشيرات والمئينات سواءً كانت حسابية أم بيانية فهي مشابهه لطريقة حساب الوسيط.

أولاً: إيجاد الربيعات والعشيرات والمئينات حسابيًا:

نستخدم الجدول التكراري المتجمع الصاعد لحساب هذه المقاييس. بعد تحديد رتبة المقياس المطلوب نستخدم الصيغة التالية:

$$\text{المقياس} = A + \left(\frac{R - F_1}{F_2 - F_1}\right) \times L$$

حيث أن:

R = رتبة المقياس

A = بداية فترة المقياس

L = طول فترة المقياس

F_1 = التكرار المتجمع الصاعد **السابق** لرتبة المقياس R

F_2 = التكرار المتجمع الصاعد **اللاحق** لرتبة المقياس R

مثال:

أوجد حسابيًا المقاييس التالية لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (٢-٢):

(أ) العشير الثاني

(ب) المئين التاسع والتسعين.

الحل:

	مستوى الهيموجلوبين		التكرار المتجمع الصاعد	
	12.95	أقل من	0	
	13.9 5	أقل من	3	
	14. 95= A	أقل من	8 = F_1	$\leftarrow R = \frac{2n}{10} = 10$
$D_2 \Rightarrow$	15. 95	أقل من	23 = F_2	
	16. 95	أقل من	39	
	17. 95 = A^*	أقل من	49= F^*_1	$\leftarrow R^* = \frac{99n}{100} = 49.5$
$P_{99} \Rightarrow$	18.95	أقل من	50= F^*_2	

(أ) حساب العشير الثاني:

$$R = \frac{2n}{10} = \frac{2 \times 50}{10} = 10$$

$A = 14.95, \quad L = 1.00, \quad F_1 = 8 \ , \ F_2 = 23$

$$D_2 = A + \left(\frac{R - F_1}{F_2 - F_1}\right) \times L = 14.95 + \left(\frac{10 - 8}{23 - 8}\right) \times 1.00 = 15.08$$

في هذا المثال نستطيع القول بأن 20% من الأشخاص يقل مستوى الهيموجلوبين لهم عن $D_2=15.08 \quad (D_2 = P_{20})$.

(ب) حساب المئين التاسع والتسعين:

$$R = \frac{99n}{100} = \frac{99 \times 50}{100} = 49.5$$

$$A^* = 17.95, \quad L^* = 1.00, \quad F^*_1 = 49 \;, \; F^*_2 = 50$$

$$P_{99} = A^* + \left(\frac{R^* - F^*_1}{F^*_2 - F^*_1}\right) \times L^* = 17.95 + \left(\frac{49.5 - 49}{50 - 49}\right) \times 1.00 = 18.45$$

في هذا المثال نستطيع القول بأن 99% من الأشخاص يقل مستوى الهيموجلوبين لهم عن $P_{99}=18.45$.

<u>ثانيًا: إيجاد الربيعات والعشيرات والمئينات بيانيًا:</u>

إن طريقة حساب الربيعات والعشيرات والمئينات بيانيًا مشابهه لطريقة حساب الوسيط إذ نستخدم المضلع التكراري المتجمع الصاعد مع ملاحظة تحديد الرتبة المناسبة للمقياس المطلوب بدلاً من رتبة الوسيط $\frac{n}{2}$.

<u>مثال:</u>

أوجد بيانيًا المقاييس التالية لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (٢-٢):

(أ) العشير الثاني

(ب) المئين التاسع والتسعين.

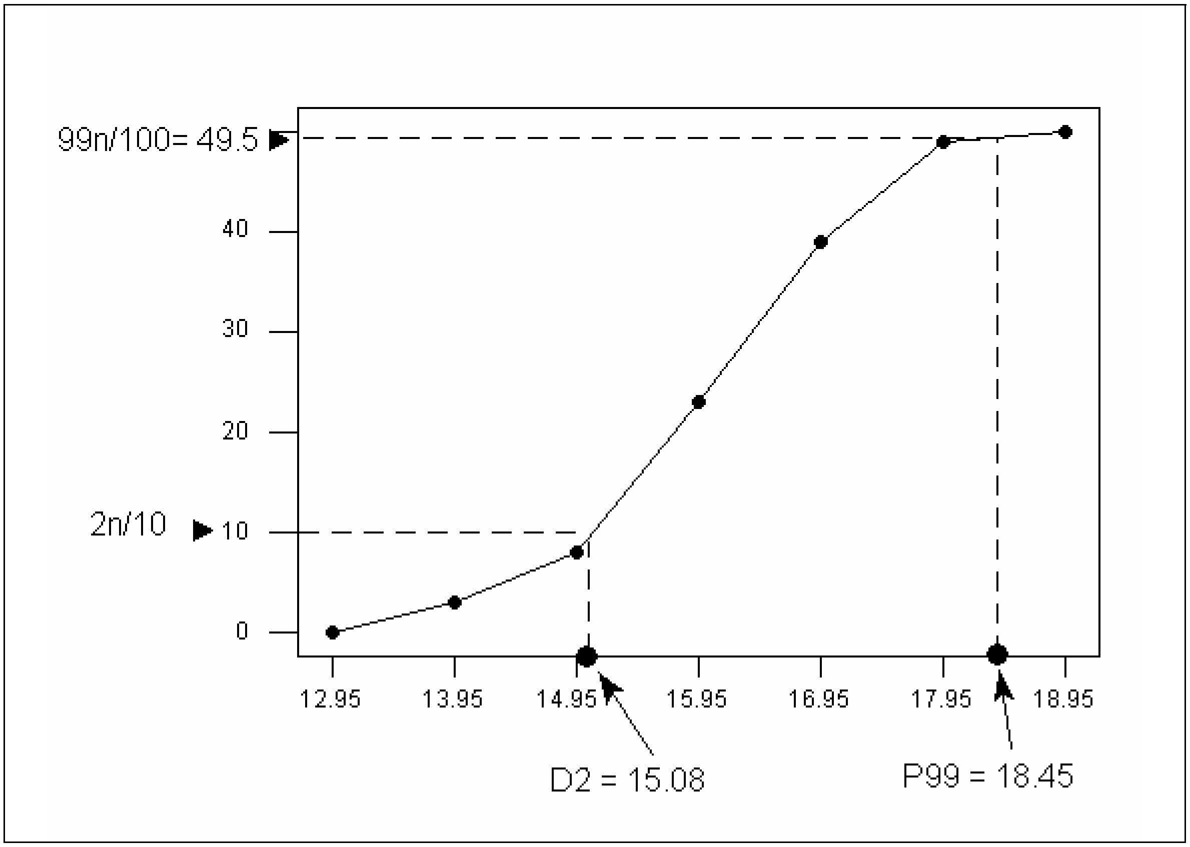
<u>الحل:</u>

رتبة العشير الثاني هي $R = \frac{2n}{10} = \frac{2 \times 50}{10} = 10$.

رتبة المئين التاسع والتسعين هي $R = \frac{99n}{100} = \frac{99 \times 50}{100} = 49.5$.

من الشكل أدناه، فإن القيم التقريبية للعشير الثاني D_2 والمئين التاسع والتسعين P_{99} هما على التوالي:

$$D_2 = 15.08$$
$$P_{99} = 18.45$$



ملاحظة:

وحدة مقاييس النزعة المركزية وكذلك الربيعات والعشيرات والمئينات هي نفس وحدة البيانات الأصلية. فإذا كانت وحدة البيانات هي الكيلوجرام فإن وحدة جميع مقاييس النزعة المركزية وكذلك الربيعات والعشيرات والمئينات هي الكيلوجرام.

٤. مقاييس التشتت (الاختلاف)
Measures of Dispersion (Variation)

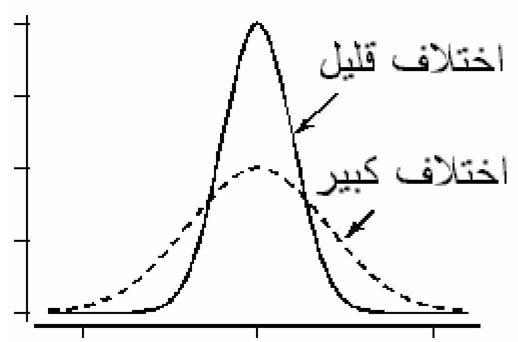
(٤-١) مقدمة:

لقد ذكرنا في الفصل السابق بعض مقاييس النزعة المركزية والتي تمثل مقاييس عددية لموضع أو مكان تركز البيانات لظاهرة ما. وقد ذكرنا بأن هذه المقاييس تستخدم لمقارنة مجموعات البيانات المختلفة. وفي الحقيقة فإن مقاييس النزعة المركزية غير كافية لإيجاد مقارنة شاملة بين مجموعات البيانات المختلفة. فقد تكون هناك مجموعات من البيانات لها نفس مقاييس النزعة المركزية (لها نفس الموضع) ولكنها تختلف في بعض الصفات الأخرى. فمثلاً المثال التالي يبين لنا مجموعتين من البيانات لهما نفس المتوسط ولكنهما مختلفتان في طبيعة تشتتهما.

مثال:

المجموعة	البيانات	المتوسط	
الأولى	59, 61, 62, 58, 60	60	-----------------●--●--●--●--●-----------
الثانية	50, 60, 66, 54, 70	60	-----------●-----●-----●-----●------●--- 60

بالرغم من أن المتوسط يساوي 60 للمجموعتين إلا أن التشتت (أو الاختلاف) بين القيم في كل مجموعة غير متساو. فمن الواضح أن بيانات المجموعة الأولى أكثر تقاربًا فيما بينها (أقل تشتتًا وتباعدًا فيما بينها) من بيانات المجموعة الثانية. لذلك دعت الحاجة لإيجاد مقاييس تقيس طبيعة تشتت (أو تفرق أو اختلاف أو تباعد) البيانات فيما بينها. هذه المقاييس تسمى مقاييس التشتت أو الاختلاف.



المضلعان التكراريان لتوزيعين لهما نفس مقاييس
النزعة المركزية ولكنهما مختلفين في التشتت

مقاييس التشتت هي مقاييس عددية تستخدم لقياس اختلاف أو تشتت البيانات. والاختلاف أو التشتت لمجموعة من البيانات هو مقدار تفرق أو تباعد أو انتشار البيانات فيما بينها. فتشتت البيانات يكون صغيرًا إذا كانت البيانات متقاربة فيما بينها والعكس بالعكس. وأما البيانات المتساوية فلا اختلاف ولا تشتت فيها. ومقاييس التشتت تستخدم لوصف مجموعة البيانات وكذلك لمقارنة مجموعات البيانات المختلفة إذ أن مقاييس النزعة المركزية لا تكفي وحدها لوصف مجموعة البيانات أو مقارنة مجموعات البيانات المختلفة. ومن أشهر مقاييس التشتت نذكر:

١. المدى: Range
٢. نصف المدى الربيعي Semi-Inter-quartile Range
٣. التباين: Variance
٤. الانحراف المعياري: Standard Deviation
٥. معامل الاختلاف (أو التغير): Coefficient of Variation

(٤-٢) المدى: Range

يعتبر المدى من أسهل مقاييس التشتت تعريفًا وحسابًا ويعطينا فكرة سريعة عن مدى تفرق البيانات. ويعرف المدى لمجموعة من البيانات بالصيغة التالية:

$$\boxed{\text{Range} = X_{max} - X_{min}}$$

حيث أن:

X_{max} = أكبر قيمة (للبيانات المفردة) = مركز الفترة العليا (للبيانات المبوبة)

X_{min} = أصغر قيمة (للبيانات المفردة) = مركز الفترة الدنيا (للبيانات المبوبة)

مثال (٤-١):

أوجد المدى للمشاهدات التالية والتي هي عبارة عن أوزان (بالكيلوجرام) مجموعة مكونة من سبعة أشخاص: 25, 30, 40, 45, 35, 55, 50

الحل:

$X_{max} = 55$
$X_{min} = 25$
$\text{Range} = X_{max} - X_{min} = 55 - 25 = 30$ (كيلوجرامًا)

مثال (٤-٢):

أوجد المدى لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (٢-٢).

<u>الحل:</u>

مستوى الهيموجلوبين	مركز الفترة x	التكرار f
12.95 – 13.95	**13.45**	3
13.9 5– 14.95	14.45	5
14. 95– 15.95	15.45	15
15. 95– 16.95	16.45	16
16. 95– 17.95	17.45	10
17. 95– 18.95	**18.45**	1

X_{max} = مركز الفترة العليا = 18.45

X_{min} = مركز الفترة الدنيا = 13.45

$$\begin{aligned} Range &= X_{max} - X_{min} \\ &= 18.45 - 13.45 \\ &= 5.00 \end{aligned}$$

<u>بعض مميزات وعيوب المدى:</u>

- **مميزات المدى:** سهل التعريف والحساب
- **عيوب المدى:**
  ١. يتأثر المدى بالقيم الشاذة أو المتطرفة.
  ٢. لا يأخذ المدى في الاعتبار جميع البيانات.

ملاحظات:

١. وحدة المدى هي نفس وحدة البيانات الأصلية.

٢. نظرًا لأن المدى يعتمد فقط على أكبر وأصغر قيمة ولا يأخذ في الاعتبار القيم الأخرى فهو مقياس غير جيد لقياس التشتت.

<u>(٤-٣) نصف المدى الربيعي Semi-Inter-quartile Range:</u>

رأينا أن المدى يتأثر كثيرًا بالقيم الشاذة أو المتطرفة. ولذلك دعت الحاجة لإيجاد مقاييس أخرى للتشتت لا تتأثر بالقيم المتطرفة. وأحد هذه المقاييس هو نصف المدى الربيعي. وحيث أن القيم المتطرفة هي تلك القيم الصغيرة جدًا أو الكبيرة جدًا فإنه عند حساب نصف المدى الربيعي لا يؤخذ في الاعتبار ربع البيانات الصغيرة (25%) ولا ربع البيانات الكبيرة (25%).

¼ n =25%   ¼ n= 25%

-/-/-/-/-/-/-/-/-/-|----------------|----------------|-/-/-/-/-/-/-/-/-/-

ربع البيانات الصغيرة **Q_1**   **Q_2**   **Q_3** ربع البيانات الكبيرة

يرمز لنصف المدى الربيعي بالرمز Q ويعرف بالصيغة التالية:

$$Q = \frac{Q_3 - Q_1}{2}$$

حيث أن Q_1 هو الربيع الأول و Q_3 هو الربيع الثالث وقد مر معنا كيفية إيجادهما للبيانات المبوبة بالطريقة الحسابية والبيانية.

مثال (٤-٣):

أوجد نصف المدى الربيعي لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (٢-٢) باستخدام:

(أ) الطريقة الحسابية

(ب) الطريقة البيانية

الحل:

(أ) الطريقة الحسابية:

	مستوى الهيموجلوبين		التكرار المتجمع الصاعد	
	12.95	أقل من	0	
	13.9 5	أقل من	3	
$Q_1 \Rightarrow$	14. 95= A	أقل من	$8 = F_1$	$\leftarrow R = \frac{n}{4} = 12.5$
$Q_3 \Rightarrow$	15. 95= A^*	أقل من	$23 = F_2 = F^*_1$	$\leftarrow R^* = \frac{3n}{4} = 37.5$
	16. 95	أقل من	$39 = F^*_2$	
	17. 95	أقل من	49	
	18.95	أقل من	50	

حساب الربيع الأول Q_1:

$$R = \frac{n}{4} = \frac{50}{4} = 12.5 \quad , \quad A = 14.95, \quad L = 1.00, \quad F_1 = 8 \quad , \quad F_2 = 23$$

$$Q_1 = A + \left(\frac{R - F_1}{F_2 - F_1}\right) \times L = 14.95 + \left(\frac{12.5 - 8}{23 - 8}\right) \times 1.00 = 15.25$$

حساب الربيع الثالث Q_3:

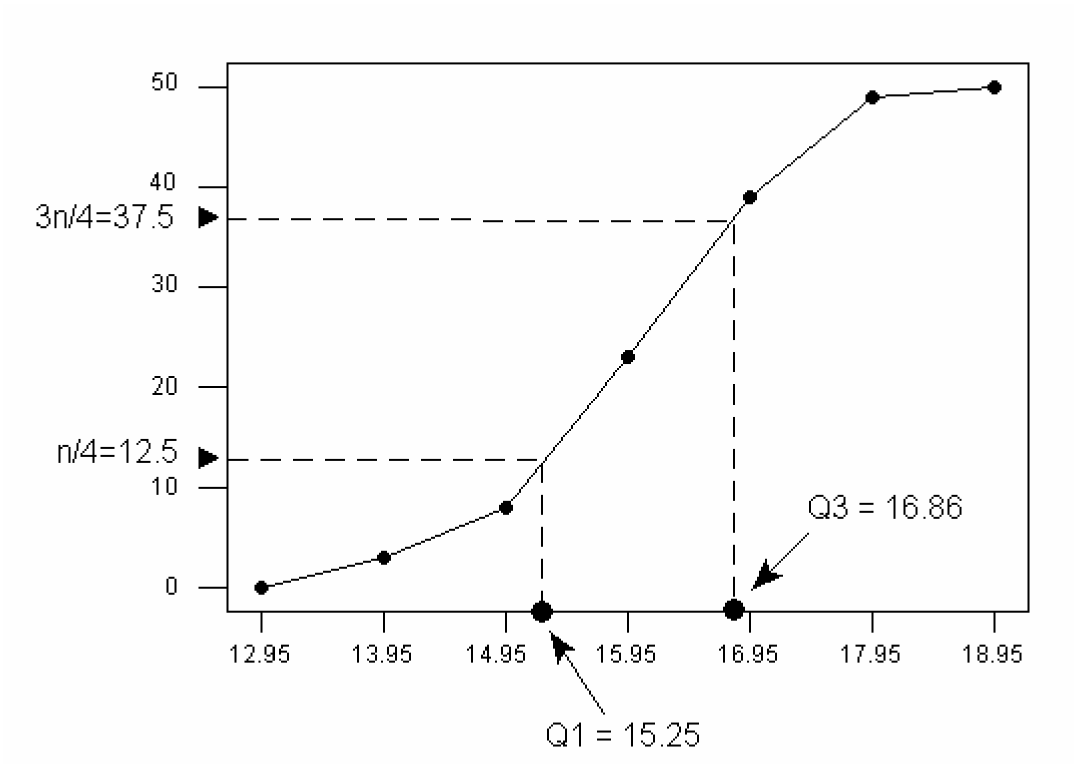
$$R^* = \frac{3n}{4} = \frac{3 \times 50}{4} = 37.5 \quad , \quad A^* = 15.95, \quad L^* = 1.00, \quad F^*_1 = 23 \quad , \quad F^*_2 = 39$$

$$Q_3 = A^* + \left(\frac{R^* - F_1^*}{F_2^* - F_1^*}\right) \times L^* = 15.95 + \left(\frac{37.5 - 23}{39 - 23}\right) \times 1.00 = 16.86$$

وباستخدام الصيغة فإن نصف المدى الربيعي هو:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{16.86 - 15.25}{2} = 1.61$$

(ب) الطريقة البيانية:



وباستخدام الصيغة فإن نصف المدى الربيعي هو:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{16.86 - 15.25}{2} = 1.61$$

بعض مميزات وعيوب نصف المدى الربيعي:

- **من المميزات:** لا يتأثر بالقيم الشاذة أو المتطرفة.
- **من العيوب :** لا يأخذ في الاعتبار جميع البيانات.

ملاحظة:

وحدة نصف المدى الربيعي هي نفس وحدة البيانات الأصلية.

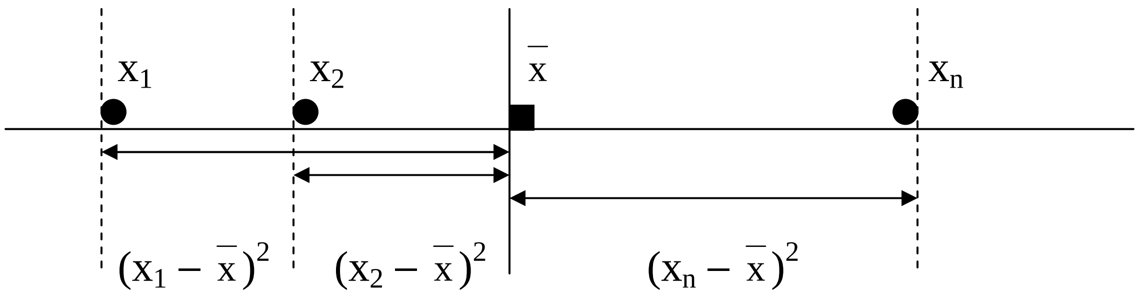
(٤-٤) التباين (Variance) والانحراف المعياري (Standard Deviation):

يعتبر التباين والانحراف المعياري من أهم وأفضل مقاييس التشتت ومن أكثرها شيوعًا واستخدامًا في التحليل الإحصائي وذلك لما يتمتعان به من خصائص وصفات إحصائية جيدة.

التباين: Variance

فكرة التباين تعتمد على تشتت أو تباعد البيانات عن متوسطها. فالتباين يكون كبيرًا إذا كانت البيانات متباعدة عن متوسطها والعكس بالعكس.

ويعرف التباين بأنه متوسط مربع انحرافات القيم عن الوسط الحسابي ويرمز له بالرمز S^2.



القيم (البيانات)	x_n	$\cdots$	x_2	x_1
انحرافات القيم عن المتوسط	$x_n - \bar{x}$	$\cdots$	$x_2 - \bar{x}$	$x_1 - \bar{x}$
مربع انحرافات القيم عن المتوسط	$(x_n - \bar{x})^2$	$\cdots$	$(x_2 - \bar{x})^2$	$(x_1 - \bar{x})^2$

الانحراف المعياري:

إن التباين من أهم وأفضل مقاييس التشتت ولكنه يقاس بوحدة البيانات الأصلية المربعة. وفي كثير من الأحيان نرغب في استخدام مقياس للتشتت يقاس بوحدة البيانات الأصلية ويتمتع بخصائص إحصائية جيدة مثل التباين. وأحد هذه المقاييس هو الانحراف المعياري. ويعرف الانحراف المعياري على أنه الجذر التربيعي للتباين ويرمز له بالرمز S.

حساب التباين والانحراف المعياري:

أولاً: التباين والانحراف المعياري للبيانات المفردة (غير المبوبة):

إذا كانت $x_1, x_2, \ldots, x_n$ عينه حجمها n وكان متوسطها هو $\bar{x}$ فإن تباين العينة يعرف كما يلي:

$$S^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

وأما الانحراف المعياري فإنه يعرف بالصيغة:

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

ملاحظات:

١. $S^2 \geq 0$ (دائمًا) وكذلك $S \geq 0$ (دائمًا).

٢. $S^2 = 0 \Leftrightarrow S = 0 \Leftrightarrow$ جميع قيم العينة متساوية (لا يوجد اختلاف بين القيم).

٣. وحدة S^2 هي وحدة البيانات الأصلية المربعة.

٤. وحدة S هي نفس وحدة البيانات الأصلية.

٥. يمكن حساب التباين بالصيغة الحسابية التالية:

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}$$

ولحساب تباين العينة باستخدام الصيغة الحسابية السابقة فإننا نحتاج إلى معرفة الكميات التالية فقط دون الحاجة لمعرفة البيانات الأصلية:

- حجم العينة = n .
- مجموع البيانات = $\sum_{i=1}^{n} x_i$.
- مجموع مربعات البيانات = $\sum_{i=1}^{n} x_i^2$.

والصيغة الحسابية السابقة تستخدم لحساب تباين العينة وذلك لسببين هما:

١. لأنها أكثر سهولة.

٢. لأنها أكثر دقة في الحساب عندما يكون هناك تقريب في حساب متوسط العينة.

مثال (٤-٤):

أوجد تباين العينة والانحراف المعياري لمجموعة الأوزان (بالكيلوجرام) التالية:

7.1, 2.5, 2.5, 5.4, 8.3

الحل:

نلخص الحل في الجدول التالي:

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	x^2
7.1	1.94	3.7636	50.41
2.5	−2.66	7.0756	6.25
2.5	−2.66	7.0756	6.25
5.4	0.24	0.0576	29.16
8.3	3.14	9.8596	68.89
$\sum_{i=1}^{n} x_i = 25.8$	0.00	$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 27.832$	$\sum_{i=1}^{n} x_i^2 = 160.96$

$n = 5$

$$\sum_{i=1}^{n} x_i = 25.8$$

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 27.832$$

$$\sum_{i=1}^{n} x_i^2 = 160.96$$

متوسط العينة هو:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{25.8}{5} = 5.16 \text{ (كيلوجرامًا)}$$

حساب تباين العينة:

(أ) باستخدام التعريف:

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{27.832}{5-1} = 6.958 \text{ (كيلوجرامًا مربعًا)}$$

(ب) باستخدام الصيغة الحسابية:

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{160.96 - \frac{(25.8)^2}{5}}{5-1} = \frac{160.96 - 133.128}{4} = 6.958$$

الانحراف المعياري هو:

$$S=\sqrt{S^2}=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}=S=\sqrt{6.958}=2.6378$$ (كيلوجرامًا)

بعض خصائص التباين والانحراف المعياري:

١. يخضع التباين والانحراف المعياري لبعض العمليات الجبرية كما يلي:

المشاهدات	الانحراف المعياري	التباين
$x_1, x_2, \dots, x_n$	S	S^2
$x_1 \pm b, x_2 \pm b, \dots, x_n \pm b$	S	S^2
$ax_1, ax_2, \dots, ax_n$	$\lvert a\rvert S$	$a^2 S^2$
$ax_1 \pm b, ax_2 \pm b, \dots, ax_n \pm b$	$\lvert a\rvert S$	$a^2 S^2$

- مثال:

المشاهدات		الانحراف المعياري	التباين
: x	2, 6, 4, 3, 5	S = 1.581	S^2= 2.5
: x+5	7, 11, 9, 8, 10	1.581	2.5
: 3x	6, 18, 12, 9, 15	$\lvert 3\rvert \times 1.581=4.743$	$9\times 2.5=22.5$
: 3x +5	11, 23, 17, 14, 20	$\lvert 3\rvert \times 1.581=4.743$	$9\times 2.5=22.5$

- مثال:

إذا كان التباين للمشاهدات $x_1, x_2, \dots, x_n$ هو 36 فإن التباين للمشاهدات $\frac{x_1-10}{2}, \frac{x_2-10}{2}, \dots, \frac{x_n-10}{2}$ هو $\left(\frac{1}{2}\right)^2 36=\frac{36}{4}=9$ وأما الانحراف المعياري فهو $\sqrt{9}=3$.

٢. إذا كان لدينا مجموعتان من البيانات بحيث أن عدد بيانات المجموعة الأولى n_1 ومتوسطها $\bar{x}_1$ وتباينها S_1^2 وكان عدد بيانات المجموعة الثانية n_2 ومتوسطها $\bar{x}_2$ وتباينها S_2^2 وإذا كان $\bar{x}_1=\bar{x}_2$ (أي أن متوسطي المجموعتين متساويان) فإن تباين المجموعة الكلية المكونة من دمج هاتين المجموعتين يمكن حسابه بالصيغة التالية:

$$\boxed{S^2=\frac{(n_1-1)\,S_1^2+(n_2-1)\,S_2^2}{n_1+n_2-1}}$$

- مثال:

أوجد تباين المجموعة الكلية المكونة من دمج المجموعتين التاليتين:

	المجموعة الأولى	المجموعة الثانية
حجم العينة	$n_1 = 4$	$n_2 = 6$
المتوسط	$\bar{x}_1 = 5$	$\bar{x}_2 = 5$
التباين	$S_1^2 = 3$	$S_2^2 = 3.5$

الحل:

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

$$= \frac{(4-1)(3) + (6-1)(3.5)}{4+6-1} = 2.944$$

ثانياً: التباين والانحراف المعياري للبيانات المبوبة:

إذا كان لدينا بيانات عددها n وكانت هذه البيانات ملخصة في جدول تكراري بحيث أن:

- عدد الفترات هو k
- مراكز الفترات هي $x_1, x_2, \ldots, x_k$
- تكرارات الفترات هي $f_1, f_2, \ldots, f_k$

بطريقة مشابهة لحسب المتوسط للتوزيع التكراري فإن التباين للتوزيع التكراري المبوب يمكن حسابه بشكل تقريبي بالصيغة التالية:

$$\boxed{S^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{n-1}}$$

حيث أن:

$$n = \sum_{i=1}^{k} f_i\,, \quad \bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum xf}{n}\ ,$$

كما يمكن استخدام الصيغة الحسابية التالية:

$$\boxed{S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} x_i^2 f_i - \frac{\left(\sum_{i=1}^{k} x_i f_i\right)^2}{n}\right) = S^2 = \frac{\sum_{i=1}^{k} x_i^2 f_i - \frac{\left(\sum_{i=1}^{k} x_i f_i\right)^2}{n}}{n-1}}$$

ويكمن تلخيص عمليتي إيجاد المتوسط والتباين باستخدام الجدول التالي:

الفترة	مركز الفترة x	التكرار f	$x\,f$	$x^2\,f$	$f\,(x-\overline{x})^2$
الفترة رقم 1	x_1	f_1	$x_1\,f_1$	$x_1^2\,f_1$	$f_1\,(x_1-\overline{x})^2$
الفترة رقم 2	x_2	f_2	$x_2\,f_2$	$x_2^2\,f_2$	$f_2\,(x_2-\overline{x})^2$
:	:	:	:	:	:
:	:	:	:	:	:
الفترة رقم k	x_k	f_k	$x_k\,f_k$	$x_k^2\,f_k$	$f_k\,(x_k-\overline{x})^2$
المجموع		$\sum f = n$	$\sum x\,f$	$\sum x^2 f$	$\sum_{i=1}^{k} f_i\,(x_i-\overline{x})^2$

مثال (٤-٥):

أوجد التباين والانحراف المعياري لمستوى الهيموجلوبين في الدم لعينة مكونة من خمسين شخصا تم تلخيص مستوى الهيموجلوبين لهم كما في مثال (٢-٢).

الحل:

مستوى الهيموجلوبين	مركز الفترة x	التكرار f	$x\,f$	$x^2\,f$	$f\,(x-\overline{x})^2$ $f\,(x-16.01)^2$
12.95 – 13.95	13.45	3	40.35	542.708	19.6608
13.9 5– 14.95	14.45	5	72.25	1044.013	12.1680
14. 95– 15.95	15.45	15	231.75	3580.538	4.7040
15. 95– 16.95	16.45	16	263.20	4329.640	3.0976
16. 95– 17.95	17.45	10	174.50	3045.025	20.7360
17. 95– 18.95	18.45	1	18.45	340.403	5.9536
المجموع		$\sum f = 50$	$\sum x\,f = 800.5$	$\sum x^2 f = 12882.33$	$\sum_{i=1}^{k} f_i\,(x_i-\overline{x})^2 = 66.320$

$$\overline{x} = \frac{\sum xf}{\sum f} = \frac{\sum xf}{n} = \frac{800\ .5}{50} = 16\ .01$$

حساب التباين باستخدام صيغة التعريف:

$$S^2 = \frac{\sum_{i=1}^{k} f_i\,(x_i-\overline{x})^2}{n-1} = \frac{66.320}{50-1} = 1.3535$$

حساب التباين باستخدام الصيغة الحسابية:

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} x_i^2 f_i - \frac{\left(\sum_{i=1}^{k} x_i f_i\right)^2}{n}\right) = \frac{1}{50-1}\left(12882.33 - \frac{(800.5)^2}{50}\right)$$

$$= \frac{1}{49}(12882.33 - 12816.005) = \frac{66.325}{49} = 1.3536$$

حساب الانحراف المعياري:

$$S = \sqrt{S^2} = \sqrt{1.3536} = 1.163$$

(٤-٥) معامل الاختلاف (التغير): Coefficient of Variation:

ذكرنا سابقًا أن التباين والانحراف المعياري من المقاييس المفيدة لقياس التشتت لتوزيع متغير ما. ولكن في كثير من الأحيان نكون مهتمين بمقارنة التشتت والاختلاف لتوزيعي متغيرين مختلفين. وبما أن التباين والانحراف المعياري مقياسان يعتمدان على وحدة البيانات فإنه يصعب استخدامهما لمقارنة تجانس المجموعات المختلفة من البيانات وذلك لاختلاف الوحدة المستخدمة. وبشكل عام فإن مقاييس التشتت التي ذكرناها آنفًا تكون غير مناسبة لمقارنة تجانس مجموعات البيانات المختلفة في الحالتين التاليتين:

١. إذا كانت وحدتا المتغيرين مختلفتين حيث لا نستطيع مقارنة الوحدات المختلفة.

٢. إذا كان متوسطا المتغيرين مختلفين وذلك لأن تباين توزيع المتغير ذي المتوسط الصغير ينزع لأن يكون صغيرًا والعكس بالعكس.

لذلك دعت الحاجة إلى مقياس لا يعتمد على وحدة المتغير ويقيس ما يسمى بالتشتت النسبي. وأحد هذه المقاييس هو ما يسمى بمعامل الاختلاف أو معامل التغير. فمعامل الاختلاف هو أحد مقاييس التشتت النسبي وهو مقياس عديم الوحدة ويستخدم لمقارنة التشتت النسبي أو التجانس لمجموعات البيانات المختلفة. فمجموعة البيانات ذات معامل الاختلاف الأكبر يكون تشتتها النسبي أكبر أي أنها تكون أقل تجانسًا والعكس بالعكس. ويعرف معامل الاختلاف للعينة التي متوسطها $\bar{x}$ وانحرافها المعياري S بالصيغة التالية:

$$\boxed{C.V. = \frac{S}{\bar{x}}}$$

مثال (٤-٦):

الجدول أدناه يتضمن بيانات إحدى الدراسات التي طبقت على خمسة أشخاص لقياس الوزن (بالكيلوجرام) والطول (بالسنتيمتر). أي البيانات أكثر تشتتًا نسبًا (أقل تجانسًا) بيانات الأوزان أم بيانات الأطوال؟

رقم الشخص	1	2	3	4	5
الوزن	69	59	65	67	65
الطول	164	162	155	165	158

الحل:

أولاً نوجد المتوسط $\bar{x}$ والانحراف المعياري S لكل من بيانات الأوزان ولبيانات الأطوال كما مر معنا سابقاً. نلخص الحسابات في الجدول التالي:

البيانات	المتوسط $\bar{x}$	الانحراف المعياري S	معامل الاختلاف $C.V. = \frac{S}{\bar{x}}$
الأوزان	65.0 kg	3.7417 kg	0.0576
الأطوال	160.8 cm	4.2071 cm	0.026

بما أن معامل الاختلاف لبيانات الأوزان أكبر من معامل الاختلاف لبيانات الأطوال فإن التشتت النسبي لبيانات الأوزان أكبر من التشتت النسبي لبيانات الأطوال. أي أن بيانات الأوزان أقل تجانسًا من بيانات الأطوال.

(٤-٦) نظرية (متراجحة) تشيبيشيف **Chebychev Inequality**:

إن نظرية تشيبيشيف من النظريات المفيدة إذا أنها تعطينا حدًا أدنى لنسبة البيانات الواقعة في فترة معينة عند معرفة متوسط البيانات وانحرافها المعياري دون الحاجة لمعرفة البيانات الأصلية أو التوزيع الذي أخذت منه العينة. ونص النظرية هو:

> إذا كان لدينا عينة من البيانات متوسطها $\bar{x}$ وانحرافها المعياري S فإن نسبة البيانات الواقعة في الفترة $(\bar{x} - kS , \bar{x} + kS)$ لايقل عن $\left(1 - \frac{1}{k^2}\right)$ حيث أن $k>1$.

$$\underbrace{\bar{x} - kS \qquad\qquad \bar{x} \qquad\qquad \bar{x} + kS}$$

نسبة البيانات الواقعة في هذه الفترة لا يقل عن $\left(1 - \frac{1}{k^2}\right)$

ملاحظات:

١. تطبق نظرية تشيبيشيف للفترات التي منتصفها (مركزها) هو المتوسط.

٢. تستخدم نظرية تشيبيشيف بطريقتين(في كلا الحالتين لابد من معرفة قيمة k):

أ- تحديد النسبة (التقريبية) لعدد البيانات الواقعة في فترة معينة.

ب- تحديد الفترة التي يقع فيها ما لا يقل عن نسبة معينة.

مثال (٤-٧):

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x} = 7$ وانحرافها المعياري $S=5$ فما هي نسبة البيانات الواقعة في الفترة $(-4 , 18)$؟

الحل:

أولاً نلاحظ أن منتصف الفترة المعطاة $(-4 \; , \; 18)$ هو المتوسط $\bar{x} = 7$ لذلك نستطيع تطبيق نظرية تشيبيشيف. والآن:

$$\begin{aligned}
(\bar{x} - kS \; , \bar{x} + kS) = (-4 , 18) \quad &\Rightarrow \quad \bar{x} + kS = 18 \\
&\Leftrightarrow \quad 7 + k(5) = 18 \\
&\Leftrightarrow \quad 5k = 11 \\
&\Leftrightarrow \quad k = 11/5 \\
&\Leftrightarrow \quad \left(1 - \frac{1}{k^2}\right) = \left(1 - \frac{1}{(11/5)^2}\right) \\
&\Leftrightarrow \quad \left(1 - \frac{1}{k^2}\right) = 0.7934
\end{aligned}$$

لذلك فإن نسبة البيانات الواقعة في الفترة $(-4 , 18)$ لا تقل عن 79.34%.

مثال (٤-٨):

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x} = 7$ وانحرافها المعياري $S=5$ فأوجد فترة يقع فيها ما لا يقل عن 75% من البيانات.

الحل:

$$\left(1-\frac{1}{k^2}\right)=0.75 \Leftrightarrow \frac{1}{k^2}=1-0.75$$
$$\Leftrightarrow \frac{1}{k^2}=0.25$$
$$\Leftrightarrow k=\sqrt{\frac{1}{0.25}}$$
$$\Leftrightarrow k=2$$

وبالتالي فإن الفترة التي يقع فيها ما لا يقل عن 75% من البيانات هي:

$$(\bar{x}-kS\ ,\ \bar{x}+kS)=(7-2\times5\ ,\ 7+2\times5)$$
$$=(7-10\ ,\ 7+10)$$
$$=(-3\ ,\ 17)$$

(٤-٧) الدرجات (القيم) المعيارية:

لتكن $x_1, x_2, \ldots, x_n$ عينه من البيانات حجمها n ومتوسطها $\bar{x}$ وانحرافها المعيـــاري S.

نعرف الدرجة (القيمة) المعيارية للمشاهدة x_i بالصيغة التالية:

$$\boxed{z_i=\frac{x_i-\bar{x}}{S}\quad i=1, 2, \ldots, n}$$

أي أن الدرجات المعيارية للبيانات $x_1, x_2, \ldots, x_n$ هي:

$$z_1=\frac{x_1-\bar{x}}{S},\ z_2=\frac{x_2-\bar{x}}{S},\ \ldots,\ z_n=\frac{x_n-\bar{x}}{S}$$

ملاحظات:

١. $z_i=\frac{x_i-\bar{x}}{S}$ هي الدرجة المعيارية للمشاهدة الأصلية x_i.

٢. المشاهدة الأصلية للدرجة المعيارية z_i هي $x_i=\bar{x}+S\,z_i$.

٣. الدرجات المعيارية هي قيم عديمة الوحدة ولذلك فإنها تستخدم للمقارنة بـــين المـــشاهدات المختلفة في المجموعات المختلفة للبيانات.

٤. متوسط الدرجات المعيارية =0.

٥. الانحراف المعياري للدرجات المعيارية يساوي =1.

مثال (٤-٩):

إذا كان لدينا مجموعة من البيانات متوسطها $\bar{x}=7$ وانحرافها المعياري $S=5$ فأوجد:

١. الدرجة المعيارية للقيمة $x=9$.

٢. القيمة الأصلية للدرجة المعيارية $z = 0.1$.

الحل:

١. الدرجة المعيارية للقيمة $x = 9$ هي:

$$z=\frac{x-\overline{x}}{S} = \frac{9-7}{5}=\frac{2}{5}=0.4$$

٢. القيمة الأصلية للدرجة المعيارية $z = 0.1$ هي:

$$x = \overline{x} + S\,z = 7 + 5\times0.1 = 7 + 0.5 = 7.5$$

مثال (٤-١٠):

إذا كانت درجة أحد الطلاب في مقرر الإحصاء تساوي 82 ودرجته في مقرر الرياضيات تساوي 89، وإذا كان متوسط درجات الطلاب في مقرر الإحصاء يساوي 75 بانحراف معياري يساوي 10 ومتوسط درجات الطلاب في مقرر الرياضيات يساوي 81 بانحراف معياري يساوي 16، ففي أي المقررين كان أداء الطالب أفضل؟

الحل:

المقرر	المتوسط $\overline{x}$	الانحراف المعياري S	الدرجة x	الدرجة المعيارية $z=\frac{x-\overline{x}}{S}$
الإحصاء	75	10	82	$z=\frac{82-75}{10}=0.7$
الرياضيات	81	16	89	$z=\frac{89-81}{16}=0.5$

بما أن الدرجة المعيارية لمقرر الإحصاء 0.7 أكبر من الدرجة المعيارية لمقرر الرياضيات 0.5 فإن أداء الطالب في مقرر الإحصاء أفضل من أدائه في مقرر الرياضيات بالرغم من أن درجته في مقرر الإحصاء أقل من درجته في مقرر الرياضيات.

٥. طرق العد
Counting Techinques

(٥-١) مقدمة:

في كثير من التطبيقات نكون مهتمين بمعرفة عدد العناصر المنتمية إلى مجموعة ما. ونستخدم طرق العد لإيجاد عدد عناصر مجموعة ما دون الحاجة إلى سرد عناصرها. وهذه الطرق تساعدنا أيضًا في إيجاد عدد الطرق المختلفة والممكنة لإجراء أي تجربة وهذا بدوره يفيدنا كثيرًا في دراسة علم الاحتمال.

(٥-٢) القواعد الأساسية لطرق العد:

هناك قاعدتان أساسيتان لطرق العد هما قاعدة الضرب وقاعدة الجمع ونذكرهما فيما يلي:

أولاً: قاعدة الضرب:

نتيجة:

إذا كان هناك عملية (أو تجربة) مكونة من عدد مقداره r من المراحل بحيث:

- المرحلة رقم 1 تتم بعدد قدره n_1 من الطرق المختلفة
- المرحلة رقم 2 تتم بعدد قدره n_2 من الطرق المختلفة
- وهكذا ...
- المرحلة رقم r تتم بعدد قدره n_r من الطرق المختلفة

فإن العملية ككل يمكن أجراؤها بعدد من الطرق المختلفة وقدره:

$$\boxed{n = n_1 \times n_2 \times \ldots \times n_r}$$

ملاحظة:

في طريقة الضرب يتم إجراء جميع المراحل معًا لإتمام العملية.

مثال (٥-١):

بكم طريقة مختلفة يمكن أن يختار أحد الطلاب ثلاثة مقررات: الأول في الإحصاء والثاني في الرياضيات والثالث في الفيزياء إذا علم أن هناك 3 مقررات مختلفة للإحصاء و 2 مقررين مختلفين للرياضيات و 2 مقررين مختلفين للفيزياء.

الحل:

العملية هي اختيار 3 مقررات وهي مكونه من ثلاث مراحل:

المرحلة الأول = اختيار مقرر الإحصاء وعدد طرق هذه المرحلة يساوي $n_1 = 3$

المرحلة الثانية = اختيار مقرر الرياضيات وعدد طرق هذه المرحلة يساوي $n_2 = 2$

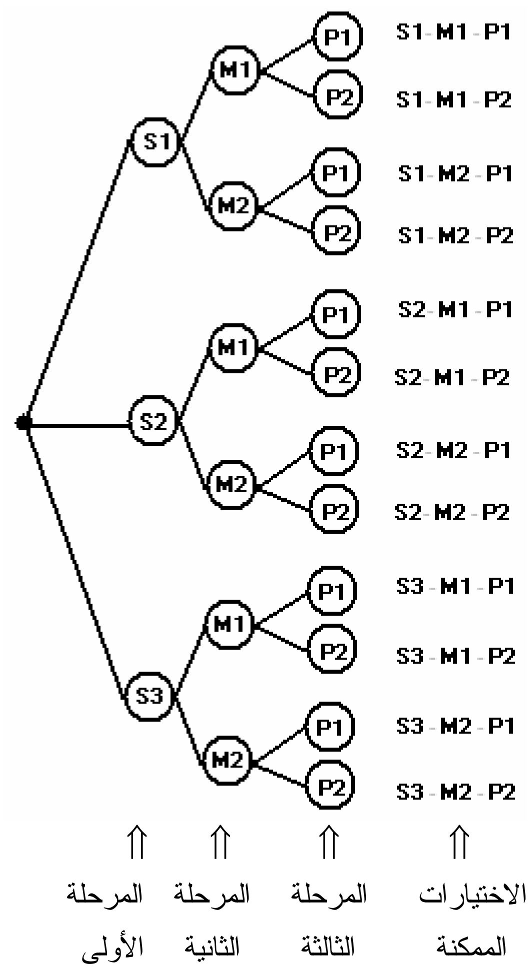
المرحلة الثالثة = اختيار مقرر الفيزياء وعدد طرق هذه المرحلة يساوي $n_3 = 2$

وباستخدام قاعدة الضرب فإن عدد الطريق المختلفة لاختيار المقررات الثلاثة يساوي:

$$n = n_1 \times n_2 \times n_3$$

$$= 3 \times 2 \times 2 = 12 \quad \text{(طريقة مختلفة)}$$

ويمكن توضيح الحل للمثال السابق باستخدام ما يسمى بشكل الشجرة البيانية كما يلي:



ثانياً: قاعدة الجمع:

نتيجة:

إذا كان هناك عدد مقداره r من العمليات بحيث:

- العملية رقم 1 تتم بعدد قدره n_1 من الطرق المختلفة
- العملية رقم 2 تتم بعدد قدره n_2 من الطرق المختلفة
- وهكذا ...

- العملية رقم r تتم بعدد قدره n_r من الطرق المختلفة

فإن عدد الطرق المختلفة لإجراء عملية واحدة فقط من هذه العمليات (العمليات متنافية) يساوي:

$$\boxed{n = n_1 + n_2 + \ldots + n_r}$$

ملاحظة:

في طريقة الجمع تكون العمليات متنافية، أي أن إجراء إحدى العمليات ينفي (أو يمنع) إجراء العمليات الأخرى.

مثال (٥-٢):

بكم طريقة مختلفة يمكن أن يختار أحد الطلاب مقررًا واحدًا فقط من الإحصاء **أو** الرياضيات **أو** الفيزياء إذا علم أن هناك 3 مقررات مختلفة للإحصاء و 2 مقررين مختلفين للرياضيات و 2 مقررين مختلفين للفيزياء.

الحل:

العملية الأول = اختيار مقرر الإحصاء وعدد طرق إجراء هذه العملية يساوي $n_1 = 3$

العملية الثانية = اختيار مقرر الرياضيات وعدد طرق إجراء هذه العملية يساوي $n_2 = 2$

العملية الثالثة = اختيار مقرر الفيزياء وعدد طرق إجراء هذه العملية يساوي $n_3 = 2$

وحيث أن العمليات متنافية وباستخدام قاعدة الجمع فإن عدد الطريق المختلفة لاختيار المقرر يساوي:

$$n = n_1 + n_2 + n_3$$
$$= 3 + 2 + 2 = 7$$ (طريق مختلفة)

(٥-٣) التباديل Permutations:

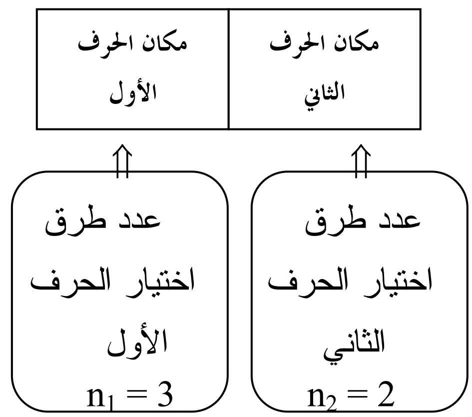
التبديلة هي ترتيبة لعدة أشياء بأخذها كلها أو بعضها في كل مرة مع مراعاة الترتيب. فعدد التباديل لمجموعة مكونة من n من الأشياء مأخوذًا r منها في كل مرة يساوي عدد الترتيبات المختلفة التي يمكن تكوينها من n من الأشياء بحيث تحوي كل ترتيبة على r من هذه الأشياء مع مراعاة الترتيب.

مثال (٥-٣):

١. كم عدد تباديل الحروف A, B, C بحيث تحوي كل ترتيبة على حرفين؟ أو بعبارة أخرى بكم طريقة يمكن ترتيب حرفين من الحروف A, B, C؟
٢. أوجد التباديل (التراتيب) المختلفة لحرفين من الحروف A, B, C.

الحل:

(١) عدد تباديل الحروف A, B, C بحيث تحوي كل ترتيبة على حرفين يمكن إيجاده بالتوضيح التالي:

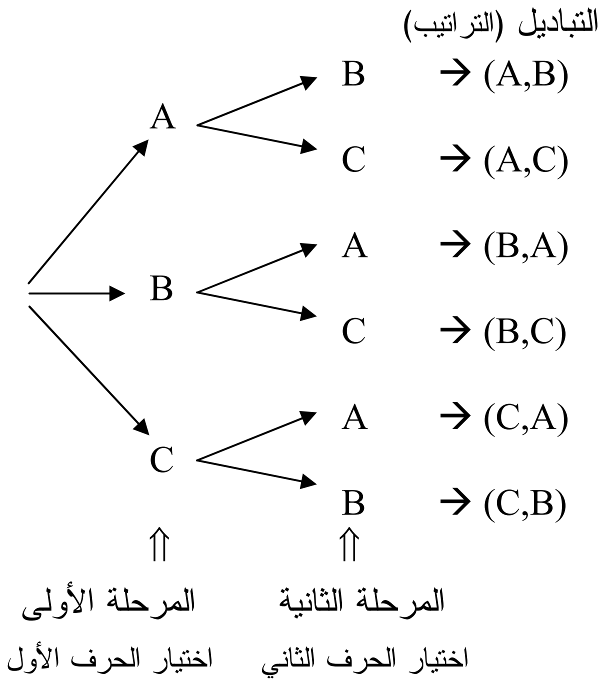


العملية مكونة من مرحلتين وبناءً على قاعدة الضرب فإن عدد طريق ترتيب حرفين من الحروف A,B,C يساوي:

$n = n_1 \times n_2 = 3 \times 2 = 6$

وعليه فإن عدد التباديل للحروف A,B,C مأخوذاً حرفين في كل مرة يساوي 6 تباديل.

(٢) يمكن إيجاد التباديل (أو التراتيب) باستخدام شكل الشجرة التالي:



لاحظ أن التبديلة أو الترتيبة (A,B) تختلف عن التبديلة (B,A)

نتيجة:

عدد تباديل n من الأشياء المختلفة مأخوذة r في كل مرة يرمز له بالرمز ${}_nP_r$ ويعطى بالصيغة التالية:

$$\boxed{{}_nP_r = n \times (n-1) \times (n-2) \times \ldots \times (n-r+1)}$$

ملاحظات:

- $n! = n \times (n-1) \times (n-2) \times \ldots \times 2 \times 1$ (يسمى مضروب العدد)
- $0! = 1$
- ${}_nP_r = n \times (n-1) \times (n-2) \times \ldots \times (n-r+1)$

$$= \frac{n(n-1)(n-2)....(n-r+1)\ (n-r)!}{(n-r)!} = \frac{n!}{(n-r)!}$$

$$\boxed{{}_nP_r = \frac{n!}{(n-r)!}}$$

- ${}_nP_n = n!$ (r=n)
- ${}_nP_1 = n$ (r=1)

مثال (٥-٤):

$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$

${}_5P_5 = 5! = 120$

${}_5P_1 = 5$

$${}_7P_3 = \frac{7!}{(7-3)!} = \frac{7!}{4!} = \frac{7\times6\times5\times4\times3\times2\times1}{4\times3\times2\times1} = \frac{5040}{24} = 210$$

$${}_7P_3 = \frac{7!}{(7-3)!} = \frac{7!}{4!} = \frac{7\times6\times5\times4!}{4!} = 7\times6\times5 = 210$$

مثال (٥-٥):

باستخدام قانون التباديل أوجد عدد التباديل المختلفة لحرفين من الحروف A, B, C. أو بعبارة أخرى بكم طريقة يمكن ترتيب حرفين من الحروف A, B, C؟

الحل:

لدينا n=3 و r=2 وعليه فإن عدد طرق ترتيب حرفين من الحروف A, B, C يساوي:

$${}_nP_r = {}_3P_2 = \frac{3!}{(3-2)!} = \frac{3!}{1!} = \frac{3\times2\times1}{1} = 6$$

مثال (٥-٦):

١. بكم طريقة يمكن أن نجلس 5 طلاب على 5 مقاعد في صف واحد؟

٢. بكم طريقة يمكن أن نجلس 5 طلاب على 3 مقاعد في صف واحد؟

الحل:

١. $n = 5,\ r = 5$

${}_nP_r = {}_5P_5 = 5! = 120$

٢. $n = 5,\ r = 3$

$${}_nP_r = {}_5P_3 = \frac{5!}{(5-3)!} = \frac{5!}{2!} = \frac{5\times4\times3\times2!}{2!} = 60$$

المقعد رقم 5	المقعد رقم 4	المقعد رقم 3	المقعد رقم 2	المقعد رقم 1
⇑	⇑	⇑	⇑	⇑
1	2	3	4	5

(٥-٤) تطبيقات على التباديل:

هناك تطبيقات كثيرة للتباديل نذكر بعضًا منها لأهميتها في دراسة علم الاحتمالات. وعلى سبيل التحديد فقد نكون مهتمين بإيجاد عدد طرق اختيار عنصر ما من مجموعة من العناصر والأمثلة على ذلك كثيرة. فقد نكون مهتمين بعدد الطرق المختلفة لسحب عينة من إنتاج أحد المصانع لفحصها. ويمكن تمثيل هذه العملية بعملية سحب r كرة من صندوق يحوي n كرة. كما تجدر الإشارة إلى أن عملية السحب تتم بطريقتين مختلفتين: الأولى تسمى السحب بإرجاع (بإحلال أو بإعادة) والثانية تسمى السحب بدون إرجاع (بدون إحلال أو بدون إعادة).

أولاً: السحب بإرجاع:

نتيجة:

إذا كان لدينا مجموعة مكونة من n من العناصر المختلفة وأردنا سحب r عنصر من هذه المجموعة بإرجاع (أي أن العنصر المسحوب يعاد مرة أخرى للمجموعة قبل سحب العنصر التالي) فإن عدد الطرق المختلفة التي يتم بها هذا السحب هو:

$$\boxed{\underbrace{n \times n \times \ \dots \times n}_{r \text{ من المرات}} = n^r}$$

سحب العنصر رقم r	. . .	سحب العنصر رقم 2	سحب العنصر رقم 1
$\Uparrow$ n	. . .	$\Uparrow$ n	$\Uparrow$ n

مثال (٥-٧):

بكم طريقة يمكن سحب كرتين بإرجاع من صندوق يحتوي على 15 كرة مختلفة؟

الحل:

لدينا $n=15$ و $r=2$ وعليه فإن عدد سحب كرتين بإرجاع من صندوق يحتوي على 15 كرة مختلفة يساوي:

$$n^r = 15^2 = 15 \times 15 = 225$$

ثانياً: السحب بدون إرجاع:

نتيجة:

إذا كان لدينا مجموعة مكونة من n من العناصر المختلفة وأردنا سحب r عنصر من هذه المجموعة بدون إرجاع (أي أن العنصر المسحوب لا يعاد مرة أخرى للمجموعة قبل سحب العنصر التالي) فإن عدد الطرق المختلفة التي يتم بها هذا السحب هو:

$$\boxed{{}_{n}P_{r} = n \times (n-1) \times (n-2) \times \ldots \times (n-r+1)}$$

سحب العنصر رقم r	. . .	سحب العنصر رقم 2	سحب العنصر رقم 1
$\Uparrow$ $n-r+1$	. . .	$\Uparrow$ $n-1$	$\Uparrow$ n

<u>مثال (٥-٥):</u>

بكم طريقة يمكن سحب كرتين بدون إرجاع من صندوق يحتوي على 15 كرة مختلفة؟

<u>الحل:</u>

لدينا $n=15$ و $r=2$ وعليه فإن عدد سحب كرتين بدون إرجاع من صندوق يحتوي على 15 كرة مختلفة يساوي:

$${}_{n}P_{r} = {}_{15}P_{2} = \frac{15!}{(15-2)!} = \frac{15!}{13!} = \frac{15 \times 14 \times 13!}{13!} = 15 \times 14 = 210$$

<u>(٥-٣) التوافيق (أو التواليف) Combinations:</u>

التوفيقة (أو التوليفة) هي كل مجموعة يمكن اختيارها من مجموعة من عدة أشياء بأخذها كلها أو بعضها دون مراعاة الترتيب.

<u>نتيجة:</u>

إذا كان لدينا مجموعة مكونة من n من العناصر المختلفة فإن عدد التوافيق التي يمكن تكوينها بحيث تحوي كل توفيقة على r عنصر يرمز له بالرمز $\binom{n}{r}$ أو بالرمز ${}_{n}C_{r}$ ويعطى بالصيغة التالية:

$$\boxed{\binom{n}{r} = {}_{n}C_{r} = \frac{{}_{n}P_{r}}{r!} = \frac{n!}{r!\,(n-r)!} \; ; \quad r = 0, 1, \ldots, n}$$

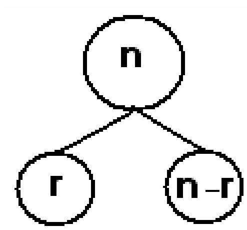
<u>ملاحظة:</u>

يمكن النظر إلى العدد $\binom{n}{r}$ على أنه:

$\binom{n}{r}$ = عدد التوافيق التي يمكن تكوينها من n عنصر مختلف بحيث تحتوي كل توفيقة على r عنصر.

= عدد طرق اختيار r عنصر من مجموعة مكونة من n عنصر مختلف.

= عدد طرق تقسيم مجموعة مكونة من n عنصر مختلف إلى مجموعتين الأولى تحوي r عنصرًا والأخرى تحوي (n–r) عنصرًا لباقية.



مثال (٥–٩):

$$\binom{5}{3} = \frac{5!}{3!\,(5-3)!} = \frac{5!}{3!\times 2!} = \frac{5\times 4\times 3!}{3!\times 2!} = 10$$

$$\binom{4}{0} = \frac{4!}{0!\,(4-0)!} = \frac{4!}{1\times 4!} = \frac{4!}{4!} = 1$$

$$\binom{7}{7} = \frac{7!}{7!\,(7-7)!} = \frac{7!}{7!\times 0!} = \frac{7!}{7!\times 1} = 1$$

$$\binom{7}{1} = \frac{7!}{1!\,(7-1)!} = \frac{7!}{1\times 6!} = \frac{7!}{6!} = \frac{7\times 6!}{6!} = 7$$

ملاحظات:

١. $\binom{n}{0} = 1 \quad , \binom{n}{n} = 1 \quad , \binom{n}{1} = n$

٢. $\binom{n}{r} = \frac{n!}{r!\,(n-r)!} = \binom{n}{n-r}$

مثال (٥–١٠):

بكم طريقة يمكن اختيار حرفين (بدون مراعاة الترتيب) من مجموعة الحروف A, B, C؟

الحل:

لدينا $n=3$ و $r=2$ وعليه فإن عدد طريقة اختيار حرفين (بدون مراعاة الترتيب) من مجموعة الحروف A, B, C يساوي:

$$\binom{n}{r}=\binom{3}{2}=\frac{3!}{2!\,(3-2)!}=\frac{3!}{2!\times 1!}=3$$

- نلاحظ أن الاختيارات (التوافيق أو التواليف) الممكنة هي: $\{A,B\},\ \{A,C\},\ \{B,C\}$.
- كما تجدر الإشارة إلى أن التوفيقة أو التوليفة $\{A,B\}$ هي نفس التوفيقة $\{B,A\}$.

(٥-٦) التباديل داخل أشياء متشابهة (متساوية):

نتيجة:

إذا كان هناك n من الأشياء مكونة من r مجموعة بحيث:

- المجموعة رقم 1 مكونة من n_1 من العناصر والمتشابهة
- المجموعة رقم 2 مكونة من n_2 من العناصر والمتشابهة
- وهكذا ...
- المجموعة رقم r مكونة من n_r من العناصر والمتشابهة

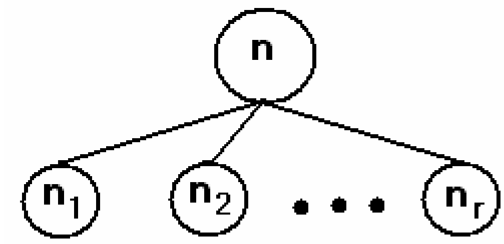
وكان $n= n_1 + n_2 + \dots + n_r$ فإن عدد التباديل المختلفة الممكنة لهذه الأشياء يساوي:

$$\boxed{\binom{n}{n_1,n_2,\cdots,n_r}=\frac{n!}{n_1!\ n_2!\ \cdots\ n_r!}}$$

ملاحظة:

يمكن النظر إلى العدد $\binom{n}{n_1,n_2,\cdots,n_r}$ على أنه:

$\binom{n}{n_1,n_2,\cdots,n_r}$ = عدد طرق تقسيم مجموعة مكونة من n عنصر مختلف إلى r مجموعة بحيث أن المجموعة رقم 1 تحتوي على n_1 عنصر و المجموعة رقم 2 تحتوي على n_2 عنصر وهكذا... والمجموعة رقم r تحتوي على n_r عنصر بحيث أن $n= n_1 + n_2 + \dots + n_r$.



مثال (٥-١١):

بكم طريقة يمكن ترتيب أحرف كلمة PROBABILITY؟

الحل:

P, R, O, B, B, A, I, I, L, T, Y

عدد تباديل أوعدد طرق ترتيب أحرف كلمة PROBABILITY يساوي:

$$\binom{11}{1,1,1,2,1,2,1,1,1} = \frac{11!}{2!\ 2!} = 9979200$$

مثال (٥-١٢):

بكم طريقة يمكن ترتيب أحرف كلمة STATISTICS؟

الحل:

S, S, S, T, T, T, A, I, I, C

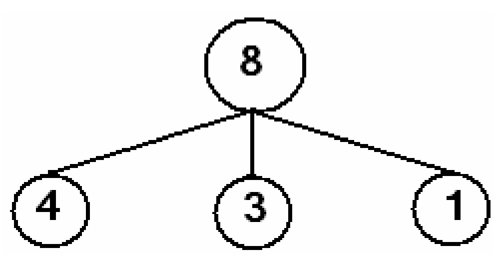
عدد تباديل أوعدد طرق ترتيب أحرف كلمة STATISTICS يساوي:

$$\binom{10}{3,3,2,1,1} = \frac{10!}{3!\ 3!\ 2!} = 50400$$

مثال (٥-١٣):

بكم طريقة يمكن توزيع 8 طلاب على النحو التالي: 4 طلاب لتخصص الإحصاء و 3 طلاب لتخصص الرياضيات و طالب واحد لتخصص الفيزياء.

الحل:



عدد الطرق لتوزيع الطلاب وفق الطريقة المذكورة يساوي:

$$\binom{8}{4,3,1} = \frac{8!}{4!\ 3!\ 1!} = 280$$