

الجمهورية الجزائرية الديمقراطية الشعبية

People's Democratic Republic of Algeria

وزارة التعليم العالي والبحث العلمي

Ministry of Higher Education and Scientific Research

University of Algiers 3

Faculty of Economic, Commercial
and Management Sciences



جامعة الجزائر 3

كلية العلوم الاقتصادية والعلوم التجارية وعلوم التسيير

مطبوعة موجهة لطلبة السنة أولى جذع مشترك

الإحصاء 1

دروس وتمارين

اعداد: د. حليلة عز الدين

السنة الجامعية: 2026-2027

الفهرس

1المقدمة

الفصل الأول: مفاهيم أساسية حول علم الإحصاء والبيانات

4.....تمهيد:

4..... مفهوم علم الإحصاء وأهميته:

4.....نشأة علم الإحصاء وتطوره:

6.....تعريف علم الإحصاء:

7..... تطبيقات علم الإحصاء في الاقتصاد وإدارة الأعمال

8.....(1)التطبيقات الكلاسيكية لعلم الإحصاء:

8.....(2)التطبيقات الحديثة لعلم الإحصاء:

10 مفهوم بعض المصطلحات الإحصائية:

12 أنواع البيانات الإحصائية:

12(1)المتغيرات الكيفية (الوصفية أو النوعية):

13(2)المتغيرات الكمية:

14 مصادر جمع البيانات:

14 المصادر المباشرة(الأولية):

15 المصادر غير المباشرة(الثانوية):

15 أسلوب جمع البيانات:

16 منهجية البحوث الإحصائية:

18 تمارين للفصل الأول:

الفصل الثاني: العرض الجدولي والبياني للبيانات الإحصائية

24 تمهيد

24 أولاً: العرض الجدولي للبيانات الإحصائية

24 أساسيات العرض الجدولي:

25 عناصر الجداول الإحصائية:

25 العرض الجدولي للمتغيرات الكمية

26 التوزيع التكراري للمتغير الكمي المنقطع

27 التوزيع التكراري للمتغير الكمي المستمر

29 جدول التوزيع التكراري النسبي والمئوي

31 التوزيع التكراري المتجمع

33 ثانياً: العرض البياني للبيانات الإحصائية (Graphical Presentation)

34 (a) العرض البياني للمتغيرات الكيفية

34 الأعمدة البيانية (Bar Chart):

35 الأعمدة البيانية المتلاصقة (Multiple Bar Graph):

36 الأعمدة البيانية المجزأة (Segmented/Stacked Bar Chart):

36 الدوائر النسبية (Pie Chart):

38	العرض البياني للمتغيرات الكمية
38	العرض البياني للمتغير الكمي المنقطع:
39	(b) العرض البياني للمتغير الكمي المستمر:
44	تمارين الفصل الثاني

الفصل الثالث: مقاييس النزعة المركزية

50	تمهيد
51	مفهوم النزعة المركزية
52	قياس النزعة المركزية
52	الوسط الحسابي (Moyenne arithmétique)
62	الوسط الهندسي (Moyenne Géométrique)
66	الوسط التوافقي (Moyenne Harmonique)
69	الوسط التربيعي (Moyenne Quadratique)
71	الوسيط (Médiane)
79	المنوال (Mode)
87	تمارين الفصل الثالث

الفصل الرابع: مقاييس التشتت

96	تمهيد
96	أولاً: المدى (Range)

96	(a) المدى في حالة البيانات غير المبوبة.....
97	المدى في حالة البيانات المبوبة.....
97	المدى في حالة متغير كمي منفصل:
98	المدى في حالة متغير كمي متصل:
98	مزايا وعيوب المدى
99	ثانياً: المدى الربيعي والانحراف الربيعي
99	(a) المدى الربيعي (Inter-Quartile Range)
100	الانحراف الربيعي (Quartile Deviation)
100	مزايا وعيوب الانحراف الربيعي
101	ثالثاً: الانحراف المتوسط (Mean Deviation)
101	(a) الانحراف المتوسط في حالة البيانات غير المبوبة.....
102	الانحراف المتوسط في حالة البيانات المبوبة.....
102	الانحراف المتوسط لمتغير كمي منفصل:
103	الانحراف المتوسط لمتغير كمي متصل:
104	مزايا وعيوب الانحراف المتوسط.....
104	رابعاً: التباين والانحراف المعياري (Variance and Standard Deviation)
104	(a) التباين والانحراف المعياري في حالة البيانات غير المبوبة.....
106	التباين والانحراف المعياري في حالة البيانات المبوبة.....

106	في حالة متغير كمي منفصل: توجد طريقتان لحساب التباين والانحراف المعياري:
107	في حالة متغير كمي متصل: تُطبَّق نفس العلاقتين السابقتين على مراكز الفئات:
108	خصائص الانحراف المعياري
108	مزايا وعيوب الانحراف المعياري
109	خامساً: معامل الاختلاف ومعامل الاختلاف الربيعي
109	(Coefficient of Variation) معامل الاختلاف
110	معامل الاختلاف الربيعي (Quartile Coefficient of Variation)
111	تمارين الفصل الرابع

الفصل الخامس: مقاييس الشكل

116	تمهيد
117	أولاً: العزوم (Moments)
117	العزوم البسيطة (Simple Moments)
119	العزوم المركزية
121	العلاقة بين العزوم المركزية والبسيطة
123	ثانياً : الالتواء (Skewness)
124	(Pearson's Coefficient of Skewness) معامل بيرسون للالتواء
127	معامل الالتواء يول وكيندال (معامل الالتواء الربيعي) (Yule & Kendall Coefficient)
129	معامل الالتواء فيشر (Fisher Skewness Coefficient)

130	 ثالثاً: التفرطح (Kurtosis)
131	(a)معامل التفرطح بيرسون (Pearson Kurtosis Coefficient)
132	معامل التفرطح فيشر (Fisher Kurtosis Coefficient)
132	معامل التفرطح كيلي (Kelly Kurtosis Coefficient)
136	 تمارين الفصل الخامس

الفصل السادس: الارتباط والانحدار الخطي البسيط

144	 تمهيد
145	 أولاً: جداول التوزيعات الثنائية (Bivariate Distributions)
145	 الجدول التكراري المشترك
146	 الجداول التكرارية الهامشية
147	 الجداول التكرارية الشرطية
148	التباين المشترك (Covariance)
149	 ثانياً: معامل الارتباط الخطي بيرسون (Pearson's Correlation Coefficient)
150	(a)معامل الارتباط بيرسون للبيانات غير المبوبة..
151	معامل الارتباط بيرسون للبيانات المبوبة (طريقة الانحرافات المختصرة)
154	 ثالثاً: معامل الارتباط الرتبي سبيرمان (Spearman's Rank Correlation)
157	 رابعاً: معامل الاقتران فاي (Coefficient de Contingence Phi)
159	 خامساً: معامل التوافق (Coefficient de Contingence)

سادساً: الانحدار الخطي البسيط (Régression Linéaire Simple)	161
(a) مفهوم الانحدار الخطي ومعادلة خط الانحدار ..	161
شكل العلاقة بين X و Y حسب إشارة معامل الانحدار	162
مثال تطبيقي شامل على الانحدار الخطي البسيط	162
الطريقة البيانية لتقدير معلمتي نموذج الانحدار	164
قياس قوة الانحدار ودراسة صلاحية النموذج	164
تمارين الفصل السادس	167

الفصل السابع: الأرقام القياسية

مفهوم الأرقام القياسية	175
أولاً: الرقم القياسي البسيط (I)	176
الرقم القياسي البسيط للأسعار	176
الرقم القياسي البسيط للكميات	178
ثانياً: الرقم القياسي التجميعي (ΣI)	180
ثالثاً: خصائص الأرقام القياسية	181
المعيار الأول: خاصية الانعكاس (Propriété de réversibilité)	181
المعيار الثاني: خاصية التحويل أو الدوران (Propriété de circularité)	182
رابعاً: الأرقام القياسية الترجيحية (Indice pondéré)	184
(a) الرقم القياسي لاسبير (Laspeyres)	184

185	(Paasche) الرقم القياسي باش
185	(Fisher) الرقم القياسي فيشر
186	 (Marshall-Edgeworth) الرقم القياسي مارشال-إدجورث
188	تمارين الفصل السابع
193	 خاتمة عامة
194	 قائمة المصادر

المقدمة

المقدمة

الحمد لله رب العالمين، والصلاة والسلام على خاتم النبيين وسيد المرسلين، نبينا محمد الهادي الأمين، الذي بعثه الله رحمة للعالمين، وعلى آله وصحبه وأنصاره وأتباعه، ومن اهتدى بهديه وعمل بسنته إلى يوم الدين.

وبعد:

يسرني أن أضع بين يدي الطالب هذه المطبوعة التي جاءت ثمرة سنوات من تدريس مقياس الإحصاء الوصفي لطلبة العلوم الاقتصادية والتجارية وعلوم التسيير، حرصنا من خلالها على تقديم المادة العلمية المقررة في هذا المقياس بأسلوب واضح ومنهجية متدرجة تراعي المستوى المعرفي لطلبة السنة الأولى جـ مشترك. فالإحصاء علم لا غنى عنه لأي طالب أو باحث في شتى الميادين والمستويات، إذ يمثل الأداة المنهجية التي تعينه على فهم الظواهر التي يدرسها، وتفسير نتائجها، واستخلاص دلالات البيانات والأرقام التي يتحصل عليها.

وبهدف تيسير استيعاب الطلبة لهذا المقياس وتذليل الصعوبات التي قد تواجههم، ارتأينا إعداد هذه المطبوعة في صورة سلسلة من المحاضرات في الإحصاء الوصفي، موجهة أساساً لطلبة السنة الأولى جـ مشترك علوم اقتصادية وتجارية وعلوم التسيير، نقدم من خلالها دروساً مبسطة ومركزة سهلة الفهم، مدعمة بأمثلة وتمارين متنوعة تُعرض حلولها النموذجية في نهاية كل فصل، تثبيتاً للمفاهيم وتدريباً للطالب على التطبيق العملي.

وقد قسّمنا محتوى هذه المطبوعة إلى سبعة فصول متكاملة: يتناول الفصل الأول المفاهيم الأساسية لعلم الإحصاء ومصطلحاته الأولية، ويعرض الفصل الثاني أساليب عرض البيانات الإحصائية جدولياً وبيانياً، في حين يُخصّص الفصل الثالث لمقاييس النزعة المركزية، ويتناول الفصل الرابع مقاييس التشتت، ويعالج الفصل الخامس مقاييس الشكل (الالتواء والتفرطح)، أما الفصل السادس فيتناول الارتباط والانحدار الخطي البسيط، ويختتم الفصل السابع هذه المطبوعة بدراسة الأرقام القياسية وتطبيقاتها.

وإننا إذ نضع هذا العمل بين يدي الطلبة، ندرك أنه كأى إنتاج علمي بشري لا يخلو من النقص والهفوات، وكل أملنا أن يسهم في تيسير هذا المقياس وتطوير مسيرة البحث العلمي لدى طلبتنا. ونسأل الله العليّ القدير أن يجعل هذا العمل خالصاً لوجهه الكريم، وأن يجعله في ميزان حسناتنا، وأن ينفع به الطلاب والدارسين

الفصل الأول

مفاهيم أساسية حول علم الإحصاء والبيانات

الفصل الأول: مفاهيم أساسية حول علم الإحصاء والبيانات

تمهيد:

يعد علم الإحصاء من الأدوات الأساسية في دراسة الظواهر المختلفة وتحليلها تحليلًا علميًا دقيقًا لما يوفره من وسائل منهجية لجمع البيانات وتنظيمها وتفسيرها، وقد ساهم التطور العلمي والتكنولوجي في توسيع مجالات استخدام الإحصاء ليشمل مختلف الميادين الاقتصادية والاجتماعية والتربوية، ويهدف هذا الفصل إلى تقديم المفاهيم الأساسية المتعلقة بعلم الإحصاء، مع إبراز مكانة الإحصاء الوصفي وأهميته في فهم الظواهر المدروسة، كما أن هذا الفصل يمهد لبقية الفصول من خلال توضيح المفاهيم والمصطلحات الضرورية للتحليل.

مفهوم علم الإحصاء وأهميته:

نشأ علم الإحصاء استجابة لحاجة الإنسان إلى تنظيم المعطيات وفهم الظواهر، ثم تطور ليشمل مجالات علمية وعملية متعددة، وتكمن أهميته في كونه أداة أساسية لتحليل البيانات ودعم البحث العلمي واتخاذ القرارات.

نشأة علم الإحصاء وتطوره:

إن كلمة الإحصاء وردت في القرآن الكريم في عدة آيات كريمات نذكر منها سورة مريم و سورة إبراهيم وسورة الكهف، وعند التفكير في ما جاء بالآيات الكريمات نجد بأن للإحصاء علاقة وثيقة بدراسة المجتمع ودراسة أجزاء منه بهدف دراسة أهم نشاطاته، ومن جانب آخر الإحصاء هو علم حصر وعد وهو علم المتوسطات والمتغيرات وعلم التقديرات والاحتمالات، وأول من استخدم الإحصاء قديما هم قدماء المصريين من الفراعنة، فهم أول من قام باستخدام وتطبيق المفاهيم البدائية للإحصاء في بناء الأهرامات وتصميمها، إذ قاموا بإجراء تعداد السكان في مصر وحصر ثرواتها واعتماد النتائج التي حصلوا عليها في تنظيم مشروع

بناء الأهرامات وفي عصر الخلافة العباسية استخدم الخليفة العباسي المأمون فكرة الحصر والعد للوقوف على عدد السكان ومعرفة مقدار الزكاة.¹

وكلمة إحصاء مشتقة من كلمة "state" اللاتينية التي تعني *statuts* وهي مجموعة الحقائق والمعلومات عن الدولة، فمنذ القدم كان الإحصاء يُزوّد قادة الدول بالمعلومات حول الضرائب التي يجب فرضها على رعاياهم وعدد الجنود المتوقع في الحرب ولم يصبح الإحصاء علمًا حقيقيًا إلا بعد الثورات البرجوازية.

ومن أهم رواد هذا العلم جون جرانت (1620-1674) وويليام بيتي (1623-1687) ففي ظل الرأسمالية لم يقتصر الاهتمام بالدراسات الإحصائية على قادة الدول فحسب بل شمل أيضًا رواد الأعمال الرأسماليين حيث ازداد استخدام الأدوات الرياضية المتقدمة لمعالجة البيانات، مما زاد من الفوائد مثل التأمين ويعتمد التأمين الجيد على التقييم الدقيق والاستدلال الرياضي السليم ومنذ القرن السابع عشر تطورت الإحصاءات الرياضية تدريجياً لتصبح فرعاً مستقلاً من فروع الرياضيات، ولها هدفاً رئيسياً وهو الحصول على معلومات أكثر موثوقية حول التقييم والرصد والقياس.

فخلال القرن السابع عشر ظهر ما يعرف بوصف الإحصاءات ثم إلى تطور الرياضيات ونظرية الاحتمالات خلال القرن الثامن عشر والتاسع عشر وكذلك السجل المدني للمواليد كما عرف القرن العشرون تطور تكنولوجيا الحاسوب²، أصبح بذلك الإحصاء علماً متكاملًا يعتمد على الأساليب الرياضية والاستدلال الإحصائي ومع ظهور الحاسوب الإلكتروني والبرمجيات الإحصائية، أصبح بالإمكان معالجة كميات ضخمة من البيانات بسرعة ودقة، مما وسع استخدام الإحصاء في الاقتصاد، الصحة، التعليم، العلوم الاجتماعية، وضبط الجودة الصناعية، وفي القرن الواحد والعشرين ومع انتشار البيانات الضخمة والذكاء الاصطناعي والحوسبة السحابية، أصبح الإحصاء أداة مركزية لتحليل المعلومات واتخاذ القرارات الدقيقة في مختلف

¹ حسن ياسين طعمة، طرق الإحصاء الوصفي، دار صفاء للنشر والتوزيع، الأردن، 2009، ص 20.

² Hunyadi, L., Vita, L., Statistics introduction history subject and task of Statistics central statistical office, 2008, p15

المجالات، وهكذا تطور الإحصاء من مجرد تسجيل الأرقام ووصفها إلى علم متقدم يعتمد على التكنولوجيا والرياضيات لتحليل البيانات واستخلاص النتائج.

تعريف علم الإحصاء:

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

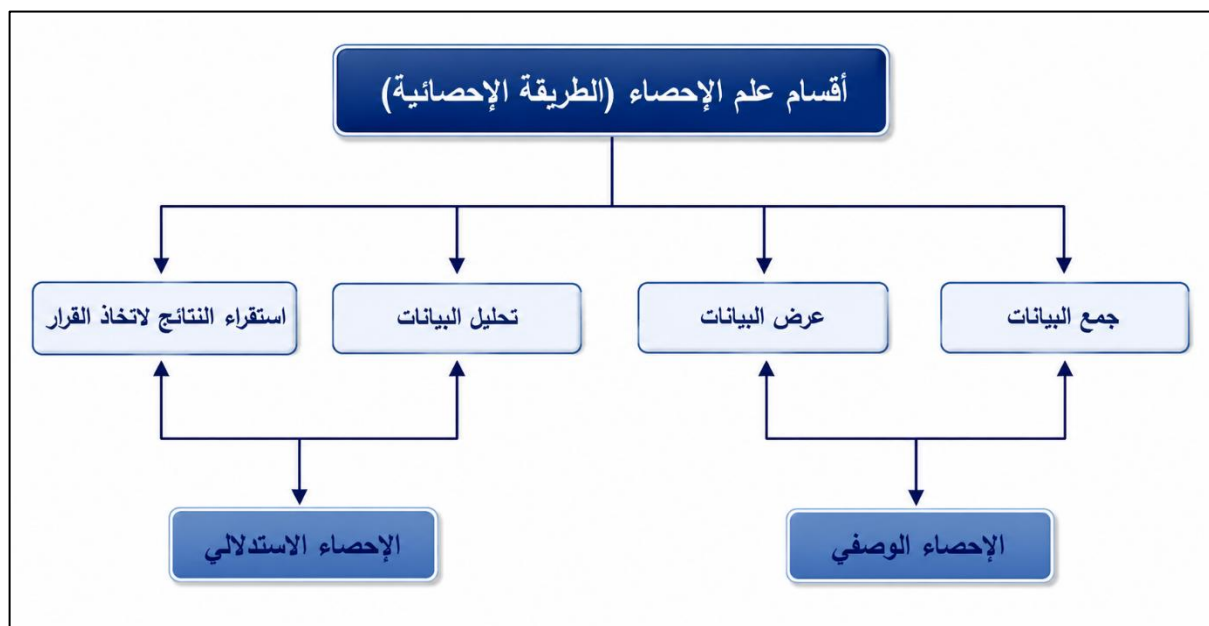
- فأما جمع البيانات فهو عملية الحصول على القياسات أو التعدادات أو قيم المشاهدات للتجارب التي يجريها الإحصائي وكلما كان جمع البيانات دقيقا كلما زادت ثقة الدارس بالاعتماد عليها ولا يكون هناك تحليل صحيح للبيانات إذا كانت هناك أخطاء في جمع تلك البيانات.
- أما تنظيم وعرض البيانات فهو عملية وضع البيانات في جداول منسقة وعرضها بطرق مناسبة كالأشكال الهندسية والرسوم البيانية والتوزيعات التكرارية.
- أما تحليل البيانات فهو عبارة عن إيجاد قيم لمقاييس واقتراحات معينة تتحدد قيمها من البيانات قيد الدراسة.
- أما استقراء النتائج واتخاذ القرارات فهو من أهم أهداف علم الإحصاء وأكثرها فائدة حيث يشمل معظم الدراسات الإحصائية والنظريات القائمة عليها والتطبيقات العلمية لها، وهو يتألف من الاستنتاجات التي يتوصل إليها الباحث من تحليل البيانات وهي غالبا ما تكون على شكل تقديرات أو تنبؤات أو تعميمات أو قرارات رفض أو قبول للفرضيات الإحصائية.

ويقسم علم الإحصاء إلى قسمين:

أ. الإحصاء الوصفي

ب. الإحصاء الاستقرائي (الاستدلالي)

أما **الإحصاء الوصفي** فيشمل جمع وتبويب البيانات الإحصائية وبيان مواصفاتها وأما **الإحصاء الاستقرائي** (الاستدلالي)، فيبحث في استقراء النتائج واتخاذ القرارات فهو يستخدم معلومات جزئية عن المجتمع لبيان مواصفات ذلك المجتمع، ففي الإحصاء الاستدلالي نقوم بسحب عينة جزئية من مجتمع كبير عادة ثم ندرس خصائص ظاهرة تهتمنا في هذه العينة ونحاول بعد هذا أن نستخلص خصائص تلك الظاهرة في المجتمع من خلال النتائج التي توصلنا إليها في العينة¹.



الشكل 1- 1: أقسام علم الإحصاء (الطريقة الإحصائية)

المصدر: أحمد عبد السميع طيبة، مبادئ الإحصاء، دار البداية للنشر والتوزيع، الأردن، 2008، ص13.

تطبيقات علم الإحصاء في الاقتصاد وإدارة الأعمال

يعد علم الإحصاء من الأدوات الأساسية في الاقتصاد وإدارة الأعمال إذ يستخدم في تحليل البيانات وتفسير الظواهر الاقتصادية ودعم اتخاذ القرارات الإدارية على أسس علمية دقيقة.

¹ نور الدين حامد، محاضرات وتمارين محلولة في الإحصاء الوصفي، دار اليازوري العلمية للنشر والتوزيع، الأردن، 2017، صص 6-8.

(1) التطبيقات الكلاسيكية لعلم الإحصاء:

- علم الاجتماع: يدخل في النظريات التي تصف الهجرة الداخلية ومدى تأثيرها على سلوك المجتمع وتصميم عمليات المعاينة التي تساعد على بناء واختبار نظريات النظم الاجتماعية، وتحليل تكاليف المساعدات والتأمينات الاجتماعية، وشرح الفرق في الاتجاه والسلوك بين جماعات من الناس تصميم وتحليل التجارب لوصف وشرح سلوك الجماعة.
- علم النفس: يدخل في نظريات التربية والمشاكل المتعلقة بقياسات القدرة على التعلم، الذكاء الصفات الشخصية السلوك الطبيعي والغير طبيعي للأشخاص وفي إيجاد مقاييس ومعايير لاستخدامها في هذه المجالات.
- السكان: يساعد في دراسة تطور مجتمع السكان عن طريق معدلات المواليد والوفيات ومعدلات الهجرة وكذلك دراسة معالم المجتمع الاجتماعية والاقتصادية والمهنية وغيرها.
- التعليم: يساهم في وضع خطط التعليم الحالية والمستقبلية وتقدير احتياجاتها من قوى بشرية ومباني ومعامل وأجهزة.
- الاقتصاد: يساعد على معرفة حجم التجارة، مصادر القوى العاملة، مستوى المعيشة، تحليل سلوك المنتج والمستهلك، ومدى تأثير السوق بتغير الأسعار والقوانين الحكومية.
- علم الأحياء: يستخدم في الأبحاث الأساسية والتجارب العملية الخاصة بتطور الحياة والوراثة.

من هذه الأمثلة يتبين لنا أهمية علم الإحصاء في كافة مجالات الحياة¹.

(2) التطبيقات الحديثة لعلم الإحصاء:

❖ أولاً- تحليل البيانات الضخمة (Big Data):

¹ جلال الصياد، مبادئ الإحصاء، دار حافظ للنشر والتوزيع، 2003، المملكة العربية السعودية، ص14

أحدث ظهور البيانات الضخمة ثورة في مجال الإحصاء إذ زود الباحثين وصانعي السياسات بكميات هائلة من البيانات الدقيقة والمتنوعة والحديثة وتركز تطبيقات البيانات الضخمة في الإحصاء على خمسة جوانب: إدارة العمليات الوحشية، وإدارة الموارد البشرية، وتحليل عمليات الصناديق، وإدارة المخاطر، وتخصيص الموارد الداخلية، وبهذا تلعب البيانات الضخمة دوراً مهماً في تحليل كميات كبيرة من المعلومات الاقتصادية والتجارية، إذ يمكن استخدام أدوات إحصائية متقدمة لاستخلاص الأنماط والعلاقات داخل هذه البيانات بغرض اتخاذ قرارات اقتصادية وإدارية أكثر دقة وكفاءة¹.

❖ ثانياً - الحوسبة السحابية (Cloud Computing):

توفر الحوسبة السحابية منصات مرنة لتخزين ومعالجة البيانات الضخمة، مما يمكن المؤسسات من استخدام القدرات الإحصائية الكبيرة دون الحاجة إلى بنية تحتية محلية مكلفة. هذا يسمح بتحليل البيانات في الوقت الحقيقي وتحسين الأداء المؤسسي².

❖ ثالثاً - تحليل الانحدار المتقدم واختبارات الفرضيات:

تساعد الأساليب الإحصائية مثل الانحدار المتعدد واختبارات الفرضيات في تحليل العلاقات المعقدة بين المتغيرات الاقتصادية ضمن مجموعات بيانات ضخمة. هذا يمكن من التنبؤ بسلوك الأسواق ودعم سياسات الأعمال³.

❖ رابعاً - التطبيقات العملية للبيانات الضخمة والحوسبة السحابية:

أمثلة على استخدام Big Data و Cloud Computing

- تحليل سلوك العملاء في التجارة الإلكترونية، لتحسين التسويق واستراتيجيات البيع.

¹ Sitong Zhu ,The Application of Big Data in Economic Statistics, 2024,p223-230

² Aws Abueid,Big Data and Cloud Computing Opportunities and Application Areas, 2024, p14509-14516

³ Sitong Zhu,opcit.

- التنبؤ المالي وتحليل المخاطر عبر نماذج إحصائية قوية.
- تحسين العمليات التشغيلية داخل الشركات من خلال تحليل بيانات الأداء اللحظي¹.

مفهوم بعض المصطلحات الإحصائية:

❖ المجتمع الإحصائي:

يعرف المجتمع بأنه مجموعة ذات خصائص مشتركة من الأفراد محل الدراسة والمجتمع قد يكون مجموعة من البشر أو مجموعة من الأشياء أو مجموعة من السلع في فترة زمنية أو مجموعة من طلاب جامعة ما وينقسم المجتمع الإحصائي إلى قسمين: القسم الأول هو عدد محدود من الأفراد في حالة حصر عدد أفرادهم مثل سكان ولاية، ما أو طلاب مرحلة معينة من مراحل التعليم أما القسم الثاني فهو عدد غير محدود من الأفراد، أي لا نهائي إذا كان لا يمكن حصر عدد أفرادهم مثل النجوم أو الكواكب أو الكائنات الحية.

❖ العينة الإحصائية:

هي جزء من المجتمع الإحصائي تُختار من المجتمع بحيث تشمل جميع صفات المجتمع وينفرد بها فرع خاص من علم الإحصاء يسمى "نظرية العينات" وقد تكون الحاجة ضرورية لأخذ العينة بدلا من دراسة المجتمع كله وبعدها يتم تعميم نتائج الدراسة على المجتمع بأكمله².

ونقتصر على ذكر ثلاثة أنواع من العينات:

(1) **العينة العشوائية:** تعتبر العينة العشوائية من أكثر العينات استعمالا وهي جزء لا على التعيين من المجتمع المدروس، وتخص العينة العشوائية المجتمعات المتجانسة أي تلك التي لها خصائص متشابهة أو متقاربة مثلا: مجتمع من الإطارات المتوسطة يشكل مجتمع

¹ Aws Abueid, opcit.

² تيلولت سامية، مبادئ في الإحصاء، دار الحديث للكتاب، الطبعة الثانية، الجزائر، 2014، ص8

متجانس من حيث المداخل المتحصل عليها، مجتمع من الطلبة يشكل مجتمعا متجانسا من حيث السن وهكذا... إذن تسحب العينة العشوائية من مجتمع متجانس.

(2) **العينة الطبقية:** تخص المجتمعات غير المتجانسة أي المجتمعات المتكونة من عدة فئات اجتماعية، مهنية وهذا بناء على عدة اعتبارات منها: مستوى الدخل، مستوى الانفاق، المستوى التعليمي، إن العينة الطبقية هي عبارة عن مجموعة من العينات العشوائية.

(3) **العينة ذات المراحل المتعددة:** تستعمل مثل هذه العينات عند دراسة المستوى التعليمي في بلد ما: مثلا لدراسة المستوى التعليمي للمرحلة النهائية للتعليم الثانوي نتبع الخطوات التالية:

- أ. نسحب ولاية بطريقة عشوائية من أصل مجموع الولايات.
- ب. نسحب دائرة من أصل الدوائر التي تتكون منها الولاية المسحوبة.
- ت. نسحب ثانوية من أصل الثانويات المتواجدة في الدائرة المختار.
- ث. يسحب قسم من الأقسام النهائية على مستوى الثانوية المسحوبة لتجرى عليه الدراسة¹.

❖ الوحدة الإحصائية:

هي كل عنصر من عناصر العينة الإحصائية مثل الطالب (وحدة إحصائية) في أحد الأقسام (عينة) من جامعة معينة (مجتمع)، وتطلق كلمة عنصر في علم الإحصاء للدلالة على الكائن الواحد، سواء أكان هذا الكائن إنسانا أو حيوانا أو شيئا، والوحدة الإحصائية في لغة الإحصاء هو كائن متحرك أو جامد يقاس بأحد المقاييس المعينة أو يعد عدًا.

❖ المتغيرة الإحصائية (الظاهرة الإحصائية أو الخاصية المدروسة):

¹ جلاطو جيلالي، الإحصاء مع تمارين ومسائل محلولة، ديوان المطبوعات الجامعية، الجزائر، 2001، ص 6.

هي مقدار له خصائص رقمية (خصائص كمية) أو غير رقمية (وصفية)، وتتغير قيمتها من عنصر لآخر من عناصر المجتمع أو العينة مثل: متغير طول الطالب أو ذكائه أو الإنتاج أو الوزن...، ونرمز للمتغيرة الإحصائية بالرمز x_i^j

أنواع البيانات الإحصائية:

تتقسم البيانات (المتغيرات) الإحصائية إلى قسمين:

1) المتغيرات الكيفية (الوصفية أو النوعية):

هي تلك المتغيرات أو الظواهر التي لا يمكن قياسها عددياً بل قياس تكراراتها فقط² ولا يمكن إجراء العمليات الحسابية عليها وتنقسم إلى نوعين وهما:

1) **المتغيرات الكيفية الرتيبة:** وهي عبارة عن متغيرات وصفية تصف عناصر في المجتمع، ولا يمكن إجراء العمليات الحسابية عليها ولكن يمكن وضع ترتيب منطقي لمفرداتها مثل التقديرات في مقياس معين.

2) **المتغيرات الكيفية الاسمية:** وهي عبارة عن متغيرات وصفية تصف عناصر في المجتمع ولا يمكن إجراء العمليات الحسابية عليها بل أيضاً لا يمكن ترتيبها في أي شكل كان مثل الحالة الاجتماعية للعمال في شركة ما، الجنسية اللون³، إلخ

¹ تيلولت سامية، مرجع سبق ذكره، ص 8.

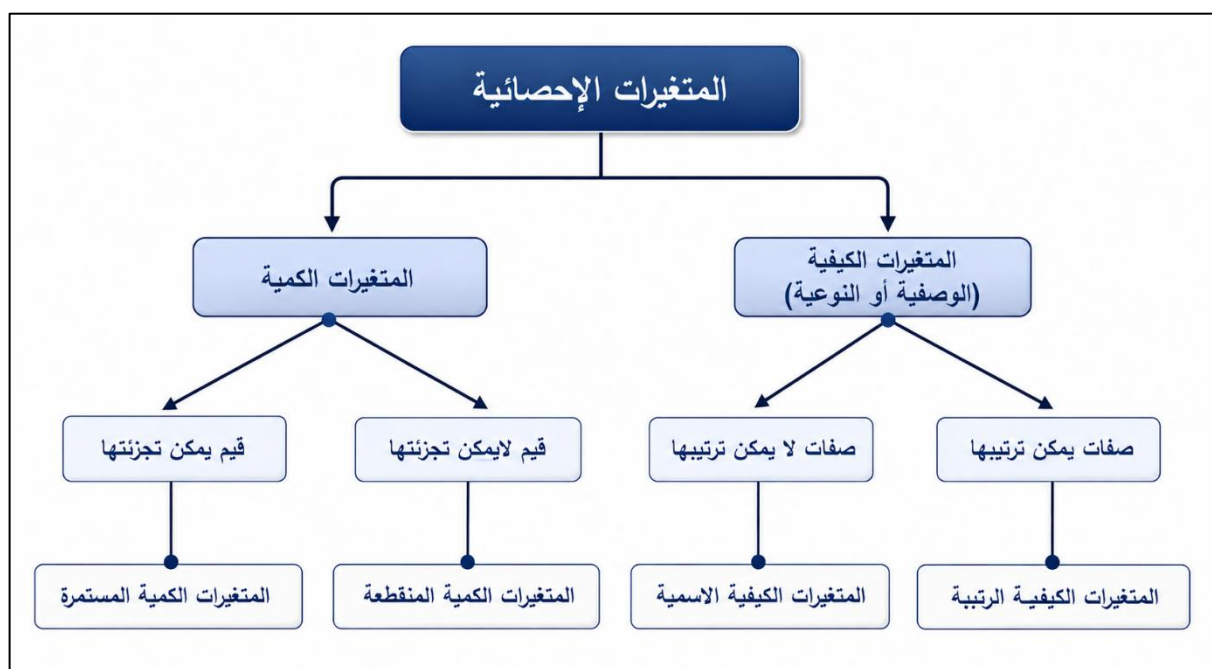
² William M. Mendenhall , statistics for engineering and the sciences, six th edition, Crc press Taylor and Frances group, new York, 2016, p07.

³ تيلولت سامية، مرجع سبق ذكره، ص 14.

2) المتغيرات الكمية:

وهي تلك الخصائص التي يمكن قياسها وهي أكثر المتغيرات انتشارا واستعمالا لأن لغة الإحصاء هي لغة الأرقام، مثل: الإنتاج، الاستهلاك، الاستثمار، الوزن، إلخ، وتنقسم المتغيرات الكمية إلى قسمين: متغيرات منقطعة ومتغيرات مستمرة.

- a. **المتغيرات المنقطعة (المنفصلة):** وهي تلك المتغيرات التي تأخذ قيما صحيحة لا يمكن تجزئتها مثل عدد الأطفال في العائلة، عدد قطع الغيار المنتجة... إلخ.
- b. **المتغيرات المستمرة (المتصلة):** وهي تلك المتغيرات التي تأخذ كل القيم الممكنة لمجال الدراسة، ونظرا للعدد غير المتناهي لهذه القيم نقسم مجال الدراسة إلى مجالات جزئية تسمى الفئات¹.



الشكل 1-2: أنواع المتغيرات الإحصائية

المصدر: من إعداد الباحثة

¹ جلاطو جيلالي، الإحصاء مرجع سبق ذكره، ص 6.

مصادر جمع البيانات:

غالباً ما تقسم مصادر جمع البيانات إلى قسمين وهما:

- أ. التقارير الرسمية التي تنشرها المؤسسات والجهات المخولة بذلك.
- ب. الأفراد والمؤسسات التي تقوم بجمع البيانات من ذوي العلاقة.

المصادر المباشرة (الأولية):

المصدر الأولي لجمع البيانات هو ذلك المصدر الذي يقوم بجمع البيانات بنفسه أو تحت إشرافه، مثل ما تقوم به دائرة الإحصاءات العامة من تسجيل للبيانات عن السكان ومن طرق المصادر المباشرة للبيانات:

- أ. المقابلات الشخصية المباشرة: حيث يقوم جامع البيانات بطرح الأسئلة ويجيب عليها الشخص المعني.
- ب. المقابلات الشخصية غير المباشرة: يقوم جامع البيانات بمقابلة شخص ثالث غير الشخص المطلوب منه البيانات.
- ت. المعلومات عن المراسلين: دورهم جمع البيانات وتقديمها للجهة المعنية بالدراسة.
- ث. عن طريق الهاتف: وهذه الطريقة شبيهة بالمقابلة الشخصية ولكن تكون الأسئلة والإجابات عن بعد عن طريق الهاتف.
- ج. الاستبيانات بالبريد أو الفاكس أو البريد الإلكتروني، حيث يرسل الباحث الاستبانة للأفراد المطلوب جمع البيانات منهم.
- ح. عن طريق الانترنت: حيث يتم طرح الموضوع قيد الدراسة على الانترنت فيجيب عليه الراغبون في المشاركة.

المصادر غير المباشرة (الثانوية):

عندما لا يستطيع الباحث من جمع البيانات بنفسه أو تحت إشرافه يلجأ إلى المصادر غير المباشرة أي إلى البيانات التي جمعها غيره: وهي معدة مسبقاً من طرف الجهات المعنية والتي غالباً ما تكون رسمية وتقسم هذه البيانات إلى قسمين وهما:

أ. المصادر المنشورة ومنها:

- التقارير والمنشورات الرسمية مثل تقارير دائرة الإحصاءات العامة أو البنك المركزي.
 - التقارير والمنشورات شبه الرسمية وهي تقارير ومنشورات تنشرها هيئات محلية مثل تقارير البلديات وغرف التجارة والصناعة المحلية.
 - التقارير والمنشورات الخاصة مثل تقارير ومنشورات الشركات والمؤسسات الخاصة.
- ب. المصادر غير المنشورة:

وهي بيانات غير منشورة لكنها مدونة في سجلات الهيئات ويكون في إمكان الباحث الرجوع إليها عند الحاجة مثل البيانات عن الموالى¹.

أسلوب جمع البيانات:

يتحدد الأسلوب المستخدم في جمع البيانات حسب الهدف من البحث وحجم المجتمع محل البحث، وهناك أسلوبين لجمع البيانات هما:

1) أسلوب الحصر الشامل:

يستخدم هذا الأسلوب إذا كان الغرض من البحث هو حصر جميع مفردات المجتمع وفي هذه الحالة يتم جمع بيانات عن كل مفردة من مفردات المجتمع بلا استثناء، كحصر جميع المزارع التي تنتج التمور

¹ سهيل أحمد سمحان، محمود حسين الوادي، مبادئ الإحصاء للاقتصاد والعلوم الإدارية، دار صفاء للنشر والتوزيع، الأردن، 2010، ص32.

أو حصر جميع البنوك الزراعية ويتميز أسلوب الحصر الشامل بالشمول وعدم التحيز ودقة النتائج ولكن يعاب عليه أنه يحتاج إلى الوقت والمجهود والتكلفة العالية.

(2) أسلوب المعاينة:

يعتمد هذا الأسبوع على معاينة جزء من المجتمع محل الدراسة، يتم اختياره بطريقة علمية سليمة ودراسته ثم تعميم نتائج العينة على المجتمع ومن ثم يتميز هذا الأسلوب بالآتي:

- تقليل الوقت والجهد.
- تقليل التكلفة.
- الحصول على بيانات أكثر تفصيلاً وخاصة إذا جمعت البيانات من خلال استمارة استبيان.
- كما أن أسلوب المعاينة يفضل في بعض الحالات التي يصعب فيها إجراء حصر شامل، مثل معاينة دم المريض، أو إجراء تعداد لعدد الأسماك في البحر، أو معاينة اللمبات الكهربائية، ولكن يعاب على أسلوب المعاينة أن النتائج التي تعتمد على هذا الأسلوب أقل دقة من نتائج أسلوب الحصر الشامل وخاصة إذا كانت العينة المختارة لا تمثل المجتمع تمثيلاً جيداً¹.

منهجية البحوث الإحصائية:

تعرف الطريقة الإحصائية بأنها الخطوات المتبعة في عمل أي دراسة أو بحث إحصائي وهذه الخطوات

هي:

¹ شرف الدين خليل، الإحصاء الوصفي، مكتبة الأنجلو المصرية، شبكة الأبحاث والدراسات، القاهرة، 2008، ص 12.

- **أولاً: جمع البيانات الإحصائية:** وهي قيم المشاهدات للتجارب التي يجريها الباحث وكلما كانت دقيقة كلما كانت القرارات المتخذة بصددتها أكثر صدقاً، وكما ذكرنا سابقاً هناك طريقتين لجمع البيانات الإحصائية هما: **طريقة المسح الشامل وطريقة المعاينة.**
- **ثانياً: تنظيم وعرض البيانات:** بعد جمع البيانات يقوم الباحث في وضع هذه البيانات في جداول مناسبة أو عرضها في رسوم بيانية أو أشكال هندسية أو توزيعات تكرارية.
- **ثالثاً: تحليل البيانات:** وهي معالجة البيانات باستعمال العلاقات الرياضية واستخراج قيم معينة تعبر عن هذه البيانات.
- **رابعاً: استقرار النتائج واتخاذ القرارات:** وهي إصدار الأحكام أو عمل الاستنتاجات الإحصائية حول المجتمع الإحصائي في ضوء النتائج المستخرجة¹.

¹ كامل فليل، فتحي حمدان، الإحصاء، دار المناهج للنشر والتوزيع: الأردن، 2012، ص 13.

تمارين للفصل الأول:

التمرين الأول:

1. ما المقصود بكلمة الإحصاء؟
2. ما هي مصادر البيانات الإحصائية؟
3. عرف: المجتمع الإحصائي، العينة الإحصائية، الوحدة الإحصائية، الظاهرة الإحصائية
4. ما هي أنواع المتغيرات الإحصائية؟

حل التمرين الأول:

1. المقصود بكلمة الإحصاء: كلمة الإحصاء لا تعني مجرد أرقام أو بيانات رقمية فقط كأعداد السكان، أعداد المواليد، بل هو علم يهتم بوصف طرق متعددة لجمع البيانات والملاحظات ومن ثم تنظيمها وعرضها باستخدام الأساليب العلمية لتحليلها واستخلاص النتائج منها بهدف الوصول إلى أخذ القرارات المناسبة.
2. مصادر البيانات الإحصائية:
 - ✓ مصادر تاريخية (مصدر غير مباشر): هي بيانات سبق تجميعها ونشرها في الكتب والجرائد أو المجلات أو مواقع الانترنت وهو ما يؤخذ من السجلات المحفوظة مثل: سجل المواليد، أو من إحصائيات منجزة مسبقا.
 - ✓ مصادر ميدانية (مصدر مباشر): هي بيانات يتم جمعها ميدانيا من المجتمع كله أو جزء منه ويشمل إما المقابلة الشخصية أو استخدام الاستبيانات الإحصائية.
3. تعريف المجتمع الإحصائي، العينة الإحصائية، الوحدة الإحصائية، الظاهرة الإحصائية
 - ✓ المجتمع الإحصائي: هو مجموعة من الملاحظات الخاصة بمجموعة من الوحدات الإحصائية ذات الخصائص المشتركة.
 - ✓ العينة الإحصائية: هي جزء من المجتمع الإحصائي، تأخذ بدلا من دراسة المجتمع ككل.
 - ✓ الوحدة الإحصائية: هي كل عنصر من عناصر العينة الإحصائية.

✓ الظاهرة الإحصائية: هي الخاصية أو الصفة أو المتغير المدروس في المجتمع الإحصائي.

4. أنواع المتغيرات الإحصائية: (الرجوع إلى الشكل (1-2): أنواع المتغيرات الإحصائية)

التمرين الثاني:

حدد المجتمع الإحصائي، الوحدة الإحصائية، المتغير الإحصائي، ونوعه في العبارات التالية:

1. الأجور الشهرية لعمال مؤسسة ما.
2. تقديرات الطلبة الجامعيين.
3. تصنيف الأحزاب السياسية حسب عدد الأصوات المكتسبة في الانتخابات.
4. أوزان التلاميذ في مدرسة ابتدائية.
5. مدة حياة المصابيح الكهربائية المنتجة في مصنع.
6. عدد السيارات في الدولة حسب الصنف.
7. جنسية مجموعة من الأشخاص.
8. الرياضة الممارسة من طرف طلبة جامعة الجزائر 3.
9. عدد الغرف المملوكة من طرف كل عائلة في حي معين.
10. تصنيف عمال مصنع حسب المؤهل.
11. علامات الطلبة المتحصل عليها في مقياس الإحصاء 1.
12. لون العيون لأفراد العائلة.

حل التمرين الثاني: تحديد المجتمع الإحصائي، الوحدة الإحصائية، المتغير الإحصائي، ونوعه

الرقم	المجتمع الإحصائي	الوحدة الإحصائية	المتغير الإحصائي	نوع المتغير الإحصائي
1	عمال المؤسسة	العامل	الأجر الشهري	كمي مستمر
2	الطلبة الجامعيين	الطالب	التقديرات	كيفي رتبي

3	الأحزاب السياسية	الحزب	عدد الأصوات	كمي منقطع
4	التلاميذ	تلميذ	الوزن	كمي مستمر
5	المصابيح الكهربائية	المصباح	مدة الحياة	كمي مستمر
6	السيارات	السيارة	الصنف	كيفي اسمي
7	الأشخاص	الشخص	الجنسية	كيفي اسمي
8	طالبة جامعة الجزائر 3	الطالب	الرياضة	كيفي اسمي
9	عائلات الحي	العائلة	عدد الغرف	كيفي اسمي
10	العمال	العامل	المؤهل	كيفي رتبي
11	الطلبة	الطالب	العلامات	كمي مستمر
12	الأفراد	الفرد	اللون	كيفي اسمي

التمرين الثالث:

قامت مؤسسة الاتصالات "موبيليس" بإجراء دراسة إحصائية بغرض معرفة مدى رضا الزبائن المتعاملين معها حول جودة الخدمات المقدمة لهم.

المطلوب:

1. ماهو الهدف العام من الدراسة؟
2. حدد المجتمع الإحصائي، الوحدة الإحصائية ؟
3. أذكر المتغير الإحصائي و نوعه ؟
4. ما هو الأسلوب المستخدم في هذه الدراسة و ما هي المصادر المعتمدة لجمع البيانات في مثل هذه الدراسة؟

حل التمرين الثالث:

الهدف العام من الدراسة معرفة مدى رضا الزبائن المتعاملين مع مؤسسة الاتصالات "موبليس" حول جودة الخدمات المقدمة لهم.

الهدف العام من الدراسة	معرفة مدى رضا الزبائن المتعاملين مع مؤسسة الاتصالات "موبليس" حول جودة الخدمات المقدمة لهم.
المجتمع الإحصائي	الزبائن المتعاملين مع مؤسسة الاتصالات "موبليس".
الوحدة الإحصائية	الزبون المتعامل مع مؤسسة الاتصالات "موبليس".
المتغير الإحصائي	رضا الزبون حول جودة الخدمات المقدمة لهم من طرف مؤسسة موبليس.
نوع المتغير الإحصائي	كيفي ترتيبي
طريقة جمع البيانات	أسلوب المعاينة تقوم هذه الطريقة على أخذ جزء من المجتمع الإحصائي يكون ممثل لهذا المجتمع.
مصادر جمع البيانات	المقابلة المباشر مع الزبائن المستهدفين أو عن طريق توزيع الاستبيان عليهم.

التمارين المقترحة للفصل الأول:

التمرين الأول:

1. عرف علم الإحصاء، وماهي أقسامه؟
2. أذكر الفرق بين الإحصاء الوصفي والإحصاء الاستدلالي؟
3. ما هو أسلوب جمع البيانات؟
4. ما هي خطوات البحث الإحصائي؟

التمرين الثاني:

صنف المتغيرات التالية حسب نوعيتها (كيفية اسمية، كيفية رتبية، كمية منقطعة، كمية مستمرة):

طول الطلبة- وزن اللاعبين- عدد الجرارات- الحالة المدنية- لون العيون- عدد المقاييس المدروسة-
الدخل السنوي- العمر- الجنس- القامة- الجنسية- درجات الحرارة- تصنيف العمال- فرع الدراسة- منطقة
الإقامة- كمية الإنتاج في مصنع- قيمة الإتصالات الهاتفية اليومية- عدد المكالمات الهاتفية في اليوم-
نسبة الكولسترول في الدم- عدد عربات القطار.

التمرين الثالث:

سجل حضور 100 مريض يوم 2018/09/12 لمصلحة أمراض العيون في المركز الاستشفائي الجامعي نفيسة حمود (بارني سابقا) بالعاصمة الجزائرية. نرسم لضغط الدم (المتغير المدروس) بالرمز (xi)، والمطلوب اختيار الإجابة الصحيحة فيما يلي:

(1) يتكون المجتمع المدروس من:

- المرضى المصابين بمرض العيون.
- كل المرضى المسجلين بمصلحة أمراض العيون.
- 100 مريض الذين حضروا يوم 2018/09/12.

(2) المتغير (xi) المدروس هو متغير:

- كمي مستمر.
- كمي منقطع.
- كيفي.

الفصل الثاني

العرض الجبري والبياني للبيانات الإحصائية

الفصل الثاني: العرض الجدولي والبياني للبيانات الإحصائية

تمهيد

تعد البيانات الإحصائية في صورتها الأولية غير صالحة للتحليل المباشر فغالبا ما تكون معقدة وكبيرة الحجم، الأمر الذي يجعل التعامل معها في شكلها الخام أمرا صعبا، لذلك يتم إعادة تنظيمها وترتيبها ضمن جداول إحصائية وفق أسس منهجية دقيقة، بما يسمح بفهمها وتحليلها بشكل أفضل، وتتنوع أساليب تصنيف المعطيات داخل الجداول بين تصنيف نوعي (كيفي) أو كمي، و سواء أكانت المتغيرات منقطعة أو مستمرة، ويساهم هذا الترتيب في عرض المعطيات بصورة مختصرة وواضحة، مما يساعد على إبراز العلاقات التي تربط بينها وتدعم عملية التحليل الإحصائي.

أولاً: العرض الجدولي للبيانات الإحصائية

أساسيات العرض الجدولي:

يقوم الباحث بتفريغ النتائج المتحصل عليها في جداول أولية بعد الانتهاء من عملية فرز الاستمارات ثم يلخصها في جداول نهائية، فالعرض الجدولي هو عبارة عن وضع المعلومات الإحصائية في جداول نهائية يحتوي كل منها على عمودين (سطين) يبين العمود (السطر) الأول قيم الظاهرة أو المتغير المدروس، وتكون هذه القيم على شكل قيم نقطية أو على شكل مجالات، أما العمود (السطر) الثاني فيحتوي على تكرارات هذه القيم أو المجالات¹.

¹ جلاطو جيلالي، مرجع سبق ذكره، ص 11

عناصر الجداول الإحصائية:

يجب على كل جدول أن ينظم حسب قواعد عامة تمكن القارئ من تفهم محتوياته واستخلاص الاتجاه العام لها، ونقصد بتنظيم الجدول أن يحتوي على كل العناصر الأساسية المطلوبة في بنائه فكل جدول يحتوي على العناصر المطلوبة التالية:

1. رقم الجدول حتى يكون عاملاً مساعداً في تمييزه عن الجداول الأخرى.
2. عنوان الجدول ويكون واضحاً يعطي وصفاً مختصراً.
3. يقسم الجدول إلى أعمدة (أفقية أو عمودية) ولكل عمود عنوان خاص به، ويوضح محتوياته بإيجاز مع توضيح وحدة قياس قرب العنوان أعلى كل عمود (هذا إذا كان الجدول في حالة تصنيف كمي)، أما إذا كانت الجداول التي تحتوي على تصنيف وصفي للبيانات فيستحسن رصد البيانات في الجداول حسب ترتيبها الجغرافي أو الأبجدي أو القابلة للترتيب.
4. يختتم كل عمود من أعمدة الجدول على المجاميع في الأسفل، ولا يكون أخطاء فيها مما يقلل من أهمية الجدول، وفي حالة وجود تقريب في الأرقام يجب الإشارة إلى ذلك أو كتابة إشارة التقريب قرب المجموع.
5. أعمدة تخصص لل تكرارات المختلفة وعادة ما يكون أول عمود هو التكرار المطلق (التكرار العادي) ثم يليه التكرار النسبي ثم المؤي ثم التكرار المتجمع الصاعد والتكرار المتجمع النازل وهكذا..¹

العرض الجدولي للمتغيرات الكمية

وهي الجداول التي نصنّف فيها ظاهرة تُقاس كمياً، أي التي تعرض تكرارات كميات الظواهر، كتوزيع عدد من الأشخاص حسب أعمارهم أو وزنهم أو أجورهم. وقد تكون هذه الجداول متصلة أو منفصلة، وفي

¹ تيلولت سامية، مرجع سبق ذكره، ص 11.

وهو جدول يتكون من عمودين: أحدهما يضم قيم المتغير الكمي المنقطع (أرقام صحيحة) مرتبة تصاعدياً، والثاني يضم عدد المفردات (التكرارات) لكل قيمة.

مثال رقم (02-01): البيانات التالية تمثل عدد مرات تأخر 45 عاملاً عن موعد الدوام في إحدى المؤسسات خلال شهر معين. المطلوب: وضع هذه البيانات في جدول تكراري.

[illegible]

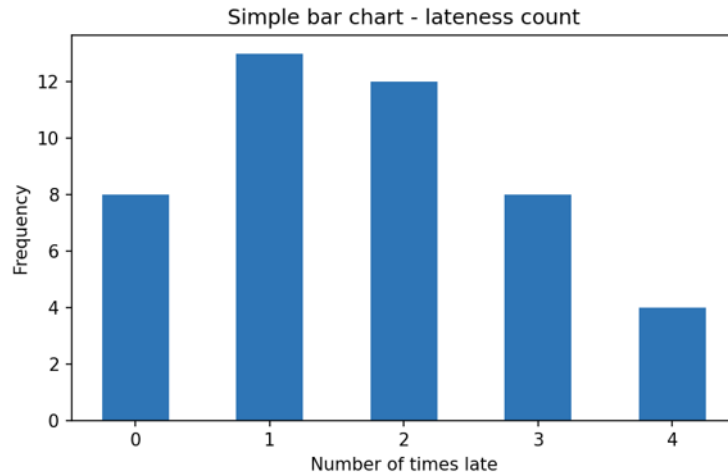
الحل: يتم ترتيب عدد مرات التأخر تصاعدياً ثم وضعها في العمود الأول من الجدول، ثم عدّ مرات تكرار كل قيمة باستخدام علامات التشطيب (////) في عمود مساعد، فنحصل على جدول التفريغ التالي:

عدد مرات التأخر	0	1	2	3	4	المجموع
التكرار f_i (عدد العمال)	8	13	12	8	4	45

وهذا هو جدول التوزيع التكراري النهائي بعد حذف عمود التفرغ المساعد.

والشكل التالي يمثل هذا التوزيع بالأعمدة البسيطة، وهو التمثيل البياني الأنسب للمتغير الكمي المنقطع:

¹ بیرى نوره، مرجع سبق ذكره، ص 13.



الشكل 2-1: أعمدة بسيطة تمثل توزيع العمال حسب عدد مرات التأخر

التوزيع التكراري للمتغير الكمي المستمر

يُلجأ إلى هذه الطريقة عندما يكون حجم البيانات كبيراً، إذ يتم تجميع البيانات في مجموعات تسمى بالفئات، كل فئة يقابلها تكرار معين. ومن أجل تشكيل جدول التوزيع التكراري في هذه الحالة نتبع الخطوات التالية¹:

1) تحديد طول الفئة (Class width): يُحدّد طول الفئة (C) بقسمة المدى (R) على عدد الفئات (K):

$$C = R/K$$

حيث المدى R هو الفرق بين أكبر قيمة $X_{\max}$ وأصغر قيمة $X_{\min}$ في المجموعة الإحصائية، أما عدد الفئات K فتوجد عدة طرق تقريبية لتحديده، أهمها:

❖ طريقة يول $K = 2.5 \times \sqrt[4]{F}$ (Yule)، حيث F عدد عناصر المجموعة الإحصائية.

¹ هشام بورمة، الإحصاء 1-دروس وأمثلة تطبيقية-، كلية العلوم الاقتصادية والتجارية وعلوم التسيير، جامعة محمد الصديق بن يحيى، جيجل، الجزائر، 2018/2017، ص 15-17

❖ طريقة ستورجس (Sturges): وتُعتبر الطريقة الوحيدة التي تأخذ بعين الاعتبار طبيعة

الظاهرة المدروسة مع كيفية توزيعها الإحصائي:

$$K = 1 + 3.322 * \log(N)$$

2 كتابة حدود الفئات (Class limits): تُكتب حدود الفئات في شكل مجالات مغلقة من

الأدنى ومفتوحة من الأعلى (أي أن القيمة الدنيا مشمولة والقيمة العليا غير مشمولة)،

انطلاقاً من الفئة الأولى التي يكون حدها الأدنى أصغر قيمة في البيانات X_{min} ، ثم

يُضاف لها طول الفئة C تباعاً حتى نغطي جميع البيانات.

3 تحديد مراكز الفئات (Class mid-point): يُحصل على مركز الفئة بجمع حديها وقسمة

النتائج على اثنين:

$$x_i = (X_{min} + X_{max})/2$$

4 وضع جدول التوزيع التكراري: نكتب في الأخير الفئات الناتجة وتكراراتها المقابلة f_i .

مثال رقم (02-02): تمثل البيانات التالية درجات 40 طالباً في اختبار مادة الرياضيات (من 25

نقطة). المطلوب: تشكيل جدول التوزيع التكراري.

12، 15، 9، 18، 14، 11، 16، 13، 10، 17، 8، 22، 6، 20، 12، 15، 9، 19، 14،

11، 16، 13، 10، 17، 21، 19، 6، 20، 12، 15، 9، 18، 14، 11، 16، 13، 10، 17، 8،

19

الحل:

1 تحديد طول الفئة:

$$R = X_{max} - X_{min} = 22 - 6 = 16 \text{ المدى}$$

عدد الفئات، بطريقة ستورجس:

$$K = 1 + 3.322 \times \log(40) = 1 + 3.322 \times 1.602 \approx 6.32 \approx 6$$

$$C = R / K = 16 / 6 \approx 2.67 \approx 3$$

ومنه طول الفئة: 3

(2) كتابة حدود الفئات: نبدأ من أصغر قيمة (6) ونضيف لها طول الفئة (3) تباعاً حتى تغطي جميع البيانات (6 فئات).

(3) تحديد مراكز الفئات و4- وضع جدول التوزيع التكراري:

الفئات (العلامات)	6-9	9-12	12-15	15-18	18-21	21-24	المجموع
مراكز الفئات x_i	7.5	10.5	13.5	16.5	19.5	22.5	/
التكرار f_i (عدد الطلبة)	4	9	9	9	7	2	40

ملاحظة: تُقدّم الجداول التكرارية المستمرة بعدة صيغ، منها:

التوزيع التكراري المنتظم وغير المنتظم: يكون منتظماً إذا كان طول جميع الفئات متساوياً، وغير منتظم في الحالة المعاكسة.

التوزيع التكراري المغلق والمفتوح: يكون مغلقاً إذا كان الحد الأدنى لأول فئة والحد الأعلى لآخر فئة محددين؛ ويكون مفتوحاً من الأعلى إذا كان الحد الأدنى لأول فئة غير محدد ("أقل من...")، أو مفتوحاً من الأسفل إذا كان الحد الأعلى لآخر فئة غير محدد ("أكثر...")، أو مفتوح الطرفين إذا كان الحدان معاً غير محددين.

جدول التوزيع التكراري النسبي والمئوي

التوزيع التكراري النسبي يبين الأهمية النسبية لكل فئة، أي نسبة ما يشكله تكرار كل فئة إلى مجموع التكرارات؛

ويُحصل على التكرار النسبي بقسمة كل تكرار على مجموع التكرارات، ويُحسب بالعلاقة التالية¹:

$$f_r = f_i / \sum f_i$$

أما التكرار المئوي فيُحصل عليه بضرب التكرار النسبي في 100:

$$f_r \% = \left(f_i / \sum f_i \right) * 100 = f_r * 100$$

مثال رقم (02-03): احسب التكرار النسبي والتكرار المئوي للجدول التالي الذي يمثل أزمدة توصيل 25 طرداً بريدياً (بالدقائق):

الفئات	20-10	30-20	40-30	50-40	60-50	المجموع
التكرار f_i	3	6	9	5	2	25

الحل:

الفئات	f_i	التكرار النسبي f_r	التكرار المئوي $f_r\%$
10-20	3	0.12	12
20-30	6	0.24	24
30-40	9	0.36	36
40-50	5	0.20	20
50-60	2	0.08	8
المجموع	25	1	100

¹ Jean Michel Josselin, Benoit Le Maux, Statistical Tools for Program Evaluation, Springer International Publishing, 2017, p.48.

التعليق: نلاحظ أن نسبة 36% من الطرود استغرق توصيلها بين 30 و 40 دقيقة وهي أكبر نسبة، أما أقل نسبة (8%) فتمثل الطرود التي استغرق توصيلها بين 50 و 60 دقيقة.

التوزيع التكراري المتجمع

التوزيع التكراري البسيط يعطينا عدد المفردات في كل فئة، لكن في بعض الأحيان نرغب في معرفة عدد المفردات التي قيمتها أقل أو أكثر من قيمة معينة؛ ويُعرّف التوزيع التكراري المتجمع بأنه التوزيع الذي يبين كمية التكرار المتجمع عند قيمة معينة من قيم المتغير، وهناك نوعان من الجداول التكرارية المتجمعة¹:

أ. **التوزيع التكراري المتجمع الصاعد ($F_{\uparrow}$):** وهو تجميع تكرار كل فئة مع جميع التكرارات

السابقة لها؛ ننطلق بالتكرار العادي للفئة الأولى كتكرار صاعد لها ($F_1 \uparrow = f_1$)، ثم نضيف

تكرار الفئة الثانية إلى التكرار الصاعد للفئة الأولى لإيجاد التكرار الصاعد للفئة الثانية،

وهكذا حتى نصل إلى التكرار الصاعد للفئة الأخيرة الذي يجب أن يساوي مجموع التكرارات.

والتكرار المجمع الصاعد له علاقة بالحد الأعلى للفئات، فهو يجيب على السؤال التالي: ما عدد

العناصر الأقل تماماً من قيمة معينة²؟

مثال رقم (02-04): بالاعتماد على معطيات المثال رقم (02-03)، كَوّن جدول التوزيع التكراري

المتجمع الصاعد.

الفئات	f_i	الحدود العليا للفئات	التكرار المتجمع الصاعد $F_{\uparrow}$
20-10	3	أقل من 20	3
30-20	6	أقل من 30	9
40-30	9	أقل من 40	18

¹ عماد توما كرش، ولاء أحمد القزاز وآخرون، مرجع سبق ذكره، ص 29.

² مهدي محمد القصاص، مبادئ الإحصاء والقياس الاجتماعي، مطبوعة جامعية، جامعة المنصورة، مصر، 2007، ص 135.

23	أقل من 50	5	50-40
25مجموع التكرارات	أقل من 60	2	60-50

التعليق: التكرار المتجمع الصاعد للفئة الثالثة يمثل عدد الطرود التي استغرق توصيلها أقل من 40 دقيقة، وهم 18 طرداً.

ب. **التوزيع التكراري المتجمع النازل ($F_{\downarrow}$):** نضع مجموع التكرارات كتكرار نازل للفئة الأولى، ثم نطرح التكرار العادي للفئة الأولى من هذا المجموع لإيجاد التكرار النازل للفئة الثانية، وهكذا حتى نصل إلى التكرار النازل للفئة الأخيرة الذي يجب أن يساوي تكرارها العادي.

والتكرار المتجمع النازل له علاقة بالحد الأدنى للفئات، فهو يجيب على السؤال: ما عدد العناصر التي تساوي أو تفوق قيمة معينة؟

مثال رقم (02-05): بالاعتماد على نفس المعطيات، كَوّن جدول التوزيع التكراري المتجمع النازل.

الفئات	f_i	الحدود الدنيا للفئات	التكرار المتجمع النازل $F_{\downarrow}$
20-10	3	20 فأكثر	25 (مجموع التكرارات)
30-20	6	20 فأكثر	22
40-30	9	30 فأكثر	16
50-40	5	40 فأكثر	7
60-50	2	50 فأكثر	2 (نفس التكرار العادي للفئة الأخيرة)

التعليق: التكرار المتجمع النازل للفئة الثالثة يمثل عدد الطرود التي استغرق توصيلها 30 دقيقة فأكثر، وهم 16 طرداً.

ت. **والتكرار المتجمع الصاعد والنازل النسبي والنسبي المئوي:** يُحسبان بنفس الطريقة المعتمدة أعلاه، ولكن بالاعتماد على التكرار النسبي بدلاً من التكرار المطلق، ثم بضرب الناتج في 100 للحصول على الصيغة المئوية.

مثال: بالاعتماد على بيانات علامات الطلبة الثلاثين (مثال 02-03 مكرر)، كَوّن جدولي التكرار المتجمع الصاعد والنازل معاً.

الفئات	f_i	$F\uparrow$	$F\downarrow$
4-0	2	2	30
8-4	6	8	28
12-8	12	20	22
16-12	7	27	10
20-16	3	30	3

التعليق: يمثل التكرار المتجمع الصاعد للفئة الثالثة (20) عدد الطلبة الذين حصلوا على أقل من 12 نقطة، بينما يمثل التكرار المتجمع النازل لنفس الفئة (22) عدد الطلبة الذين حصلوا على 8 نقاط فأكثر.

تطبيقات عملية للتوزيعات التكرارية المتجمعة: تُستخدم هذه الجداول كثيراً في التعليم (لمعرفة عدد أو نسبة الطلبة الناجحين، أي الذين حصلوا على العلامة الدنيا للنجاح فأكثر)، وفي الاقتصاد (لتحديد نسبة الأسر التي يقل دخلها عن عتبة معينة، كخط الفقر)، وفي الصناعة (لمراقبة الجودة، كتحديد عدد القطع التي تتجاوز مدة صلاحيتها حداً معيناً).

ثانياً: العرض البياني للبيانات الإحصائية (Graphical Presentation)

لا يستطيع القارئ غير المتخصص إدراك الجداول التكرارية بسهولة، خاصة إذا احتوت على مصطلحات متخصصة؛ لذلك يعطي الرسم البياني فكرة عامة مرئية بالعين المجردة، ويكون أكثر جاذبية وأسهل فهماً واستيعاباً. ويختلف أسلوب العرض البياني للبيانات حسب نوع المتغير المدروس (كمي أو كمي، منقطع أو مستمر).

a) العرض البياني للمتغيرات الكيفية

أكثر العروض والتمثيلات البيانية انتشاراً لوصف المتغيرات الكيفية هي الأعمدة البيانية والدوائر النسبية.

الأعمدة البيانية (Bar Chart):

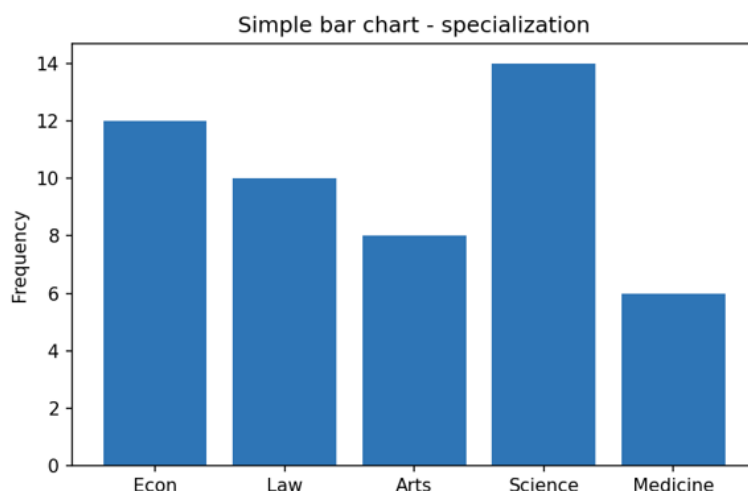
وهي مستطيلات غير متلاصقة تفصلها مسافات متساوية، يمثل المحور الأفقي فيها الظاهرة المدروسة بينما يمثل المحور العمودي التكرارات لكل مفردة.

مثال رقم (02-06): يمثل الجدول التالي التخصص الدراسي لعينة من 50 طالباً في إحدى الجامعات:

التخصص	اقتصاد	حقوق	آداب	علوم	طب	المجموع
التكرار	12	10	8	14	6	50

المطلوب: اعرض هذه البيانات بالتمثيل المناسب.

الحل: المتغير الإحصائي هو التخصص الدراسي، وهو متغير كيفي، وبالتالي يمكن تمثيله بالأعمدة البيانية:



الشكل 2-2: أعمدة بيانية تمثل توزيع الطلبة حسب التخصص الدراسي

الأعمدة البيانية المتلاصقة (Multiple Bar Graph):

تُستخدم عادة للمقارنة بين ظاهرتين ومقارنة التطور بينهما، وهي عبارة عن عمودين متلاصقين لكل قراءتين متناظرتين، تُلوّن أعمدة الظاهرة الأولى بلون خاص وأعمدة الظاهرة الثانية بلون مختلف حتى يمكن المقارنة بسهولة¹.

مثال رقم (02-07): إذا أدخلنا على المثال السابق (02-06) جنس الطلبة بالإضافة إلى التخصص، كما يوضحه الجدول التالي:

التخصص	اقتصاد	حقوق	آداب	علوم	طب	المجموع
ذكور	7	4	2	9	2	24
إناث	5	6	6	5	4	26

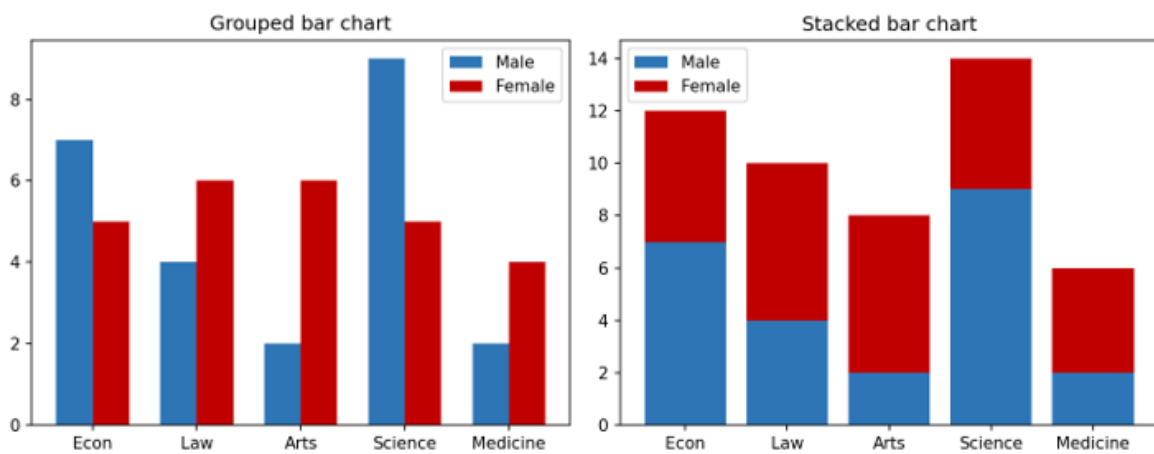
¹ سهيل أحمد سمحان، محمود حسين الوادي، مبادئ الإحصاء للاقتصاد والعلوم الإدارية، الطبعة الأولى، دار صفاء للنشر والتوزيع، عمان، 2010، ص58.

المطلوب: مثل هذه المعطيات بالأعمدة المتلاصقة.

الأعمدة البيانية المجزأة (Segmented/Stacked Bar Chart):

أعمدة بيانية بسيطة، لكن كل عمود منها يوضح قراءات أخرى لظاهرتين أو أكثر، بحيث يكون كل جزء في العمود متناسباً مع تكراره، ويمثل طول العمود الكامل المجموع الكلي للتكرارات.

مثال رقم (02-08): مثل الأعمدة المتلاصقة والأعمدة المجزأة لبيانات المثال (02-07):



الشكل رقم 2-3: أعمدة متلاصقة (يسار) وأعمدة مجزأة (يمين) لتوزيع الطلبة حسب التخصص والجنس

الدوائر النسبية (Pie Chart):

تمثل الدائرة مجموع القيم الكلية للظاهرة، وتُقسَّم إلى قطاعات جزئية تُميّز بألوان أو ظلال مختلفة. ويُستخدم هذا النوع من الرسوم عندما يكون الهدف مقارنة الأجزاء بالنسبة للمجموع الكلي، أو عندما تكون الأجزاء المقارنة قليلة العدد نسبياً، أو لتوضيح التطور النسبي لأجزاء الظاهرة عبر فترات زمنية مختلفة.

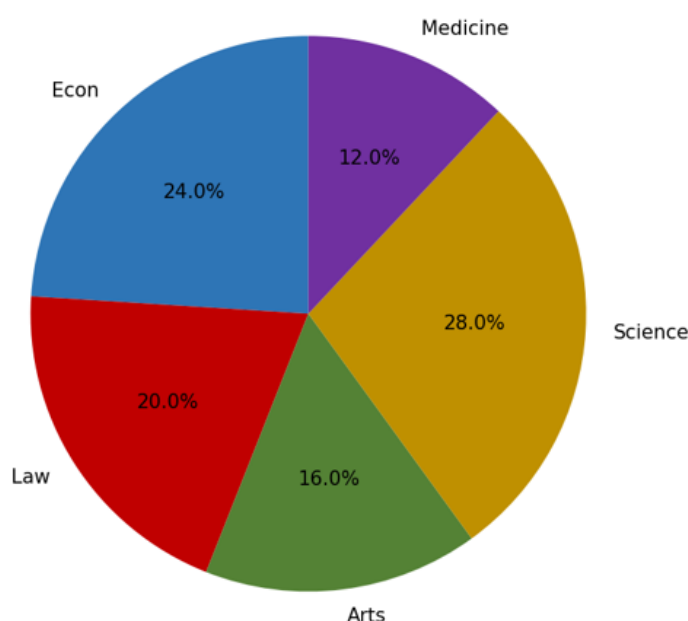
ولرسم الدائرة النسبية نتبع الخطوات التالية: نرسم دائرة بمقياس رسم مناسب، نحسب التكرار النسبي لكل فئة، ثم نحدد زاوية كل فئة بضرب تكرارها النسبي في 360° .

مثال رقم (02-09): مثل معطيات المثال رقم (02-06) بدائرة نسبية.

الحل: يجب حساب التكرار النسبي ثم الزوايا لكل تخصص:

التخصص	اقتصاد	حقوق	آداب	علوم	طب	المجموع
التكرار	12	10	8	14	6	50
التكرار النسبي	0.24	0.20	0.16	0.28	0.12	1
قياس الزاوية	86.4°	72°	57.6°	100.8°	43.2°	360°

Pie chart - specialization



الشكل 2-4: دائرة نسبية تمثل توزيع الطلبة حسب التخصص الدراسي

تطبيقات عملية للعرض البياني الكيفي: تُستخدم الأعمدة البيانية والدوائر النسبية بكثرة في التقارير الإدارية والاقتصادية (لعرض حصص السوق بين الشركات المنافسة)، والاستطلاعات الرأي (لعرض نتائج استبيان حول قضية معينة)، والتقارير الصحية (لعرض توزيع المرضى حسب نوع المرض). ويُفضّل

استعمال الدائرة النسبية عندما يكون عدد الفئات محدوداً (عادة أقل من 6 أو 7)، وإلا أصبح قراءة الدائرة صعبة ومربكة، فيُستحسن عندئذ اللجوء إلى الأعمدة البيانية.

العرض البياني للمتغيرات الكمية

تُستخدم هنا أنواع مختلفة من العروض البيانية، بحسب نوعية المتغير المدروس (منقطع أو مستمر).

العرض البياني للمتغير الكمي المنقطع:

يمكن هنا استخدام الأعمدة البسيطة، والعرض البياني للتكرارات المتجمعة الصاعدة والنازلة.

أ. **الأعمدة البسيطة (Simple Bar Chart):** يعتمد هذا النوع على رسم معلم، حيث نضع على محور الفواصل قيم المتغير، وعلى محور الترتيب التكرارات المناسبة لها، ثم نرسم خطوطاً عمودية مستقيمة تتناسب أطوالها مع التكرار المقابل (كما سبق تمثيله في الشكل رقم 01-02 أعلاه لمثال عدد مرات التأخر).

ب. **لعرض البياني للتكرار المتجمع الصاعد والنازل:** بالنسبة للمتغير المنقطع، يُمثل التكرار المتجمع الصاعد والنازل بمنحنيين متدرجين (سلمي الشكل)، حيث يمثل المحور الأفقي قيم المتغير والمحور العمودي التكرار المتجمع المقابل¹.

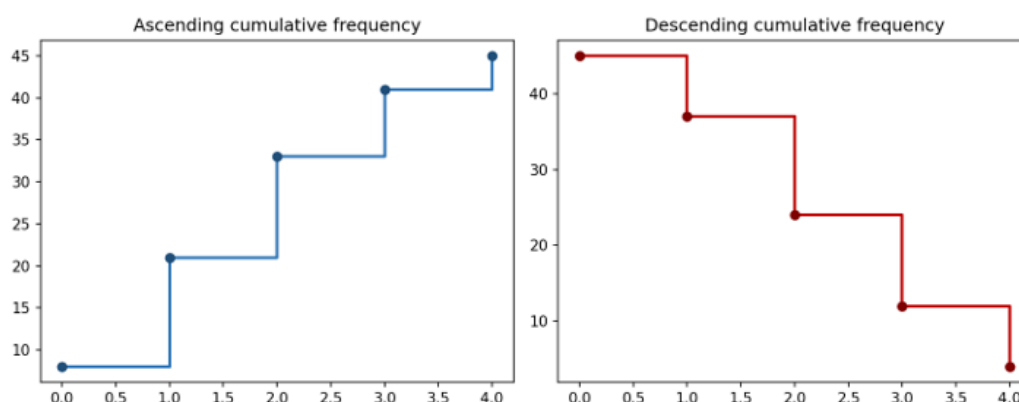
مثال رقم (02-10): مثل التكرار المتجمع الصاعد والنازل لبيانات المثال رقم (01-02) (عدد مرات التأخر لـ 45 عاملاً).

الحل: نحسب أولاً التكرار المتجمع الصاعد والنازل:

عدد مرات التأخر	0	1	2	3	4
التكرار f_i	8	13	12	8	4

¹ Timothy C. Urdan, Statistics in Plain English, Third Edition, Taylor and Francis Group, New York, 2010, p.07

45	41	33	21	8	$\uparrow F$
4	12	24	37	45	$\downarrow F$



الشكل 2-5: التكرار المتجمع الصاعد (يسار) والنازل (يمين) لعدد مرات التأخر

(b) العرض البياني للمتغير الكمي المستمر:

يمكن هنا استخدام المدرج التكراري، المضلع التكراري، المنحنى التكراري، ومنحنى التكرارات المتجمعة الصاعدة والنازلة (الأوجيف).

أ. المدرج التكراري (Histogram): نرسم المدرج التكراري على محورين متعامدين، أحدهما

أفقي يمثل الفئات والثاني عمودي يمثل التكرار، ونرسم مستطيلات متلاصقة على الفئات،

قاعدتها طول الفئة وارتفاعها تكرار هذه الفئة.

وهناك حالتان لرسم المدرج التكراري: حالة الفئات المتساوية الطول (توزيع تكراري منتظم)، حيث

يكون ارتفاع كل مستطيل هو التكرار المطلق مباشرة؛ ومن الملاحظ أن المساحة تحت المدرج التكراري

تساوي مجموع مساحات المستطيلات، أي طول الفئة مضروباً في مجموع التكرارات¹.

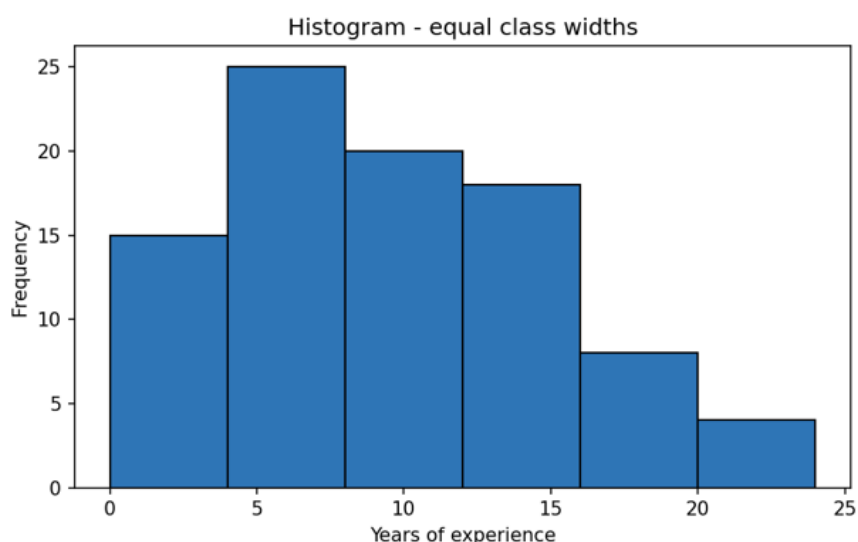
¹ محمد عبد العال النعيمي، الإحصاء المتقدم في دعم القرار بالتركيز على منظمات الأعمال الإنتاجية، الطبعة الأولى، مؤسسة الوراق للنشر والتوزيع، عمان، 2007، ص 330.

مثال رقم (11-02): يبين الجدول التالي توزيع 90 موظفاً في إحدى المؤسسات حسب مدة الخبرة المهنية (بالسنوات):

فئات الخبرة	4-0	8-4	12-8	16-12	20-16	24-20	المجموع
عدد الموظفين	15	25	20	18	8	4	90

المطلوب: ارسم المدرج التكراري.

الحل: نلاحظ أننا في حالة فئات متساوية الطول (4 سنوات)، وبالتالي نرسم المدرج التكراري مباشرة بارتفاعات تساوي التكرار المطلق لكل فئة:



الشكل 2-6: مدرج تكراري لتوزيع الموظفين حسب مدة الخبرة المهنية

أما في حالة الفئات غير متساوية الطول (توزيع تكراري غير منتظم)، فيلزم تعديل التكرارات قبل رسم المدرج، ومن أكثر الطرق استخداماً قسمة تكرار كل فئة على طولها للحصول على التكرار المعدل¹.

¹ صالح رشيد بطارسة، الإحصاء والاحتمالات، الطبعة الأولى، دار أسامة للنشر والتوزيع، عمان، 2010، ص82.

مثال رقم (12-02): بتعديل على معطيات المثال (11-02)، أصبح توزيع 84 موظفاً كما يلي:

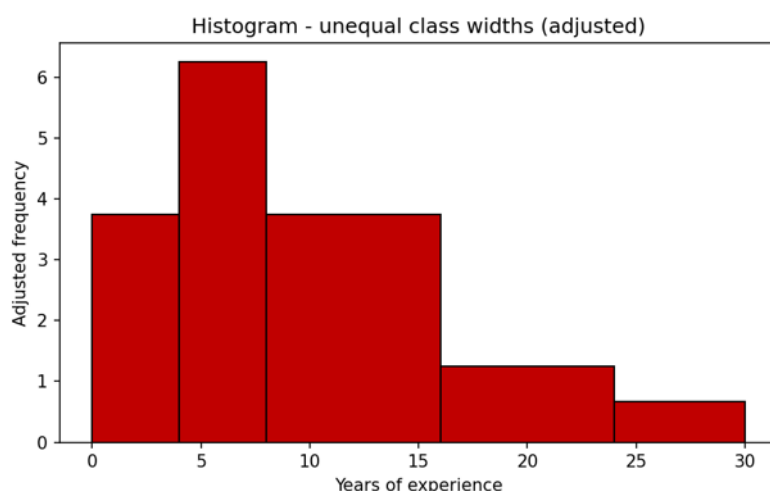
فئات الخبرة	4-0	8-4	16-8	24-16	30-24	المجموع
عدد الموظفين	15	25	30	10	4	84

المطلوب: ارسم المدرج التكراري.

الحل: الفئات غير متساوية الطول، لذا يجب أولاً حساب التكرار المعدل لكل فئة (بقسمة التكرار على

طول الفئة):

فئات الخبرة	4-0	8-4	16-8	24-16	30-24	المجموع
عدد الموظفين	15	25	30	10	4	84
طول الفئة	4	4	8	8	6	/
التكرار المعدل	3.75	6.25	3.75	1.25	0.667	/



الشكل 2-7: مدرج تكراري لفئات غير متساوية الطول (بالتكرار المعدل)

ب. المضلع التكراري (Frequency Polygon): وسيلة أخرى لتمثيل التوزيع التكراري بيانياً،

ويمكن رسمه بإحدى طريقتين: الطريقة الأولى (على المدرج التكراري) بتحديد منتصفات

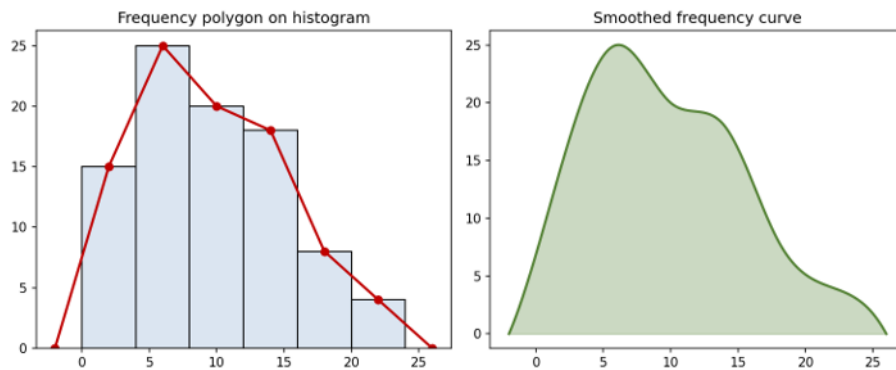
القواعد العليا لمستطيلات المدرج ثم توصيلها بمستقيمات، مع إضافة فئة قبل الأولى وفئة

بعد الأخيرة بتكرار معدوم لإغلاق المضلع؛ والطريقة الثانية (خارج المدرج) برسم نقاط فوق كل مركز فئة بارتفاع يساوي تكرارها، ثم توصيل هذه النقاط.

ت. المنحنى التكراري (Frequency Curve): يُرسم على نفس المحورين، لكن بدلاً من توصيل النقاط بخطوط مستقيمة، تُوصَل بمنحنى انسيابي باليد، مع إغلاقه بنفس طريقة المضلع (بإضافة فئة سابقة ولاحقة بتكرار معدوم).

مثال رقم (13-02): مثل المضلع التكراري (على المدرج) والمنحنى التكراري لبيانات المثال (02-11).

الحل: نضيف فئتين افتراضيتين بتكرار معدوم: فئة (4- 0) مركزها 2 (نقطة الانطلاق)، وفئة (24-28) مركزها 26 (نقطة النهاية):



الشكل 2-8: المضلع التكراري على المدرج (يسار) والمنحنى التكراري (يمين)

ملاحظة: يُفضَّل المضلع التكراري على المدرج عند الرغبة في مقارنة توزيعين على نفس الرسم البياني (كمقارنة أداء فوجين من الطلبة)، إذ يصعب ذلك برسم مدرجين متراكبين، بينما يسهل تماماً رسم مضلعين متراكبين بألوان مختلفة. أما المنحنى التكراري فيُستعمل أكثر في الدراسات النظرية لتقريب شكل التوزيع من منحنى رياضي معروف (كالتوزيع الطبيعي).

ث. **المنحنى المتجمع الصاعد (Ogive - Less than):** نرسم محورين متعامدين، الأفقي

للحدود العليا للفئات والعمودي للتركرات المتجمعة الصاعدة، ثم نحدد نقاطاً إحداثياتها السينية

الحدود العليا والصادية التكرارات المتجمعة الصاعدة، ونوصل بينها.

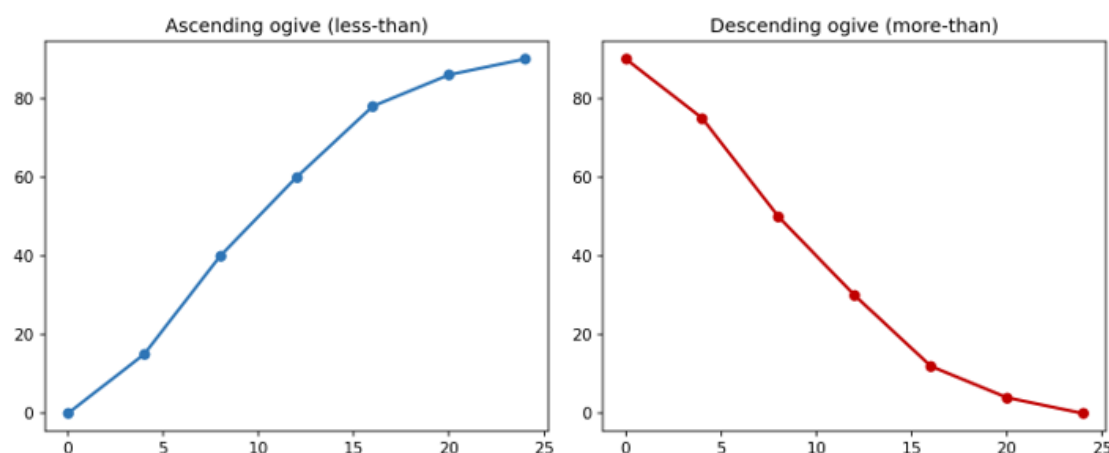
ج. **المنحنى المتجمع النازل (Ogive - More than):** بنفس المنطق، لكن باستعمال الحدود

الدنيا للفئات على المحور الأفقي والتكرارات المتجمعة النازلة على المحور العمودي.

مثال رقم (02-14): مثل منحنى التكرار المتجمع الصاعد والنازل لبيانات المثال (02-11).

الحل: نحسب أولاً التكرار المتجمع الصاعد والنازل:

فئات الخبرة	4-0	8-4	12-8	16-12	20-16	24-20
f_i	15	25	20	18	8	4
$\uparrow F$	15	40	60	78	86	90
$\downarrow F$	90	75	50	30	12	4



الشكل 2-9: المنحنى المتجمع الصاعد (يسار) والنازل (يمين) لمدة الخبرة المهنية

تطبيقات عملية للعرض البياني الكمي المستمر: يُستخدم المدرج التكراري والمنحنيات المتجمعة بكثرة

في الدراسات الاقتصادية (توزيع الدخل والأسعار)، والدراسات الصحية (توزيع الأوزان والأطوال حسب

الفئة العمرية)، والدراسات البيئية (توزيع درجات الحرارة أو معدلات التلوث)؛ كما يُستعمل منحني الأوجيف الصاعد بشكل خاص لتحديد الوسيط والربيعيات بيانياً، كما سيرد تفصيله في الفصل الموالي.

خلاصة: اختيار التمثيل البياني المناسب

يلخص الجدول التالي أهم أنواع التمثيل البياني المدروسة في هذا الفصل، ومتى يُستعمل كل منها:

نوع التمثيل	نوع المتغير المناسب	الاستعمال الأساسي
الأعمدة البيانية البسيطة	كيفي أو كمي منقطع	مقارنة تكرارات فئات مستقلة
الأعمدة المتلاصقة/المجزأة	كيفي (مع متغير ثانٍ)	مقارنة توزيعين فرعيين معاً
الدائرة النسبية	كيفي (عدد فئات محدود)	إبراز نسبة كل جزء من المجموع الكلي
المدرج التكراري	كمي مستمر	توزيع تكراري بفئات متصلة
المضلع والمنحنى التكراري	كمي مستمر	مقارنة عدة توزيعات، أو تقريب شكل نظري
منحنى الأوجيف (المتجمع)	كمي منقطع أو مستمر	تحديد عدد/نسبة العناصر دون أو فوق قيمة معينة، وإيجاد الوسيط بيانياً

تمارين الفصل الثاني

التمرين الأول:

البيانات التالية تمثل الحالة الاجتماعية لعينة من 12 موظفاً في إحدى المؤسسات:

أعزب، متزوج، متزوج، أعزب، مطلق، متزوج، أعزب، أرمل، متزوج، متزوج، أعزب، مطلق.

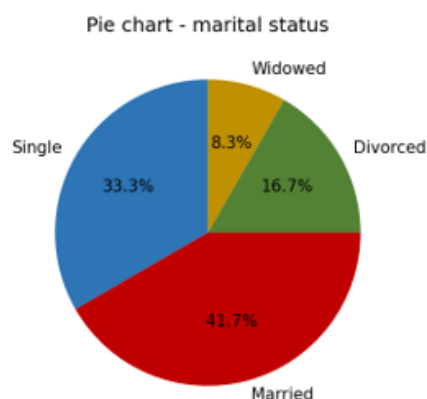
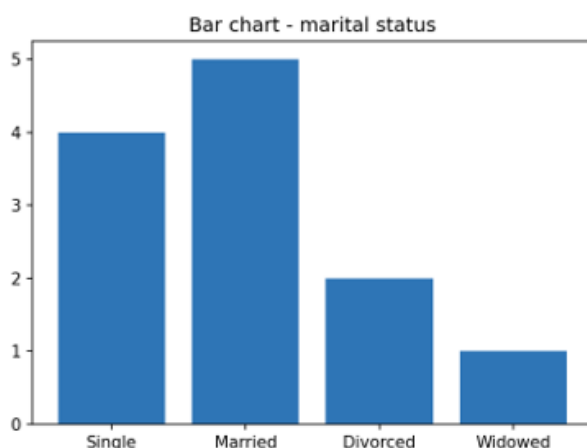
المطلوب: حدد المتغير الإحصائي ونوعه، كوّن جدول التوزيع التكراري ومثله بيانياً.

الحل:

1) المتغير الإحصائي: الحالة الاجتماعية للموظفين. نوعه: كيفي.

الحالة الاجتماعية	أعزب	متزوج	مطلق	أرمل	المجموع
التكرار	4	5	2	1	12
التكرار النسبي	0.333	0.4170	0.1670	0.0830	1
قياس الزاوية	120°	150°	60°	30°	360°

(3) التمثيل البياني: بما أن المتغير كيفي، يمكن تمثيله بالأعمدة البيانية والدائرة النسبية:



الشكل 0-1): أعمدة بيانية ودائرة نسبية لتوزيع الموظفين حسب الحالة الاجتماعية

التمرين الثاني:

لدينا عينة عشوائية تتكون من 42 أسرة موزعة حسب عدد السيارات المملوكة لكل منها كما يلي:

,1 ,2 ,3 ,1 ,0 ,2 ,1 ,4 ,1 ,2 ,3 ,0 ,1 ,2 ,1 ,0 ,2 ,1 ,3 ,2 ,1 ,0 ,2 ,1 ,1 ,0
 ,4 ,2 ,0 ,1 ,2 ,3 ,1 ,2 ,0 ,1 ,4 ,2 ,3 ,1 ,2 ,0

المطلوب: حدد المتغير الإحصائي ونوعه؛ كَوْن جدول التوزيع التكراري مشتملاً على التكرار النسبي والمئوي والتكرار المتجمع الصاعد والنازل؛ مثَل جدول التوزيع التكراري بالتمثيل المناسب؛ ومثَل منحني التكرار المتجمع الصاعد والنازل.

الحل:

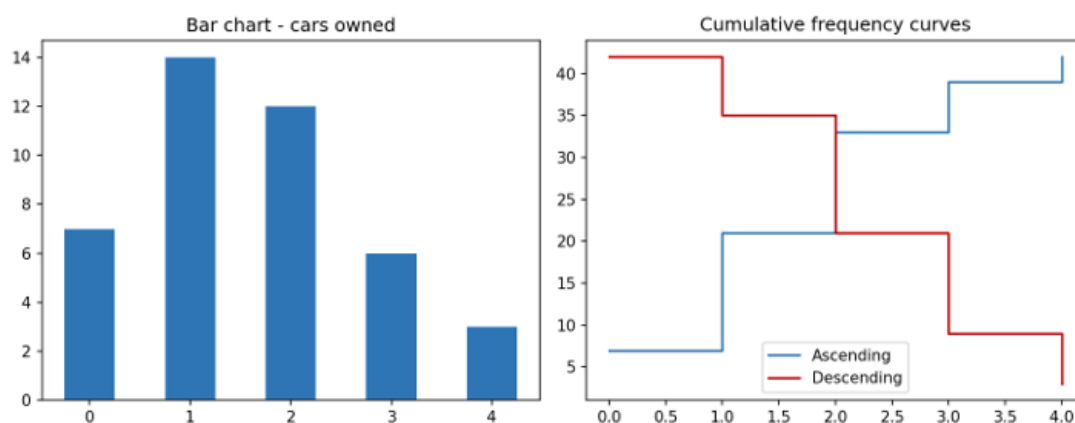
(1) المتغير الإحصائي: عدد السيارات المملوكة لكل أسرة. نوعه: كمي منقطع.

(2) جدول التوزيع التكراري:

عدد السيارات x_i	0	1	2	3	4	المجموع
التكرار f_i	7	14	12	6	3	42
التكرار النسبي fr	0.167	0.333	0.286	0.143	0.071	1
التكرار المئوي $fr\%$	16.7	33.3	28.6	14.3	7.1	100
$F \uparrow$	7	21	33	39	42	/
$F \downarrow$	42	35	21	9	3	/
عدد السيارات x_i	0	1	2	3	4	المجموع

(3) التمثيل البياني: بما أن المتغير كمي منقطع، يمكن تمثيله بالأعمدة البسيطة، إضافة إلى

منحني التكرار المتجمع الصاعد والنازل:



الشكل (2-0): أعمدة بسيطة ومنحني التكرار المتجمع لعدد السيارات المملوكة

التمرين الثالث:

سُجِّلَت أوزان 42 طرداً بريدياً (بالكيلوغرام) فكانت كما يلي:

22، 25، 19، 28، 24، 21، 26، 23، 20، 27، 18، 32، 16، 30، 22، 25، 19، 29،
24، 21، 26، 23، 20، 27، 31، 29، 16، 30، 22، 25، 19، 28، 24، 21، 26، 23، 20،
27، 18، 29، 17، 31.

المطلوب: حدد المتغير الإحصائي ونوعه؛ كوّن جدول التوزيع التكراري؛ احسب التكرار النسبي والمئوي؛ ارسم المدرج التكراري؛ كوّن جدول التكرار المتجمع الصاعد؛ وارسم منحنى التكرار المتجمع الصاعد.

الحل:

(1) المتغير الإحصائي: وزن الطرد البريدي. نوعه: كمي متصل.

(2) تكوين جدول التوزيع التكراري:

$$R = X_{\max} - X_{\min} = 32 - 16 = 16 \text{ المدى}$$

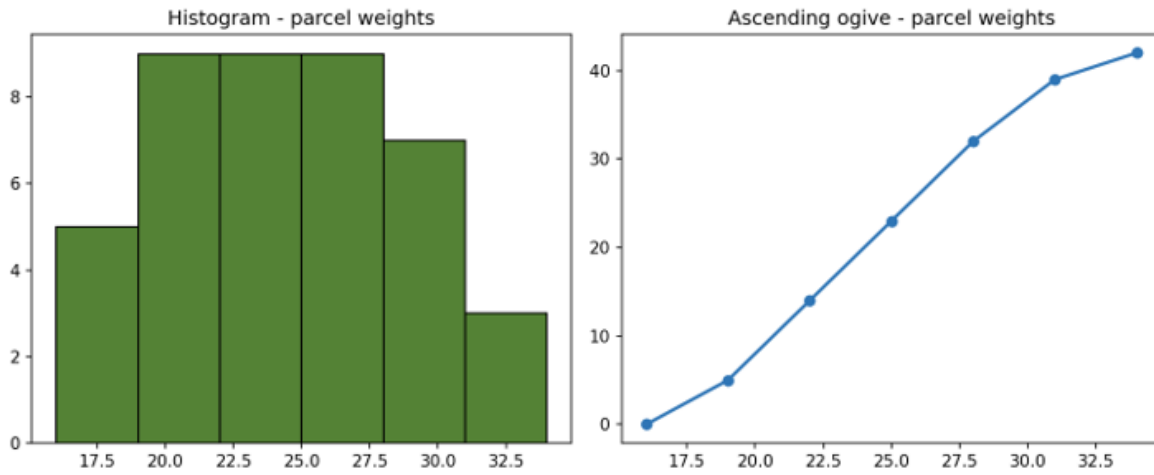
$$K = 1 + 3.322 \times \log(42) \approx 1 + 3.322 \times 1.623 \approx 6.39 \approx 6 \text{ عدد الفئات (ستورجس):}$$

6

$$C = R/K = 16/6 \approx 2.67 \approx 3 \text{ طول الفئة:}$$

الفئات	16-19	19-22	22-25	25-28	28-31	31-34	المجموع
التكرار f_i	5	9	9	9	7	3	42
التكرار النسبي	0.119	0.214	0.214	0.214	0.167	0.071	1
التكرار المئوي	11.9	21.4	21.4	21.4	16.7	7.1	100
$\uparrow F$	5	14	23	32	39	42	/

(3) المدرج التكراري ومنحنى التكرار المتجمع الصاعد:



الشكل 3-0: مدرج تكراري ومنحنى تكرار متجمع صاعد لأوزان الطرود البريدية

الفصل الثالث

مقاييس النزعة المركزية

الفصل الثالث: مقاييس النزعة المركزية

تمهيد

تناولنا في الفصل السابق طرق عرض البيانات الإحصائية جدولياً وبيانياً، بهدف تقديم وصف عام وسريع للظاهرة المدروسة وترتيب المعلومات الإحصائية بشكل منظم ومقروء. وقد رأينا كيف يتيح الجدول الإحصائي والتمثيل البياني (كالمدرج التكراري والمضلع التكراري) تكوين فكرة أولية عن شكل التوزيع وعن المجالات التي تتركز فيها أغلب المشاهدات.

غير أن هذا العرض، على أهميته المنهجية، يبقى محدوداً من عدة جوانب أساسية، أهمها:

- لا يصلح للاستخدام في العرض الشفهي للنتائج أمام جمهور أو لجنة، إذ يصعب استحضار جدول كامل أو منحنى بياني شفهاً.
- لا يتيح تحليل المعطيات تحليلاً كمياً دقيقاً يسمح بالمقارنة الموضوعية بين ظاهرتين أو أكثر.
- لا يمكن الاعتماد عليه في مجال الاستقراء الإحصائي (النتبؤ واتخاذ القرار)، لأن اتخاذ القرار يستلزم مؤشرات رقمية دقيقة ومختصرة يمكن توظيفها في نماذج رياضية وإحصائية لاحقة¹.

لهذه الأسباب مجتمعة، طُوّرت مقاييس إحصائية خاصة تصف التوزيعات التكرارية وصفاً كمياً دقيقاً، وتحلل شكل المنحنيات الممثلة لها، وتتيح كذلك المقارنة الموضوعية بين توزيعات مختلفة، بحيث يصف كل مقياس منها التوزيع من زاوية معينة ولغرض معين. ونتناول في هذا الفصل الصنف الأول والأهم من هذه المقاييس، وهو "مقاييس النزعة المركزية"، على أن نخصص الفصل الموالي لمقاييس التشتت التي تكمل هذه الصورة الإحصائية.

¹ جلاطو جيلالي، مرجع سبق ذكره، ص 30.

ولتوضيح فائدة هذه المقاييس بمثال بسيط، لنتخيل باحثاً يدرس دخل الأسر في حيّين سكنيين مختلفين، وتوصل إلى أن كلا الحيين يضم نفس العدد من الأسر (100 أسرة)؛ فهل يكفي هذا وحده للمقارنة بينهما؟ بالطبع لا، فقد يكون متوسط الدخل في الحي الأول أعلى بكثير منه في الحي الثاني، أو قد تتركز أغلب الأسر في الحي الأول حول دخل معين بينما تتوزع أسر الحي الثاني بشكل متفاوت جداً. ولمعرفة ذلك بدقة، لا بد من حساب مقياس يلخص "القيمة النموذجية" لدخل كل حي، وهذا بالضبط ما توفره مقاييس النزعة المركزية.

مفهوم النزعة المركزية

تنزع البيانات الإحصائية في أغلب الظواهر الطبيعية والاجتماعية والاقتصادية إلى التمرکز حول قيمة معينة، وكلما ابتعدنا عن هذه القيمة كلما تناقص عدد المشاهدات المحيطة بها؛ وتُعرف هذه الظاهرة بالنزعة المركزية، وهي خاصية شبه عامة في أغلب التوزيعات التكرارية المدروسة في الواقع العملي¹.

ومقاييس النزعة المركزية هي المقاييس التي تسعى لتقدير قيمة تتمركز حولها أغلب البيانات، بحيث تمثل هذه القيمة التوزيع بأكمله أفضل تمثيل ممكن، وتُغني بذلك الباحث عن استعراض جميع مفردات الظاهرة واحدة واحدة².

وتُعرف مقاييس النزعة المركزية أيضاً بمقاييس التمرکز أو مقاييس الموضع أو المتوسطات، وتعني كلها التعبير عن مجموعة كبيرة من القيم بقيمة واحدة ممثلة لها بطريقة مبسطة ووجيزة، ويتم ذلك عملياً بواسطة أحد المتوسطات الإحصائية المتعارف عليها.

وترتبط فكرة المتوسطات بمفهوم "الوسط" المتداول في اللغة العامة، حيث يُقال "خير الأمور أوسطها"، أي أن أفضل القيم هي التي تقع بين الحد الأدنى والحد الأعلى دون إفراط ولا تفريط. أما في الاصطلاح

¹ جلاطو جيلالي، نفس المرجع، ص30.

² سهيل أحمد سمحان ومحمود حسين الوادي، مرجع سبق ذكره، ص71.

الإحصائي، فإن المتوسط هو القيمة التي تعبر عن مجموعة كبيرة من القيم بقيمة واحدة، تُستنبط منها وفق علاقة رياضية محددة¹.

ومن أبرز مقاييس النزعة المركزية المتعارف عليها في الإحصاء الوصفي، والتي سيتناولها هذا الفصل بالتفصيل:

- الوسط الحسابي، ويرمز له بالرمز (X) .
- الوسط الهندسي، ويرمز له بالرمز (MG) .
- الوسط التوافقي، ويرمز له بالرمز (MH) .
- الوسط التربيعي، ويرمز له بالرمز (MQ) .
- الوسيط، ويرمز له بالرمز (Me) .
- الربيعيات، ويرمز لها بالرمز (Q_i) ، وكذلك العشيريات (D_i) والمئينيات (P_i) .
- المنوال، ويرمز له بالرمز (Mo) .

ولكل واحد من هذه المقاييس مجال تطبيق يناسبه بحسب طبيعة الظاهرة المدروسة ونوع البيانات المتوفرة عنها (مبوبة أو غير مبوبة، منقطعة أو مستمرة)، وهو ما سنفصله تباعاً.

قياس النزعة المركزية

الوسط الحسابي (Moyenne arithmétique)

يُعرف الوسط الحسابي أيضاً بالمتوسط، وهو أبسط أنواع المتوسطات وأكثرها استعمالاً في الحياة العملية والبحث العلمي على حد سواء، ويُعرّف بأنه حاصل قسمة مجموع قيم البيانات على عددها f في حالة العينة، F في حالة المجتمع الإحصائي الكامل².

¹ تيلولت سامية، مرجع سبق ذكره، ص 51-52.

² وليد إسماعيل السيفو وآخرون، مرجع سبق ذكره، ص 102.

ويرمز له بالرمز (X) ، وهو المقياس الذي يتبادر عادة إلى الذهن عند الحديث عن "المتوسط" في اللغة اليومية.

أ. طرق حساب الوسط الحسابي

نظراً لأهمية الوسط الحسابي البالغة في التحليل الإحصائي، توجد عدة طرق لحسابه بحسب طبيعة البيانات المتاحة؛ فيمكن حسابه من البيانات الأولية (غير المبوبة) مباشرة، كما يمكن حسابه من بيانات مبوبة في جدول تكراري، سواء تعلق الأمر بمتغير كمي منقطع أو متغير كمي مستمر. ونعرض فيما يلي هذه الطرق تباعاً مع أمثلة تطبيقية على كل منها.

1. حساب الوسط الحسابي من البيانات غير المبوبة

• الطريقة المباشرة (الوسط الحسابي البسيط):

يُستعمل هذا الأسلوب في حالة البيانات غير المبوبة، أي عندما تكون لجميع مشاهدات المتغير المدروس نفس درجة الأهمية، ولا داعي لترجيحها بأي وزن¹.

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

مثال رقم (03-01): تمثل السلسلة التالية رواتب 10 موظفين في مؤسسة خاصة (بالدينار الجزائري)، وقد جُمعت هذه البيانات مباشرة من كشف الرواتب الشهري دون أي تبويب مسبق:

850، 920، 1100، 1250، 1400، 980، 890، 1150، 1050، 1300

المطلوب: حساب متوسط الرواتب.

الحل: بتطبيق الطريقة المباشرة، نجمع القيم العشر جميعها ثم نقسم الناتج على عددها:

¹ جلاطو جيلالي، مرجع سبق ذكره، ص 30.

$$\bar{X} = (850+920+1100+1250+1400+980+890+1150+1050+1300) / 10 = 10890$$

$$\text{دج} / 10 = 1089$$

أي أن متوسط راتب الموظف الواحد في هذه المؤسسة يقدر بـ 1089 دج.

• طريقة الانحرافات البسيطة (طريقة الوسط الفرضي):

قد يكون حساب مجموع القيم مباشرة مرهقاً إذا كانت الأرقام كبيرة؛ ولتسهيل الحساب يمكن اختيار قيمة تقريبية نفترضها وسطاً مبدئياً، تسمى الوسط الفرضي (X_0)، ثم يكون الوسط الحسابي الحقيقي مساوياً لهذا الوسط الفرضي مضافاً إليه متوسط انحرافات قيم المتغير عنه¹.

$$\bar{X} = x_0 + \sum (x_i - x_0) / n$$

مثال رقم (03-02): نعيد حل المثال السابق نفسه بطريقة الوسط الفرضي، بافتراض $X_0 = 1000$ (وهي قيمة مقربة قريبة من وسط مجال الرواتب):

الحل: نحسب أولاً انحراف كل راتب عن الوسط الفرضي ($x_i - 1000$)، فنحصل على القيم التالية:
 -150، -80، 100، 250، 400، -20، -110، 150، 50، 300، ومجموع هذه الانحرافات يساوي 890.

$$\bar{X} = 1000 + 890 / 10 = 1000 + 89 = 1089 \text{ دج}$$

ونلاحظ تطابق النتيجة تماماً مع الطريقة المباشرة (1089 دج)، وهذا أمر منطقي لأن طريقة الوسط الفرضي ليست سوى إعادة صياغة جبرية للطريقة المباشرة، تُسهّل الحساب اليدوي دون أن تغير جوهر العملية.

2. حساب الوسط الحسابي من البيانات المبوبة

¹ جلاطو جيلالي، نفس المرجع، ص31.

• في حالة متغير كمي منقطع:

تختلف أهمية قياسات المتغير الإحصائي من قيمة إلى أخرى في غالب الأحيان بحسب عدد مرات تكرارها، فتختلف بذلك أهميتها النسبية (ترجيحها) داخل التوزيع الكلي؛ ويُستعمل في هذه الحالة الوسط الحسابي المرجح، الذي يأخذ بعين الاعتبار تكرار كل قيمة:

$$\bar{X} = \sum (x_i * n_i) / \sum n_i$$

مثال رقم (03-03): يمثل الجدول التالي عدد الجلسات التدريبية التي حضرها 100 موظف خلال شهر معين، ضمن برنامج تكويني داخلي نظّمته مؤسستهم:

عدد الجلسات (x_i)	2	4	6	8	10	المجموع
عدد الموظفين (f_i)	10	20	30	25	15	100

المطلوب: حساب متوسط عدد الجلسات التي حضرها الموظف الواحد.

الحل: بالطريقة المباشرة، نضيف عموداً لحساب حاصل ضرب كل قيمة في تكرارها المقابل، ثم نجمع هذا العمود:

x_i	f_i	$f x_i$
2	10	20
4	20	80
6	30	180
8	25	200
10	15	150
المجموع	100	630

$$\bar{X} = 630 / 100 = 6.3 \text{ جلسة}$$

أي أن الموظف الواحد حضر في المتوسط 6.3 جلسة تدريبية خلال الشهر.

ويمكن التحقق من هذه النتيجة بطريقة الانحرافات (الوسط الفرضي)، بأخذ $X_0 = 6$ (وهي القيمة المقابلة لأكبر تكرار):

$f_i \times (x_i - X_0)$	$(x_i - X_0)$	f_i	x_i
-40	-4	10	2
-40	-2	20	4
0	0	30	6
50	2	25	8
60	4	15	10
30	/	100	المجموع

$$\bar{X} = 6 + 30 / 100 = 6.3 \text{ جلسة}$$

وتتطابق النتيجة مرة أخرى مع الطريقة المباشرة، وهو ما يؤكد صحة الحساب.

• في حالة متغير كمي مستمر:

عندما تكون بيانات المتغير مبوبة في فئات (متغير كمي مستمر)، لا يمكن التعامل مع كل فئة إلا من خلال مركزها الحسابي (منتصفها)، بافتراض أن القيم داخل الفئة الواحدة موزعة بانتظام حولها. وتوجد ثلاث طرق أساسية لحساب الوسط الحسابي في هذه الحالة: الطريقة المباشرة، طريقة الانحرافات البسيطة، وطريقة الانحرافات المختصرة (طريقة الترميز).

مثال رقم (03-04): يمثل الجدول التالي توزيع الإنفاق الشهري على الاتصالات (بالدينار الجزائري)

لعينة من 100 أسرة، استُخلصت من استمارة استبيان ميداني:

الفئات	20-0	40-20	60-40	80-60	100-80	120-100	140-120	160-140
عدد الأسر	6	12	18	24	20	12	6	2

المطلوب: حساب متوسط الإنفاق الشهري على الاتصالات بالطرق الثلاث.

الحل:

1. الطريقة المباشرة (الوسط المرجح): نستعمل مراكز الفئات (x'_i) بدلاً عن حدودها، ونطبق

نفس علاقة الوسط المرجح المستعملة في حالة المتغير المنقطع:

الفئات	f_i	x'_i	$f_i \times x'_i$
0-20	6	10	60
20-40	12	30	360
40-60	18	50	900
60-80	24	70	1680
80-100	20	90	1800
100-120	12	110	1320
120-140	6	130	780
140-160	2	150	300
المجموع	100	/	7200

$$\bar{X} = 7200 / 100 = 72 \text{ دج}$$

2. طريقة الانحرافات البسيطة (الوسط الفرضي): نختار X من بين مراكز الفئات، ولكون الفئة

(60-80) تقابل أكبر تكرار (24) نأخذ $X_0 = 70$:

الفئات	f_i	x'_i	$(x'_i - X_0)$	$f_i \times (x'_i - X_0)$
0-20	6	10	-60	-360
20-40	12	30	-40	-480
40-60	18	50	-20	-360
60-80	24	70	0	0
80-100	20	90	20	400
100-120	12	110	40	480
120-140	6	130	60	360
140-160	2	150	80	160
المجموع	100	/	/	200

$$\bar{X} = 70 + 200 / 100 = 72 \text{ دج}$$

3. طريقة الانحرافات المختصرة (طريقة الترميز): وهي تحسين لطريقة الانحرافات البسيطة،

تهدف إلى تبسيط الأرقام المتداولة أكثر، وذلك بقسمة كل انحراف على طول الفئة الثابت

K، ثم ضرب الناتج النهائي في هذا الطول لاسترجاع الوحدة الأصلية¹:

$$\bar{X} = X_0 + \frac{\left[\sum \frac{n_i(x_i - X_0)}{K} \right]}{\sum n_i * K}$$

بأخذ $K = 20$ (طول الفئة)، تصبح أعمدة الانحرافات المقسومة على K كما يلي: -3، -2، -1،

0، 1، 2، 3، 4، ومجموع حاصل ضربها في التكرارات المقابلة يساوي 10:

$$\bar{X} = 70 + (10 / 100) \times 20 = 70 + 2 = 72 \text{ دج}$$

وهكذا تتطابق نتيجة الطرق الثلاث تماماً (72 دج)، وهو ما يؤكد أن اختلاف الطريقة المستعملة لا

يغير شيئاً من جوهر النتيجة، وإنما يقتصر أثره على تبسيط العمليات الحسابية اليدوية فقط.

ب. الوسط الحسابي المرجح لأوساط حسابية

قد يجد الباحث نفسه أمام حالة مختلفة قليلاً: إذ تتوفر لديه مجموعتان من البيانات أو أكثر، وقد

سبق حساب الوسط الحسابي لكل منها على حدة، ويرغب في معرفة الوسط الحسابي الإجمالي لهذه

المجموعات مجتمعة دون العودة إلى البيانات الأولية؛ ويُستخدم في هذه الحالة الوسط الحسابي المرجح

لأوساط حسابية، وفق العلاقة التالية:

$$\bar{X} = (n_1\bar{X}_1 + n_2\bar{X}_2 + n_3\bar{X}_3) / (n_1 + n_2 + n_3)$$

مثال رقم (03-05): يبين الجدول التالي عدد الموظفين ومتوسط الراتب الشهري في ثلاثة فروع

لمؤسسة اقتصادية واحدة:

فروع المؤسسة	الفرع الأول	الفرع الثاني	الفرع الثالث
--------------	-------------	--------------	--------------

¹ راجع المرجعين: تيلولت سامية، مرجع سبق ذكره ص 57 وجلاطو جيلالي، مرجع سبق ذكره، ص 33.

عدد الموظفين	120	100	60
متوسط الراتب (دج)	42000	38000	45000

المطلوب: حساب متوسط الرواتب في المؤسسة ككل، دون الرجوع إلى قوائم رواتب كل موظف على حدة.

الحل: نضرب عدد موظفي كل فرع في متوسط راتبه، ثم نجمع النواتج ونقسمها على العدد الإجمالي للموظفين:

$$\bar{X} = (120 \times 42000 + 100 \times 38000 + 60 \times 45000) / (120 + 100 + 60)$$

$$\bar{X} = (5\,040\,000 + 3\,800\,000 + 2\,700\,000) / 280 = 11\,540\,000 / 280 \approx 41214.29 \text{ دج}$$

ويُلاحظ أن هذا الوسط العام (41214.29 دج) أقرب إلى متوسط الفرع الأول (42000 دج) الذي يضم أكبر عدد من الموظفين، وهذا منطقي لأن الوسط المرجح يعطي وزناً أكبر للفروع الأكثر عدداً.

مثال رقم (03-06): سجّلت ثلاث مدارس متوسط علامات تلاميذها في اختبار وطني موحد كما يلي:

المدسة	المدسة الأولى	المدسة الثانية	المدسة الثالثة
عدد التلاميذ	40	55	35
متوسط العلامة (من 20)	14.5	12.8	16.2

المطلوب: احسب متوسط علامات التلاميذ الثلاثة عبر المدارس مجتمعة.

الحل:

$$\bar{X} = (40 \times 14.5 + 55 \times 12.8 + 35 \times 16.2) / (40 + 55 + 35) = 1851 / 130 \approx 14.24 \text{ نقطة}$$

ونلاحظ أن هذا المتوسط العام (14.24) أقرب إلى متوسط المدرسة الثانية (12.8) نسبياً من متوسط المدرسة الثالثة (16.2)، رغم أن الفارق بينهما وبين المتوسط العام ليس متساوياً حسابياً، وذلك لأن المدرسة الثانية تضم أكبر عدد من التلاميذ (55) فيكون تأثيرها على المتوسط العام أكبر نسبياً من تأثير المدرستين الأخريين

ت. مميزات، عيوب وخصائص الوسط الحسابي

(1) مميزات الوسط الحسابي:

يتميز بسهولة حسابه وفهمه ودقة قياسه في وصف البيانات، خاصة إذا كانت موزعة توزيعاً متماثلاً أو قريباً من التوزيع الطبيعي (الجرسي الشكل).

قيمه ثابتة ومحددة لأنها محسوبة وفق معادلة رياضية صريحة، ولذا يُنظر إليه على أنه قيمة غير متحيزة نسبياً عن متوسط المجتمع الأصلي الذي أخذت منه العينة.

يمكن الاستفادة منه في حسابات أخرى لاحقة، مثل إيجاد مجموع قيم جميع أفراد العينة بضرب الوسط الحسابي في عدد أفراد هذه العينة مباشرة، دون الحاجة للرجوع إلى القيم الأصلية¹.

(2) عيوب الوسط الحسابي:

تتأثر قيمته بجميع قيم المشاهدات دون استثناء، ولذا فهو يتأثر بشدة بالقيم المتطرفة (الشاذة)، مما قد يجعله مضللاً في بعض الحالات كوصف متوسط ثروة مجموعة من الأفراد تضم مليارديراً واحداً وسط عمال بسطاء؛ ويُستحسن عندئذ استخدام مقياس آخر أقل حساسية للقيم المتطرفة، كالوسيط².

(3) خصائص الوسط الحسابي:

¹ سعيد السيد على إسماعيل، مرجع سبق ذكره، ص 88.

² نور الدين حامد، مرجع سبق ذكره، ص 75-76.

من أهم الخصائص الرياضية للوسط الحسابي، والتي تُستغل كثيراً في البراهين والتطبيقات الإحصائية اللاحقة:

مجموع انحرافات القيم عن وسطها الحسابي يساوي دائماً صفرًا: $\sum (x_i - \bar{X}) = 0$ في حالة البيانات غير المبوبة، و $\sum f_i (x_i - \bar{X}) = 0$ في حالة البيانات المبوبة؛ وهذه الخاصية هي أساس طريقة الوسط الفرضي المشروحة أعلاه¹.

يتأثر الوسط الحسابي بالعمليات الحسابية الأربع المطبقة على قيم المتغير: فإذا عُدلت جميع القيم وفق العلاقة الخطية $y = ax \pm b$ ، فإن الوسط الحسابي الجديد يصبح: $\bar{Y} = a\bar{X} \pm b^2$.

الوسط الحسابي لقيمة ثابتة يساوي هذه القيمة الثابتة نفسها، مهما كان عدد مرات تكرارها.

الوسط الحسابي لجداء مجموعة قيم في عدد ثابت يساوي جداء الوسط الحسابي لهذه القيم في هذا العدد الثابت³.

مجموع مربعات انحرافات القيم عن وسطها الحسابي هو أقل ما يمكن مقارنة بأي قيمة مرجعية أخرى غير الوسط الحسابي؛ وهذه الخاصية تُعرف بخاصية "أقل المربعات".

الوسط الحسابي لمجموع عدة عينات (بعد ضمها في عينة واحدة) هو نفسه مجموع الأوساط الحسابية المرجحة بأحجام هذه العينات، وهذا ما استُغل في المثال الخاص بالوسط الحسابي المرجح لأوساط حسابية أعلاه⁴.

تطبيقات عملية للوسط الحسابي: يُستخدم الوسط الحسابي في جل المجالات الاقتصادية والاجتماعية تقريباً: في الاقتصاد لحساب متوسط الدخل الفردي أو متوسط الإنفاق الأسري؛ في التعليم لحساب معدلات

¹ سهيل أحمد سمحان ومحمود حسين الوادي، مرجع سبق ذكره، ص 85.

² كامل فليفل وفتحي حمدان، نفس المرجع، ص 52.

³ موساوي عبد النور وبركان يوسف، مرجع سبق ذكره، ص 41.

⁴ لولت سامية، مرجع سبق ذكره، ص 59-60.

الطلبة؛ في الصحة لحساب متوسط وزن أو طول فئة عمرية معينة؛ وفي إدارة الأعمال لحساب متوسط الإنتاجية اليومية أو متوسط زمن إنجاز مهمة معينة. ويبقى الشرط الأساسي لصحة استعماله هو غياب قيم متطرفة قوية قد تُحرف الصورة الحقيقية للظاهرة.

الوسط الهندسي (Moyenne Géométrique)

في عدة حالات تكون قيم الظاهرة المدروسة عبارة عن نسب أو معدلات تتغير من فترة إلى أخرى بشكل تراكمي (مضاعف)، كمعدلات النمو السنوية أو معدلات الفائدة المركبة؛ وهي الحالات التي نرغب فيها بدراسة معدل تغير الظاهرة عبر الزمن. وفي مثل هذه الحالات، لا يصف الوسط الحسابي البسيط الظاهرة وصفاً سليماً، إذ يبالغ عادة في تقدير المعدل الحقيقي؛ لذا دعت الضرورة إلى إيجاد متوسط آخر أنسب لهذا النوع من البيانات، يُسمى بالوسط الهندسي، ويرمز له بالرمز (MG).

يُعرف الوسط الهندسي لمجموعة من القيم الموجبة عددها f بالجذر النوني لحاصل ضرب هذه القيم جميعها¹.

أ. طرق حساب الوسط الهندسي

❖ من البيانات غير المبوبة:

$$MG = \sqrt[n]{(x_1 * x_2 * \dots * x_n)}$$

ولما كان حساب الجذر النوني لحاصل ضرب عدة قيم عملية مرهقة يدوياً، خاصة عندما يكبر عدد القيم (f)، يُلجأ عملياً إلى اللوغاريتم لتحويل عملية الضرب إلى عملية جمع أبسط؛ فإذا أخذنا لوغاريتم الطرفين في العلاقة أعلاه نحصل على:

¹ محمد بوهزة، محاضرات في الإحصاء الوصفي، دار المحمدية العامة، الجزائر، 2011، ص 62.

$$\log MG = \log \left[\sqrt[n]{(x_1 * x_2 * \dots * x_n)} \right] = \frac{1}{n} * \log(x_1 * x_2 * \dots * x_n)$$

$$= \left(\frac{1}{n} \right) * [\log(x_1) + \log(x_2) + \dots + \log(x_n)]$$

وبإيجاز، تصبح العلاقة العملية المستعملة لحساب الوسط الهندسي كما يلي:

$$\log MG = (1/n) \sum \log(x_i)$$

ثم نُخرج الوسط الهندسي بأخذ اللوغاريتم العشري العكسي (الأس) للنتيجة.

مثال رقم (03-07): أرادت مؤسسة تقييم أداء خمس وحدات إنتاجية تابعة لها من خلال عدد الوحدات المنتجة يومياً؛ وكانت القيم المسجلة كما يلي (بالآلاف): 4، 8، 12، 6، 9. احسب الوسط الهندسي لهذه القيم.

$$\text{الحل: حاصل ضرب القيم الخمس} = 4 \times 8 \times 12 \times 6 \times 9 = 20736$$

$$MG = \sqrt[5]{20736} \approx 7.30 \text{ (آلاف وحدة)}$$

وللتحقق بطريقة اللوغاريتم: مجموع لوغاريتمات القيم =

$$\log(4) + \log(8) + \log(12) + \log(6) + \log(9) \approx 4.3167$$

وبقسمته على f=5 نحصل على 0.8633، وبأخذ الأس العشري لهذه القيمة:

$$MG = 10^{0.8633} \approx 7.30 \text{، وهي نفس النتيجة.}$$

❖ من البيانات المبوبة:

$$\log MG = (1/N) \sum n_i \log(x_i)$$

مثال رقم (03-08): متغير كمي منقطع: يمثل الجدول التالي عدد ساعات العمل الإضافي الأسبوعية لعينة من 60 عاملاً في ورشة صناعية:

ساعات العمل الإضافي (x_i)	1	2	3	4	5	المجموع
عدد العمال (f_i)	6	14	20	12	8	60

المطلوب: حساب الوسط الهندسي لعدد ساعات العمل الإضافي.

الحل:

x_i	f_i	$\log(x_i)$	$f_i \times \log(x_i)$
1	6	0	0
2	14	0.3010	4.214
3	20	0.4771	9.542
4	12	0.6021	7.225
5	8	0.6990	5.592
المجموع	60	/	26.573

$$MG = 10^{(26.573/60)} = 10^{0.4429} \approx 2.77 \text{ ساعة}$$

مثال رقم (03-09): يمثل الجدول التالي عدد الشهادات التدريبية التي حصل عليها 70 موظفاً خلال السنة:

عدد الشهادات (x_i)	1	2	3	4	المجموع
عدد الموظفين (f_i)	15	25	18	12	70

الحل: مجموع ($f_i \times \log(x_i)$) لجميع القيم يساوي 23.339:

$$MG = 10^{(23.339/70)} = 10^{0.3334} \approx 2.15 \text{ شهادة}$$

مثال رقم (10-03): متغير كمي مستمر: يمثل الجدول التالي توزيع المسافات (بالكيلومتر) التي تقطعها 48 شاحنة نقل يومياً لصالح إحدى شركات التوزيع:

المسافة (كم)	20-10	30-20	40-30	50-40	60-50	المجموع
عدد الشاحنات	4	10	16	12	6	48

الحل: باستعمال مراكز الفئات (15، 25، 35، 45، 55)، يكون مجموع $(f_i \times \log(x'_i))$ لجميع الفئات مساوياً لـ 73.670:

$$MG = 10^{(73.670/48)} = 10^{1.5348} \approx 34.26 \text{ كم}$$

ب. مجالات تطبيق الوسط الهندسي

تُعد مجالات تطبيق الوسط الهندسي أضيّق مقارنة بالوسط الحسابي، إذ يُستخدم أساساً في قياس النزعة المركزية للظواهر المتزايدة (أو المتناقصة) هندسياً عبر الزمن، ومن أهم هذه المجالات¹:

حساب الأرقام القياسية (كالرقم القياسي للأسعار)، إذ يُعد الوسط الهندسي من أفضل المتوسطات المستعملة في هذا المجال بالذات.

حساب معدلات النمو والتغير عبر الزمن، مثل معدل الفائدة المركبة، معدل نمو السكان، ومعدل النمو الاقتصادي السنوي لدولة معينة.

ت. خصائص الوسط الهندسي

من أبرز خصائص الوسط الهندسي التي تميّزه عن غيره من المتوسطات²:

- يدخل في حسابه جميع قيم التوزيع دون استثناء، لكنه أقل تأثراً بالقيم المتطرفة الكبيرة جداً مقارنة بالوسط الحسابي، بسبب استخدام اللوغاريتم الذي "يُلطّف" الفروق الكبيرة بين القيم.

¹ جلاطو جيلالي، مرجع سبق ذكره، ص 37.

² علي عبد السلام العماري و علي حسين العجيلي، الإحصاء والاحتمالات: النظرية والتطبيق، منشورات ELGA، مالطا، 2000، ص 57.

- لا يمكن حسابه من جداول التوزيع التكراري المفتوحة من أحد الطرفين أو كليهما.
- ليس له معنى رياضي إذا كانت إحدى قيم المتغير سالبة أو معدومة، لأن لوغاريتم العدد السالب أو الصفر غير معرف.
- قيمته تكون دائماً أصغر من أو تساوي قيمة الوسط الحسابي لنفس البيانات $(MG \leq X)$ ، والمساواة لا تتحقق إلا إذا كانت جميع القيم متساوية فيما بينها.

الوسط التوافقي (Moyenne Harmonique)

يُعرّف الوسط التوافقي لمجموعة من القيم الموجبة عددها f بأنه مقلوب الوسط الحسابي لمقلوبات هذه القيم، ويرمز له بالرمز $(MH)^1$.

ويستخدم الوسط التوافقي بصفة خاصة في الحالات التي تتضمن معدلات مركبة تشترك في بسط ثابت ومقام متغير، كمعدلات السرعة (كم لكل ساعة) عندما تكون المسافة ثابتة والزمن متغيراً، أو متوسط الأسعار عندما يكون المبلغ المنفق ثابتاً وكمية السلعة المشتراة متغيرة.

أ. طرق حساب الوسط التوافقي

(1) من البيانات غير المبوبة:

$$MH = n / \sum (1/x_i)$$

مثال رقم (03-11): احسب الوسط التوافقي للقيم الأربع التالية: 25، 40، 50، 20

الحل: نحسب أولاً مقلوب كل قيمة: $0.04=25/1$ ، $0.025=40/1$ ، $0.02=50/1$ ، $0.05=20/1$ ، ومجموع هذه المقلوبات = 0.135

$$MH = 4 / 0.135 \approx 29.63$$

¹ نور الدين حامد، مرجع سبق ذكره، ص 79.

(2) من البيانات المبوبة:

$$MH = \sum n_i / \sum (n_i / x_i)$$

مثال رقم (12-03): متغير كمي منقطع: يمثل الجدول التالي عدد القطع المطلوبة في 50 طلبية استلمت من طرف مصنع للأثاث خلال أسبوع معين:

عدد القطع (x_i)	3	5	7	9	المجموع
عدد الطلبيات (f_i)	12	18	15	5	50

الحل:

f_i / x_i	f_i	x_i
4	12	3
3.6	18	5
2.143	15	7
0.556	5	9
10.298	50	المجموع

$$MH = 50 / 10.298 \approx 4.86 \text{ قطعة}$$

مثال رقم (13-03): متغير كمي مستمر: يمثل الجدول التالي توزيع أزمنة التوصيل (بالدقائق) لعينة من 48 طلبية عبر تطبيق للتوصيل السريع:

زمن التوصيل (د)	15-5	25-15	35-25	45-35	المجموع
عدد الطلبيات	8	20	14	6	48

الحل: باستعمال مراكز الفئات، يكون مجموع (f_i / x'_i) لجميع الفئات مساوياً لـ 2.417:

$$MH = 48 / 2.417 \approx 19.86 \text{ دقيقة}$$

مثال رقم (03-14): يمثل الجدول التالي أزمنة تنقل (بالدقائق) لعينة من 50 موظفاً من منازلهم إلى مقر عملهم:

زمن التنقل (د)	20-10	30-20	40-30	50-40	المجموع
عدد الموظفين	10	18	14	8	50

الحل: باستعمال مراكز الفئات، مجموع (f_i/x'_i) لجميع الفئات يساوي 1.964:

$$MH = 50 / 1.964 \approx 25.45 \text{ دقيقة}$$

ب. خصائص الوسط التوافقي

الوسط التوافقي أقل تأثراً بالقيم المتطرفة الكبيرة من الوسط الحسابي، بل هو أكثر المتوسطات الثلاثة (الحسابي، الهندسي، التوافقي) حساسية للقيم الصغيرة.

قيمة الوسط التوافقي تكون دائماً أصغر من أو تساوي قيمة الوسط الهندسي، التي بدورها أصغر من أو تساوي الوسط الحسابي، أي: $MH \leq MG \leq X$.

يُستخدم في حالات خاصة جداً كحساب متوسط السرعة بدلالة وحدة الزمن، أو متوسط الأسعار عند شراء كميات متفاوتة بمبالغ متساوية، وهي حالات لا يصلح فيها الوسط الحسابي البسيط إطلاقاً¹.

لا يمكن حسابه من الجداول المفتوحة، ولا في حال وجود قيمة معدومة ضمن بيانات المتغير (لأن القسمة على صفر غير ممكنة رياضياً).

¹ سعيد السيد على إسماعيل، مرجع سبق ذكره، ص 101-102.

الوسط التربيعي (Moyenne Quadratique)

الوسط التربيعي لمجموعة من القيم هو الجذر التربيعي للوسط الحسابي لمربعات هذه القيم، ويرمز له بالرمز $(MQ)^1$.

ويُفيد هذا المقياس بشكل خاص عندما نرغب في حساب متوسط مجموعة من القيم التي قد تكون سالبة أو موجبة (كالانحرافات عن وسط ما)، مع تفادي تعويض الإشارات السالبة بعضها ببعض، إذ إن تربيع القيم يزيل أثر الإشارة قبل حساب المتوسط.

أ. طرق حساب الوسط التربيعي

(1) من البيانات غير المبوبة:

$$MQ = \sqrt{(\sum x_i^2 / n)}$$

مثال رقم (03-15): احسب الوسط التربيعي للقيم التالية: 6، 10، 8، 4، 12

الحل: مجموع مربعات القيم = $360 = 144 + 16 + 64 + 100 + 36$

$$MQ = \sqrt{(360/5)} = \sqrt{72} \approx 8.49$$

(2) من البيانات المبوبة:

$$MQ = \sqrt{(\sum n_i x_i^2 / \sum n_i)}$$

مثال رقم (03-16): متغير كمي منقطع: يمثل الجدول التالي عدد القطع المعيبة المكتشفة في 48

دفعة إنتاج بمصنع للأجهزة الكهرومنزلية:

عدد القطع المعيبة (x_i)	1	2	3	4	5	6	المجموع
-----------------------------	---	---	---	---	---	---	---------

¹ حمد راتول، مرجع سبق ذكره، ص 123.

عدد الدفعات (f_i)	5	9	14	10	7	3	48
-----------------------	---	---	----	----	---	---	----

الحل: مجموع ($f_i \times x_i^2$) لجميع القيم يساوي 610:

$$MQ = \sqrt{(610/48)} = \sqrt{12.708} \approx 3.56 \text{ قطعة}$$

مثال رقم (03-17): متغير كمي مستمر: يمثل الجدول التالي توزيع استهلاك الكهرباء الشهري (بالكيلوواط ساعة) لعينة من 56 أسرة:

الاستهلاك (ك.و.س)	40-0	80-40	120-80	160-120	200-160	المجموع
عدد الأسر	5	15	22	10	4	56

الحل: باستعمال مراكز الفئات، مجموع ($f_i \times x_i'^2$) لجميع الفئات يساوي 601600:

$$MQ = \sqrt{(601600/56)} = \sqrt{10742.86} \approx 103.65 \text{ ك.و.س}$$

ب. مجالات تطبيق الوسط التربيعي

يُعد الوسط التربيعي، كالوسط التوافقي، معقداً نسبياً في الحساب اليدوي ونادر الاستخدام المباشر في وصف الظواهر، غير أنه لا يمكن تعويضه في حالات معينة تتضمن قيماً سالبة نرغب في معرفة متوسطها بغض النظر عن إشارتها، أو عند حساب مقاييس التشتت (كالانحراف المعياري) التي سنتناولها في الفصل الموالي، والتي تعتمد أصلاً على فكرة الوسط التربيعي لانحرافات القيم عن وسطها الحسابي.

ت. العلاقة بين المتوسطات الأربعة

بالنسبة لأي مجموعة من القيم الموجبة غير المتساوية فيما بينها، فإن الوسط التوافقي أصغر من الوسط الهندسي، والوسط الهندسي أصغر من الوسط الحسابي، والوسط الحسابي بدوره أصغر من الوسط التربيعي؛ أي أن العلاقة العامة الثابتة رياضياً هي¹:

¹ تيلولت سامية، مرجع سبق ذكره، ص 82.

$$MH < MG < \bar{X} < MQ$$

مثال رقم (03-18): بأخذ معطيات جدول استهلاك الكهرباء الشهري (متغير كمي مستمر) المستعمل أعلاه لحساب الوسط التربيعي، وبعد حساب المتوسطات الأربعة على نفس البيانات، توصلنا إلى النتائج التالية:

- ك.و.س $MQ \approx 103.65$
- ك.و.س $\bar{X} = 95$
- ك.و.س $MG \approx 83.66$
- ك.و.س $MH \approx 68.83$

وعليه فإن العلاقة ($MH < MG < \bar{X} < MQ$) محققة تماماً في هذا المثال: $83.66 > 68.83$ و $103.65 > 95$ ، وهذا يوضح عملياً كيف يتزايد ترتيب المتوسطات الأربعة بشكل ثابت بغض النظر عن طبيعة البيانات المدروسة، طالما أنها قيم موجبة غير متساوية.

الوسيط (Médiane)

هو المقياس الثاني من مقاييس النزعة المركزية من حيث الأهمية والاستعمال بعد الوسط الحسابي، ويُعرّف بأنه القيمة التي يتساوى عندها عدد القيم الأقل منها مع عدد القيم الأكبر منها، بعد ترتيب جميع البيانات تصاعدياً أو تنازلياً؛ ومن أهم ميزاته أنه لا يتأثر بالقيم الشاذة (المتطرفة) على عكس الوسط الحسابي¹.

يرمز للوسيط بالرمز (Me)، وتختلف طريقة حسابه بحسب طبيعة البيانات (مبوبة أو غير مبوبة) وعدد المشاهدات (فردية أو زوجية).

أ. طرق حساب الوسيط

¹ سهيل أحمد سمحان ومحمود حسين الوادي، مرجع سبق ذكره، ص 76.

أ. من البيانات غير المبوبة:

- حالة سلسلة فردية العدد: إذا كان عدد القيم f فردياً، فإن الوسيط هو القيمة التي رتبها بعد الترتيب التصاعدي تساوي $t = (f+1)/2$.

مثال رقم (03-19): تمثل البيانات التالية أعمار 15 متدرباً التحقوا بدورة تكوينية متخصصة في إحدى المؤسسات:

28، 34، 22، 30، 26، 19، 33، 25، 31، 27، 24، 29، 36، 21، 32

المطلوب: تحديد السن الوسيط لهؤلاء المتدربين.

الحل: عدد القيم $f = 15$ (فردية)، ورتبة الوسيط $t = (15+1)/2 = 8$. وبعد الترتيب التصاعدي للقيم نحصل على السلسلة: 19، 21، 22، 24، 25، 26، 27، 28، 29، 30، 31، 32، 33، 34، 36

$$\text{سنة } Me = x_8 = 28$$

أي أن نصف المتدربين أعمارهم أقل من 28 سنة، ونصفهم الآخر أعمارهم أكبر منها.

- حالة سلسلة زوجية العدد: إذا كان عدد القيم f زوجياً، فلا توجد قيمة وسطى واحدة، ويُحسب الوسيط كمتوسط حسابي للقيمتين اللتين رتبتهما $(f/2)$ و $(f/2 + 1)$.

مثال رقم (03-20): بإضافة متدرب سادس عشر عمره 23 سنة إلى المجموعة السابقة، يصبح عدد المتدربين $f = 16$ (زوجي). والمطلوب إعادة حساب السن الوسيط.

الحل: بعد الترتيب التصاعدي (16 قيمة): 19، 21، 22، 23، 24، 25، 26، 27، 28، 29، 30، 31، 32، 33، 34، 36.

رتبنا الوسيط: $t_1 = 16/2 = 8$ و $t_2 = 9$ ، أي القيمتين $x_8 = 27$ و $x_9 = 28$:

$$\text{سنة } Me = (x_8 + x_9) / 2 = (27 + 28) / 2 = 27.5$$

ب. من البيانات المبوبة:

- حالة متغير كمي منقطع: نحسب أولاً التكرار المتجمع الصاعد لجميع القيم، ثم نحدد رتبة الوسيط $t = F/2$ ، ثم نبحث عن أول تكرار متجمع صاعد يساوي هذه الرتبة أو يفوقها مباشرة؛ والقيمة المقابلة له هي الوسيط.

مثال رقم (03-21): يمثل الجدول التالي عدد أيام التغيب عن العمل لعينة من 46 موظفاً خلال شهر معين، وفق سجلات مصلحة الموارد البشرية:

أيام التغيب (x_i)	0	1	2	3	4	5	المجموع
عدد الموظفين (f_i)	15	10	9	6	4	2	46

الحل: التكرار المتجمع الصاعد هو: 15، 25، 34، 40، 44، 46. رتبة الوسيط: $t = F/2 = 23$. نجد أن التكرار المتجمع 25 هو أول تكرار يفوق 23 مباشرة (بما أن $25 > 23 > 15$)، وهو يقابل القيمة 1.

$$Me = 1 \text{ يوم}$$

أي أن نصف الموظفين تقريباً لم يتغيّبوا سوى يوم واحد على الأكثر خلال الشهر.

- حالة متغير كمي مستمر: تُحدّد أولاً الفئة الوسيطة (وهي الفئة التي يقع عندها التكرار المتجمع الصاعد المساوي أو المتجاوز مباشرة لرتبة $F/2$)، ثم يُحسب الوسيط بالاستيفاء الخطي داخل هذه الفئة:

$$M_e = A + \left[\left(\frac{N}{2} - F \uparrow \right) / nM_e \right] * K$$

مثال رقم (03-22): يمثل الجدول التالي توزيع 100 موظف حسب أجورهم الشهرية (وحدة القياس: 10^3 دج) في إحدى المؤسسات العمومية:

الأجر	30-20	40-30	50-40	60-50	70-60	80-70	المجموع
عدد الموظفين	12	22	28	20	11	7	100

الحل: التكرار المتجمع الصاعد: 12، 34، 62، 82، 93، 100. رتبة الوسيط: $F/2 = 50$. نجد أن التكرار المتجمع 62 يفوق 50 مباشرة ($62 > 50 > 34$)، فالفئة الوسيطة هي $[50-40]$:

$$Me = 40 + [(50 - 34) / 28] \times 10 \approx 45.71 (10^3 \text{ دج})$$

مثال رقم (03-23): يمثل الجدول التالي توزيع 80 طلبية حسب مسافة التوصيل (بالكيلومتر):

مسافة التوصيل (كم)	10-0	20-10	30-20	40-30	50-40	المجموع
عدد الطلبيات	8	20	26	18	8	80

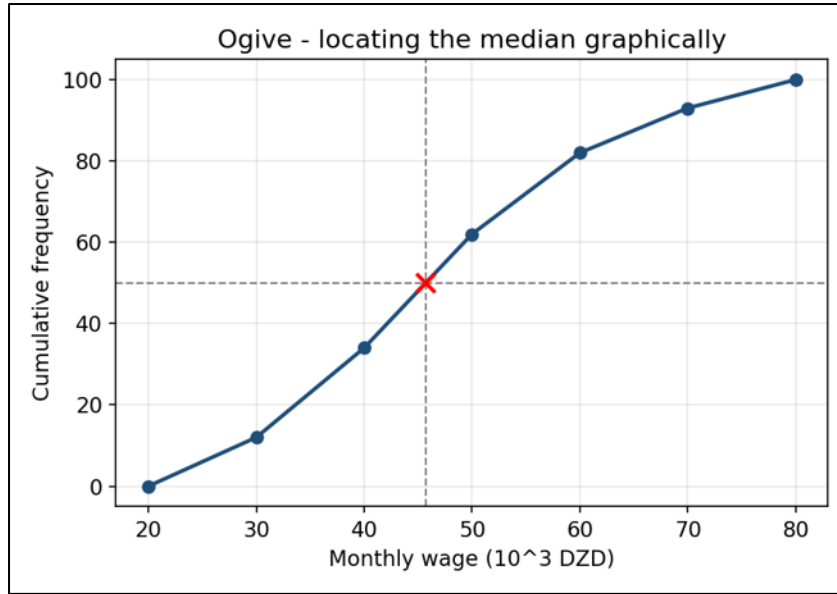
الحل: التكرار المتجمع الصاعد: 8، 28، 54، 72، 80. رتبة الوسيط: $F/2 = 40$. نجد أن 54 يفوق 40 مباشرة ($54 > 40 > 28$)، فالفئة الوسيطة $[30-20]$:

$$Me = 20 + [(40 - 28) / 26] \times 10 \approx 24.62 \text{ كم}$$

ب. الطريقة البيانية لإيجاد الوسيط

يمكن تحديد قيمة الوسيط بيانياً دون اللجوء إلى الحساب الجبري، وذلك برسم منحنى التكرار المتجمع الصاعد (الأوجيف)، ثم رسم مستقيم أفقي من رتبة الوسيط ($F/2$) على المحور العمودي حتى يلامس المنحنى، ثم إسقاط عمودي من نقطة التماس هذه إلى المحور الأفقي؛ ونقطة تقاطع هذا العمود مع المحور الأفقي تعطي القيمة التقريبية للوسيط. ويمكن أيضاً استخدام تقاطع منحنى التكرار المتجمع الصاعد والنازل معاً لتحديد الوسيط بنفس الطريقة.

توضيح: نطبق هذه الطريقة على بيانات مثال الأجور الشهرية أعلاه (الجدول ذي الفئات 20-80، $F=100$)؛ والشكل التالي يمثل منحنى التكرار المتجمع الصاعد لهذا التوزيع، مع تحديد نقطة الوسيط ($Me \approx 45.71$) بيانياً:



الشكل 3-1: تحديد الوسيط بيانياً بواسطة منحنى التكرار المتجمع الصاعد

ونلاحظ أن النقطة التي تقابل التكرار المتجمع 50 (نصف حجم العينة) على المنحنى تسقط عمودياً عند القيمة 45.71 على محور الأجور، وهي نفس القيمة التي توصلنا إليها جبرياً بالاستيفاء الخطي، مما يؤكد صحة الطريقتين.

ت. خصائص الوسيط

لا يتأثر الوسيط بالقيم الشاذة (المتطرفة)، عكس الوسط الحسابي الذي يتأثر بها بشدة.

يتأثر الوسيط بعدد المشاهدات، إذ تتغير رتبته عند حذف أو إضافة قيم جديدة للسلسلة، وقد تتغير قيمته تبعاً لذلك.

يمكن إيجاده من الجداول المفتوحة من أحد الطرفين أو كليهما، وكذلك بيانياً كما تم بيانه أعلاه¹.

¹ سهيل أحمد سمحان ومحمود حسين الوادي، نفس المرجع، ص 87-88.

يُستعمل الوسيط في عدة مجالات تطبيقية مهمة، منها دراسة توزيع الأجور والأسعار (لتقادي تأثير الرواتب الاستثنائية)، ودراسة الوفيات والعمر المتوقع، ودراسة توزيع الثروات¹.

يقسم الوسيط المدرج التكراري إلى مساحتين متساويتين تماماً، وهذه الخاصية الهندسية هي التي تُستغل في الطريقة البيانية لتحديده.

ث. أشباه الوسيط

بما أنه يمكن تقسيم بيانات أي ظاهرة إلى أي عدد من الأقسام المتساوية (وليس فقط قسمين متساويين كما في حالة الوسيط)، فإنه يمكن التعامل مع القيم التي تقسم هذه البيانات بنفس منطق التعامل مع الوسيط تماماً، مع تغيير رتبة القيمة المطلوبة فقط. وأشباه الوسيط الأكثر استعمالاً هي: الربعيات، العشيريات والمئينيات.

1) الربعيات (Les Quartiles):

يُرمز لها بالرمز (Q_i) ، وهي ثلاث قيم (Q_1, Q_2, Q_3) تقسم التوزيع المرتب تصاعدياً إلى أربعة أقسام متساوية العدد، علماً أن الربع الثاني (Q_2) هو نفسه الوسيط (Me) .

- في حالة البيانات غير المبوبة: نتبع نفس خطوات حساب الوسيط بالضبط، مع تعديل صيغة الرتبة فقط: $t = i(f+1)/4$ ، حيث $i = 1$ لحساب الربع الأول، و $i = 3$ لحساب الربع الثالث.

مثال رقم (03-24): بأخذ بيانات درجات 15 طالباً في اختبار مادة معينة (بيانات مستقلة عن مثال أعمار المتدربين السابق)، وهي كما يلي:

12، 15، 9، 18، 7، 14، 11، 16، 13، 10، 17، 8، 19، 6، 20

¹ جلاطو جيلالي، مرجع سبق ذكره، ص 45

المطلوب: تحديد الربع الأول والربع الثالث لهذا التوزيع.

الحل: بعد الترتيب التصاعدي: 6، 7، 8، 9، 10، 11، 12، 13، 14، 15، 16، 17، 18، 19، 20 (f=15).

رتبة الربع الأول: $t_1 = 1 \times 16/4 = 4$

رتبة الربع الثالث: $t_3 = 3 \times 16/4 = 12$

نقطة $Q_1 = x_4 = 9$

نقطة $Q_3 = x_{12} = 17$

أي أن ربع الطلبة تحصلوا على أقل من 9 نقاط، بينما ربعهم الآخر تحصل على أكثر من 17 نقطة.

- في حالة البيانات المبوبة (متغير كمي مستمر): نتبع نفس منطق حساب الوسيط بالاستيفاء الخطي، مع تعديل الرتبة إلى $(iF/4)$:

$$Q_i = A + \left[\left(\frac{iN}{4} - F \uparrow \right) / nQ_i \right] * K$$

مثال: يمثل الجدول التالي توزيع 80 عاملاً حسب زمن إنجاز مهمة إنتاجية معينة (بالدقائق) في خط إنتاج واحد:

زمن الإنجاز (د)	25-15	35-25	45-35	55-45	65-55	75-65	المجموع
عدد العمال	10	18	22	16	9	5	80

الحل: التكرار المتجمع الصاعد: 10، 28، 50، 66، 75، 80. رتبة الربع الأول: $F/4 = 20$.

نجد أن 28 يفوق 20 مباشرة ($28 > 20 > 10$)، فالفترة الربيعية الأولى هي [25-35]:

$$Q1 = 25 + [(20 - 10) / 18] \times 10 \approx 30.56 \text{ دقيقة}$$

رتبة الربع الثالث: $F/4 = 603$. نجد أن 66 يفوق 60 مباشرة ($66 > 60 > 50$)، فالفئة الربعية الثالثة هي $[55-45]$:

$$Q3 = 45 + [(60 - 50) / 16] \times 10 = 51.25 \text{ دقيقة}$$

مثال رقم (03-25): بأخذ بيانات مسافات التوصيل الثمانين المستعملة أعلاه في مثال الوسيط المكرر ($F=80$):

رتبة الربع الأول: $F/4 = 20$. نجد أن 28 يفوق 20 مباشرة ($28 > 20 > 8$)، فالفئة الربعية الأولى $[20-10]$:

$$Q1 = 10 + [(20 - 8) / 20] \times 10 = 16 \text{ كم}$$

رتبة الربع الثالث: $F/4 = 603$. نجد أن 72 يفوق 60 مباشرة ($72 > 60 > 54$)، فالفئة الربعية الثالثة $[40-30]$:

$$Q3 = 30 + [(60 - 54) / 18] \times 10 \approx 33.33 \text{ كم}$$

تطبيقات عملية للوسيط والربيعيات: تُستخدم هذه المقاييس بكثرة في الدراسات الاجتماعية والاقتصادية، مثل تحديد "خط الفقر" اعتماداً على نسبة من وسيط الدخل، أو تصنيف تلاميذ مدرسة إلى أرباع حسب نتائجهم (الربع الأعلى تحصيلاً، الربع الأدنى، وهكذا) لتوجيه برامج الدعم التربوي المناسبة لكل فئة.

(2) العشيريات (Les Déciles):

وهي تسع قيم ($D1$ إلى $D9$) تقسم التوزيع المرتب تصاعدياً إلى عشرة أجزاء متساوية العدد، تقابل رتبها على التوالي 10%، 20%، ...، 90% من مجموع التكرارات؛ وتُحسب بنفس منطق حساب الوسيط تماماً، مع استبدال الرتبة برتبة $(i \times F/10)$.

ت. المئينات (Les Percentiles):

وهي تسع وتسعون قيمة (P1 إلى P99) تقسم التوزيع المرتب تصاعدياً إلى مائة جزء متساوية العدد، وتُحسب بنفس المنطق أيضاً، مع استبدال الرتبة برتبة $(i \times F/100)$.

ملاحظة مهمة: العشير الخامس والمئين الخمسون يتطابقان دوماً مع الوسيط ومع الربع الثاني، أي: $D5 = P50 = Q2 = Me$ ، مهما اختلفت طبيعة البيانات المدروسة.

مثال تطبيقي على العشيرات والمئينات: بأخذ بيانات جدول الأجور الشهرية المستعمل سابقاً (الفئات 20-80، $F=100$ ، والتكرار المتجمع الصاعد: 12، 34، 62، 82، 93، 100)، احسب العشير الثالث (D3) والمئين الخامس والسبعين (P75).

الحل: رتبة العشير الثالث: $D3 = 3F/10 = 30$. نجد أن التكرار المتجمع 34 يفوق 30 مباشرة $(12 < 30 < 34)$ ، فالقئة العشرية [30-40]:

$$D3 = 30 + [(30 - 12) / 22] \times 10 \approx 38.18 \text{ (دج } 10^3 \text{)}$$

رتبة المئين 75: $P75 = 75F/100 = 75$. نجد أن التكرار المتجمع 82 يفوق 75 مباشرة (62 < 75 < 82)، فالقئة المئينية [50-60]:

$$P75 = 50 + [(75 - 62) / 20] \times 10 = 56.5 \text{ (دج } 10^3 \text{)}$$

ونلاحظ أن قيمة المئين 75 (56.5) تساوي بالضبط قيمة الربع الثالث (Q3) المحسوبة سابقاً لنفس البيانات، وهذا منطقي تماماً لأن الربع الثالث ليس إلا حالة خاصة من المئينات تقابل الرتبة 75%، أي: $Q3 = P75$ دائماً.

المنوال (Mode)

المنوال هو المقياس الثالث من مقاييس النزعة المركزية، ويُعرّف بأنه القيمة الأكثر تكراراً أو شيوعاً بين مشاهدات المتغير المدروس، ويمكن حسابه لجميع أنواع البيانات (كمية أو وصفية على حد سواء)،

بخلاف باقي المتوسطات التي تقتصر على البيانات الكمية فقط. وقد يوجد أكثر من منوال واحد لنفس السلسلة، فتكون عندئذ ثنائية أو متعددة المنوال، وقد لا يوجد منوال على الإطلاق إذا تساوت جميع التكرارات¹.

يرمز للمنوال بالرمز (Mo).

أ. طرق حساب المنوال

(1) من البيانات غير المبوبة:

يكفي حصر عدد مرات تكرار كل قيمة، والمنوال هو القيمة صاحبة أكبر تكرار.

مثال رقم (03-26): لدينا السلاسل الإحصائية الثلاث التالية، المستقلة عن بعضها البعض:

A: 5 ، 3 ، 9 ، 6 ، 3 ، 8 ، 3

B: 4 ، 6 ، 9 ، 1 ، 7 ، 9 ، 2 ، 4 ، 7

C: 3 ، 9 ، 15 ، 7 ، 11

المطلوب: حساب منوال كل سلسلة والتعليق على النتيجة.

الحل:

بالنسبة للسلسلة A: القيمة 3 هي الأكثر تكراراً (تتكرر 3 مرات مقابل مرة واحدة لكل قيمة أخرى)، ومنه $Mo = 3$ ، وتكون السلسلة أحادية المنوال.

¹ تيلولت سامية، مرجع سبق ذكره، ص 68.

بالنسبة للسلسلة B: نلاحظ أن القيم 4، 7، 9 تتكرر مرتين لكل منها، وهو أكبر تكرار في هذه السلسلة، فتكون السلسلة متعددة المنوال (ثلاثية): $Mo1 = 4$ ، $Mo2 = 7$ ، $Mo3 = 9$.

بالنسبة للسلسلة C: جميع القيم تتكرر مرة واحدة فقط، ولا توجد أي قيمة أكثر تكراراً من غيرها، فلا يوجد منوال لهذه السلسلة.

ث. من البيانات المبوبة:

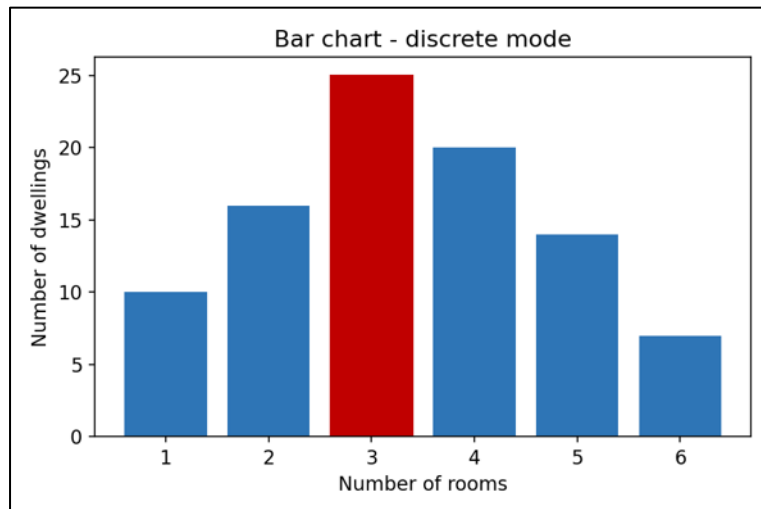
- في حالة متغير كمي منقطع: يكون المنوال حسابياً هو قيمة المتغير المقابلة لأكبر تكرار في الجدول التكراري مباشرة، دون حاجة لأي حساب إضافي.

مثال رقم (03-27): يمثل الجدول التالي توزيع 92 سكناً حسب عدد الغرف، في حي سكني معين:

عدد الغرف (x_i)	1	2	3	4	5	6	المجموع
عدد المساكن (f_i)	10	16	25	20	14	7	92

الحل: أكبر تكرار هو 25 ويقابل القيمة 3، إذن $Mo = 3$ غرف؛ أي أن أغلب المساكن في هذا الحي هي مساكن من ثلاث غرف.

والشكل التالي يمثل بيانياً هذا التوزيع بواسطة أعمدة بسيطة، مع إبراز العمود المنوالي بلون مغاير:



الشكل 3-2: تمثيل بياني لتوزيع المساكن حسب عدد الغرف، وتحديد المنوال

- في حالة متغير كمي مستمر: تُحدّد أولاً الفئة المنوالية، وهي الفئة المقابلة لأكبر تكرار، ثم يُحسب المنوال داخلها بطريقة بيرسون (طريقة الفروق):

$$M_o = A + [D_1 / (D_1 + D_2)] * K$$

حيث:

- A الحد الأدنى للفئة المنوالية
- D_1 = الفرق بين تكرار الفئة المنوالية والفئة السابقة لها
- D_2 = الفرق بين تكرار الفئة المنوالية والفئة الموالية لها
- K طول الفئة.

مثال رقم (03-28): يمثل الجدول التالي زمن انتظار 40 زبوناً في صالة انتظار بأحد مكاتب البريد (بالدقائق):

زمن الانتظار	5-0	10-5	15-10	20-15	25-20	المجموع
عدد الزبائن	5	9	14	8	4	40

الحل: الفئات متساوية الطول ($K=5$)، وأكبر تكرار هو 14 في الفئة الثالثة [15-10]، فهي إذن الفئة المنوالية:

$$D_1 = 14 - 9 = 5, D_2 = 14 - 8 = 6$$

$$M_o = 10 + [5 / (5+6)] \times 5 \approx 12.27 \text{ دقيقة}$$

والشكل التالي يمثل المدرج التكراري لهذا التوزيع، مع إبراز الفئة المنوالية بلون مغاير:



الشكل 3-3: المدرج التكراري لأزمنة الانتظار، وتحديد الفئة المنوالية

مثال رقم (03-29): يمثل الجدول التالي توزيع 56 طالباً حسب علاماتهم في اختبار مادة معينة (من 50 نقطة):

العلامات	10-0	20-10	30-20	40-30	50-40	المجموع
عدد الطلبة	4	10	22	14	6	56

الحل: الفئات متساوية الطول ($K=10$)، وأكبر تكرار هو 22 في الفئة الثالثة [30-20]، فهي إذن الفئة المنوالية:

$$D1 = 22 - 10 = 12, D2 = 22 - 14 = 8$$

$$Mo = 20 + [12 / (12+8)] \times 10 = 26 \text{ نقطة}$$

تطبيقات عملية للمنوال: يُستخدم المنوال بكثرة في دراسات السوق وتفضيلات المستهلكين (كتحديد المقاس أو اللون الأكثر مبيعاً)، وفي دراسة النقل العام (تحديد الوقت الأكثر ازدحاماً)، وبصفة عامة في كل ظاهرة وصفية (كيفية) لا يمكن حساب وسط حسابي لها أصلاً.

ب. خصائص المنوال

من أهم خصائص المنوال التي يمتاز بها عن باقي مقاييس النزعة المركزية¹:

- أسهل مقاييس النزعة المركزية حساباً وأبسطها فهماً، إذ لا يتطلب أي عملية حسابية معقدة في حالة البيانات غير المبوبة أو المتغير المنقطع.
- لا يتأثر بالقيم المتطرفة إطلاقاً، وهو غير قابل للعمليات الجبرية المعقدة (كإيجاد منوال مجموع متغيرين من منوالي كل منهما على حدة).
- يمكن حسابه بيانياً بواسطة المدرج التكراري، وكذلك من الجداول المفتوحة من أحد الطرفين أو كليهما، بخلاف الوسط الحسابي والهندسي والتوافقي.
- يُعد أفضل مقاييس النزعة المركزية لوصف الظواهر الوصفية (الكيفية) التي لا يمكن حساب متوسط حسابي لها أصلاً، كاللون المفضل أو التخصص الأكثر إقبالاً.

ت. العلاقة بين مقاييس النزعة المركزية الثلاثة

تختلف العلاقة بين الوسط الحسابي والوسيط والمنوال بحسب شكل التوزيع التكراري (منحنى تكراراته):

- إذا كان التوزيع التكراري متماثلاً تماماً (منحنى جرسى الشكل، كالتوزيع الطبيعي)، فإن المتوسطات الثلاثة تتطابق:

$$\bar{X} = M_e = M_o$$

- إذا كان منحنى التوزيع ملتوياً نحو اليمين (التواء موجب، أي أن الذيل الطويل للمنحنى يقع على اليمين)، فإن الترتيب يكون: $M_o < M_e < \bar{X}$.
- إذا كان منحنى التوزيع ملتوياً نحو اليسار (التواء سالب، أي أن الذيل الطويل يقع على اليسار)، فإن الترتيب يكون معكوساً: $M_o > M_e > \bar{X}$.

¹ علي عبد السلام العماري و علي حسين العجيلي، مرجع سبق ذكره، ص 67.

وتفيد العلاقة التقريبية التجريبية التالية، التي وضعها الإحصائي كارل بيرسون، في تقدير أحد هذه المقاييس الثلاثة تقديراً تقريبياً عند تعذر حسابه مباشرة (كأن يكون الجدول مفتوحاً من أحد طرفيه ويمنع حساب الوسط الحسابي مثلاً):

$$M_o \approx 3M_e - 2\bar{X}$$

مثال تطبيقي: لتوضيح هذه العلاقة التجريبية بمثال رقمي، لنفترض توزيعاً لمبالغ مشتريات شهرية لعينة من العملاء (سيرد تفصيله في التمرين الخامس لاحقاً)، حسبنا له: $M_e = 44.44$ و $\bar{X} = 45.33$. فإذا رغبتنا في تقدير المنوال بالعلاقة التجريبية دون حسابه مباشرة، نطبق العلاقة كما يلي:

$$M_o \approx 3 \times 44.44 - 2 \times 45.33 = 133.32 - 90.66 = 42.66$$

وهذه القيمة التقريبية (42.66) قريبة جداً من القيمة الفعلية للمنوال المحسوبة مباشرة بطريقة بيرسون لنفس البيانات (44 كما سيظهر في حل التمرين الخامس)، والفارق البسيط بينهما مقبول تماماً كون العلاقة تجريبية (تقريبية) وليست دقيقة رياضياً بشكل مطلق؛ وتبقى فائدتها الأساسية في الحالات التي يتعذر فيها حساب المنوال مباشرة، كالجداول المفتوحة من الطرفين.

خلاصة: يلخص الجدول التالي أهم مقاييس النزعة المركزية المدروسة في هذا الفصل، ومجال استعمال كل منها:

المقياس	يتأثر بالقيم المتطرفة؟	يُحسب من الجداول المفتوحة؟	أنسب مجال استعمال
الوسط الحسابي	نعم، بشدة	لا	البيانات المتماثلة دون قيم شاذة
الوسط الهندسي	قليلاً	لا	معدلات النمو والنسب
الوسط التوافقي	قليلاً	لا	السرعات والأسعار (نسب مركبة)

الوسيط	لا	نعم	توزيعات ملتوية أو بها قيم شاذة
المنوال	لا	نعم	البيانات الوصفية (الكيفية)

خطوات عملية لاختيار المقياس المناسب: يواجه الباحث أحياناً صعوبة في تحديد أي مقياس نزعة مركزية يستعمل؛ ويمكن الاسترشاد بالقواعد العملية التالية: إذا كانت البيانات كمية ومتماثلة تقريباً دون قيم شاذة، يُفضّل الوسيط الحسابي لدقته. إذا كانت البيانات تحتوي قيماً متطرفة أو كان التوزيع ملتوياً بشدة، يُفضّل الوسيط لأنه أكثر مقاومة لهذه القيم. إذا كانت البيانات عبارة عن معدلات نمو متتالية، يُفضّل الوسيط الهندسي. إذا كانت البيانات عبارة عن معدلات مركبة (سرعة، سعر بوحدة الكمية)، يُفضّل الوسيط التوافقي. أما إذا كانت البيانات وصفية (كيفية) لا تقبل الترتيب أو الجمع أصلاً، فلا خيار سوى المنوال.

غير أن هذه المقاييس، على أهميتها، تبقى غير كافية وحدها لوصف البيانات وصفاً شاملاً، إذ إنها لا تخبرنا شيئاً عن مدى تشتت أو تقارب القيم حول القيمة المركزية؛ فقد يتساوى توزيعان في وسطهما الحسابي تماماً، ومع ذلك يكون أحدهما شديد التجانس (قيمه متقاربة جداً من الوسيط) بينما يكون الآخر شديد التباين (قيمه متباعدة جداً عن الوسيط). ولسدّ هذه الثغرة بالذات، سنخصص الفصل الموالي لدراسة مقاييس التشتت، التي تكمل مقاييس النزعة المركزية وتمنحنا صورة إحصائية أكثر اكتمالاً ودقة عن الظاهرة المدروسة.

تمارين الفصل الثالث

التمرين الأول:

لتكن لدينا السلاسل الإحصائية الثلاث التالية، المتعلقة بعدد الزبائن الذين استقبلهم ثلاثة موظفي استقبال مختلفين خلال أسبوع عمل واحد (6 أيام):

A: 6، 2، 2، 7، 4، 9، 2

B: 5، 3، 2، 6، 3

C: 3، 3، 5، 12، 7، 10

المطلوب:

1. احسب الوسط الحسابي لكل سلسلة بطريقتين مختلفتين.
2. احسب الوسيط والمنوال لكل سلسلة.
3. إذا اعتبرنا هذه السلاسل الثلاث مجموعة واحدة D (عدد الزبائن الذين استقبلهم الموظفون الثلاثة مجتمعين)، احسب لها الوسط الحسابي، الوسيط والمنوال.

الحل:

الموظف	f	مجموع القيم	الوسط الحسابي	الوسيط	المنوال
A	7	32	4.57	4	2
B	5	19	3.8	3	3
C	6	40	6.67	6	3

بالنسبة للوسط الحسابي، يمكن حسابه إما بالطريقة المباشرة (جمع القيم وقسمة الناتج على عددها)، أو بطريقة الوسط الفرضي (افتراض وسط تقريبي ثم إضافة متوسط الانحرافات عنه)، وتعطي الطريقتان دوماً نفس النتيجة كما سبق بيانه بالتفصيل في الدرس.

أما الوسيط، فيُحسب بترتيب قيم كل سلسلة تصاعدياً ثم أخذ القيمة الوسطى (لأن f فردي في السلسلتين A و C) أو متوسط القيمتين الوسطيتين (لأن f زوجي في السلسلة B).

والمنوال هو القيمة الأكثر تكراراً في كل سلسلة: القيمة 2 تتكرر 3 مرات في السلسلة A، والقيمة 3 تتكرر مرتين في كل من السلسلتين B و C.

3- بدمج السلاسل الثلاث في مجموعة واحدة D (18 قيمة إجمالاً، ومجموعها الكلي 91):

$$X(D) = 91 / 18 \approx 5.06$$

وبعد ترتيب قيم D تصاعدياً بالكامل (18 قيمة، وهو عدد زوجي)، يقع الوسيط بين الرتبتين 9 و 10 (القيمتان 4 و 5 على التوالي):

$$Me(D) = (4+5)/2 = 4.5$$

أما المنوال، فبحصر تكرارات جميع القيم الثماني عشرة نجد أن القيمتين 2 و 3 تتكرر كل واحدة منهما أربع مرات، وهو أكبر تكرار مسجل في كامل المجموعة، فتكون المجموعة D ثنائية المنوال: $Mo_1 = 2$ ، $Mo_2 = 3$.

التمثيل البياني: يمكن تمثيل توزيع المجموعة D بمدرج تكراري بسيط (أعمدة) يوضح تكرار كل قيمة على حدة، وسيظهر فيه عمودان متساويا الارتفاع عند القيمتين 2 و 3 يفوقان في الارتفاع باقي الأعمدة، وهو ما يؤكد بيانياً ازدواجية المنوال المستنتجة حسابياً.

التمرين الثاني:

يمثل التوزيع التالي مساحة عينة من 100 شقة سكنية معروضة للبيع في مدينة معينة (بالمتر المربع):

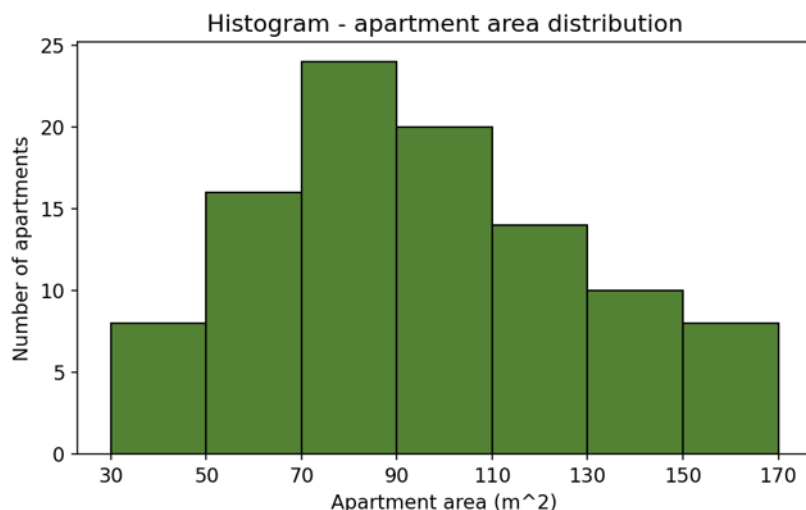
الفئات	50-30	70-50	90-70	-90	-110	-130	-150
				110	130	150	170
عدد الشقق	8	16	24	20	14	10	8

المطلوب:

1. حدد نوعية الخاصية المدروسة (المتغير الإحصائي).
2. مثل هذه المعطيات بيانياً بواسطة المدرج التكراري.
3. احسب متوسط مساحة الشقق بطريقتين مختلفتين.
4. احسب الوسيط.
5. احسب المنوال.

الحل:

1. المتغير المدروس هو مساحة الشقة السكنية، وهو متغير كمي مستمر، لأنه قابل للقياس بأي درجة من الدقة (يمكن أن تكون المساحة 85 م² أو 85.5 م² أو 85.25 م² وهكذا).
2. يُمثل هذا التوزيع بيانياً بواسطة المدرج التكراري (Histogramme): وهو عبارة عن مجموعة مستطيلات متجاورة بدون فواصل بينها، قاعدة كل مستطيل تساوي طول فئته (20 م² في هذا المثال) وارتفاعه يتناسب طردياً مع تكرارها. والشكل التالي يمثل هذا المدرج:



الشكل 1-0: المدرج التكراري لتوزيع مساحات الشقق السكنية

3. حساب المتوسط: نستعمل مراكز الفئات ونطبق الطريقة المباشرة:

الفئات	f_i	x'_i	$f_i \times x'_i$	$F \uparrow$
30-50	8	40	320	8
50-70	16	60	960	24
70-90	24	80	1920	48
90-110	20	100	2000	68
110-130	14	120	1680	82
130-150	10	140	1400	92
150-170	8	160	1280	100
المجموع	100	/	9560	/

$$\bar{X} = 9560 / 100 = 95.6^2 \text{ م}$$

وللتحقق بطريقة الانحرافات، بأخذ $X_0 = 100$ (وهو مركز قريب من الفئة الأكبر تكراراً)، نجد

أن مجموع $(f_i \times (x'_i - 100))$ يساوي -440، وبالتالي:

$$\bar{X} = 100 + (-440/100) = 100 - 4.4 = 95.6^2 \text{ م}$$

وتتطابق النتيجة تماماً مع الطريقة المباشرة، كما هو متوقع.

4. حساب الوسيط: رتبته $F/2 = 50$. نجد أن التكرار المتجمع 68 يفوق 50 مباشرة (48)

$(68 > 50 > 60)$ ، فالفئة الوسيطة هي $[90-110]$:

$$Me = 90 + [(50 - 48) / 20] \times 20 = 92^2$$

5. حساب المنوال: أكبر تكرار هو 24 في الفئة $[70-90]$ ، فهي إذن الفئة المنوالية:

$$D1 = 24 - 16 = 8, D2 = 24 - 20 = 4$$

$$Mo = 70 + [8 / (8+4)] \times 20 \approx 83.33^2$$

ويلاحظ أن الترتيب $Mo < Me < X$ ($95.6 > 92 > 83.33$) محقق في هذا المثال، مما

يدل على أن منحنى توزيع مساحات الشقق ملتوٍ قليلاً نحو اليمين (نحو المساحات الكبيرة).

التمرين الثالث:

قطعت سيارة مسافة معينة على ثلاث مراحل متساوية الطول تماماً: قطعت المرحلة الأولى بسرعة

50 كم/سا، والمرحلة الثانية بسرعة 65 كم/سا، والمرحلة الثالثة بسرعة 75 كم/سا.

المطلوب: احسب الوسط الإحصائي المناسب لسرعة هذه السيارة خلال الرحلة كاملة، مع تعليل

اختيارك لهذا المقياس دون غيره.

الحل:

بما أن المسافات المقطوعة في كل مرحلة من مراحل الرحلة الثلاث متساوية تماماً، بينما تختلف

الأزمنة المستغرقة في قطع كل مرحلة (باعتبار أن الزمن = المسافة ÷ السرعة، وهي علاقة عكسية)،

فإن الوسط المناسب لحساب متوسط سرعة السيارة هنا هو الوسط التوافقي حصراً، وليس الوسط الحسابي

البسيط كما قد يُظن للوهلة الأولى؛ وذلك لأن الوسط التوافقي هو المقياس الوحيد الذي يأخذ بعين

الاعتبار تفاوت الأزمنة المستغرقة في كل مرحلة رغم تساوي المسافات:

$$MH = 3 / (1/50 + 1/65 + 1/75) \approx 61.58 \text{ كم/سا}$$

وهذه القيمة (61.58 كم/سا) هي متوسط السرعة الفعلي والدقيق للسيارة خلال كامل الرحلة. ونلاحظ أنها أقل من الوسط الحسابي البسيط لنفس السرعات الثلاث، والذي يساوي $3/(75+65+50) = 63.33$ كم/سا؛ والفارق بين النتيجتين، وإن بدا بسيطاً، يعكس حقيقة أن السيارة أمضت وقتاً أطول نسبياً في المرحلة الأبطأ (50 كم/سا)، مما يخفض المعدل الحقيقي للسرعة عن المتوسط الحسابي البسيط الذي يتجاهل هذا التفاوت الزمني.

التمرين الرابع:

لدينا التوزيع التالي، والذي يوضح 110 أسرة تحصلت على إعانة اجتماعية من مصلحة الشؤون الاجتماعية، مصنفة حسب عدد الأطفال المتكفل بهم في كل أسرة:

عدد الأطفال	1	2	3	4	5	6	المجموع
عدد الأسر	8	20	25	30	18	9	110

المطلوب:

1. أوجد متوسط عدد الأطفال المتكفل بهم في هذه الأسر، بطريقتين مختلفتين.
2. حدد منوال ووسيط هذا التوزيع.
3. احسب الوسط الهندسي والوسط التوافقي والوسط التربيعي لعدد الأطفال.
4. أوجد الربيع الأول والربيع الثالث لهذا التوزيع.

الحل:

1. المتوسط الحسابي:

$F \uparrow$	$f_i \times x_i$	f_i	x_i
8	8	8	1
28	40	20	2

53	75	25	3
83	120	30	4
101	90	18	5
110	54	9	6
/	387	110	المجموع

$$\bar{X} = 387 / 110 \approx 3.52 \text{ طفل}$$

وللتحقق بطريقة الوسط الفرضي $X_0 = 3$ ، نحسب أعمدة الانحرافات المرجحة $f_i \times (x_i - 3)$ بنفس أسلوب الدرس، فنحصل على مجموع يساوي 57، ومنه:

$$\bar{X} = 3 + 57/110 \approx 3.52 \text{ طفل}$$

2. الوسيط: رتبته $F/2 = 55$. نجد أن التكرار المتجمع 83 يفوق 55 مباشرة ($55 > 53$)

($83 >$)، فالوسيط $Me = 4$ أطفال.

المنوال: أكبر تكرار هو 30 عند القيمة 4، فالمنوال $Mo = 4$ أطفال أيضاً، وهو ما يعني تطابق الوسيط والمنوال في هذا المثال.

3. الوسط الهندسي والتوافقي والتربيعي: بتطبيق نفس العلاقات المشروحة في الدرس على

معطيات هذا الجدول، نحصل على:

- طفل $MG \approx 3.20$
- طفل $MH \approx 2.83$
- طفل $MQ \approx 3.77$

ويلاحظ أن العلاقة النظرية ($MH < MG < X < MQ$) محققة تماماً هنا: $3.20 > 2.83 >$

$3.52 > 3.77$ ، وهذا يؤكد صحة الحسابات.

4. الربيعيات: رتبة الربيع الأول $F/4 = 27.5$. نجد أن التكرار المتجمع 28 يفوق 27.5

مباشرة، فالربيع الأول

$$Q1 = 2. \text{أطفال}$$

رتبة الربيع الثالث $F/4 = 82.53$. نجد أن التكرار المتجمع 83 يفوق 82.5 مباشرة، فالربيع

الثالث

$$Q3 = 4. \text{أطفال}$$

أي أن ربع الأسر لديها طفلان على الأكثر، بينما ربعها الآخر لديها أربعة أطفال فأكثر.

الفصل الرابع

مقابلة
التشنت

الفصل الرابع: مقاييس التشتت

تمهيد

لا تكفي مقاييس النزعة المركزية وحدها للحكم على طبيعة توزيع تكراري أو مجموعة بيانات، إذ قد تتساوى ظاهرتان في الوسط الحسابي أو الوسيط بينما يختلف سلوك مفرداتهما اختلافاً بيّناً. من هنا تبرز الحاجة إلى أدوات إحصائية أخرى تكشف درجة تقارب أو تباعد القيم عن بعضها، وتحدد مدى اعتماديتها على مقياس النزعة المركزية في تمثيل البيانات. فإذا أخذنا مجموعتين من القيم، الأولى (4، 5، 6) والثانية (0، 5، 10)، فإن الوسط الحسابي لكليهما يساوي 5، غير أن القيمة 5 تمثل بدقة مفردات المجموعة الأولى، في حين أنها لا تعكس التفاوت الكبير الحاصل في المجموعة الثانية.

من هذا المنطلق، تُستخدم مقاييس التشتت لبيان درجة التجانس أو التباين بين المشاهدات: فكلما ارتفعت قيمة مقياس التشتت دلّ ذلك على تباعد أكبر بين القيم، والعكس صحيح. وأبرز هذه المقاييس: المدى، المدى الربيعي والانحراف الربيعي، الانحراف المتوسط، التباين والانحراف المعياري، ومعامل الاختلاف. وسنتناول في هذا الفصل كل واحد منها بالتعريف والقانون والتطبيق

أولاً: المدى (Range)

(a) المدى في حالة البيانات غير المبوبة

يُعدّ المدى أبسط مقاييس التشتت على الإطلاق، ويُحسب في حالة البيانات غير المبوبة بطرح أصغر قيمة في السلسلة الإحصائية من أكبر قيمة فيها، وفق العلاقة التالية¹:

$$R = x_{max} - x_{min}$$

¹ Karen A. Randolph , Laural. Myers , basic statistics in multivariate analysis ,pocket guides to social work research methods ,new york,2013,p29

مثال رقم (04-01): سُجِّلت درجات الحرارة العظمى (بالدرجة المئوية) خلال ثمانية أيام متتالية في مدينتين، كما يلي:

$$x : 5 , 9 , 14 , 2 , 18 , 11 , 20 , 7$$

$$y : 10 , 15 , 15 , 15 , 25 , 25 , 25 , 28$$

المطلوب: احسب مدى درجات الحرارة في كل مدينة y و x ، وعلّق على النتيجة.

الحل:

$$R_x = 20 - 2 = 18^\circ$$

$$R_y = 28 - 10 = 18^\circ$$

يلاحظ تساوي المدى في المدينتين رغم اختلاف نمط توزيع القيم بينهما، إذ تتمركز درجات حرارة المدينة (y) حول قيم متقاربة بينما تتوزع درجات المدينة (x) بشكل أكثر تشتتاً. وهذا يبيّن أن المدى قد يكون مضللاً أحياناً كونه يعتمد على القيمتين الطرفيتين فقط، وهما كثيراً ما تكونان قيمتين شاذتين، ومن هنا ظهرت الحاجة إلى مقاييس تشتت أكثر دقة.

المدى في حالة البيانات المبوبة

المدى في حالة متغير كمي منفصل:

يُحسب المدى هنا بالفرق بين أكبر قيمة للمتغير وأصغر قيمة له، وفق نفس العلاقة السابقة:

$$R = x_{max} - x_{min}$$

مثال رقم (04-02): احسب المدى للتوزيع التكراري التالي، الذي يمثل عدد الأخطاء المسجلة في عيّنة من الفواتير:

10	8	6	4	2	المتغير (عدد الأخطاء)
3	8	15	10	4	التكرارات

الحل:

$$R = 10 - 2 = 8$$

المدى في حالة متغير كمي متصل:

يُحسب المدى هنا بالفرق بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى:

$$R = x_{max} - x_{min}$$

حيث: Xmax الحد الأعلى للفئة الأخيرة، Xmin الحد الأدنى للفئة الأولى

مثال رقم (03-04): يبين الجدول التالي توزيع 50 علبة إنتاج في أحد المصانع حسب وزنها (بالكيلوغرام):

الوزن (كغ)	15-10	20-15	25-20	30-25	35-30	40-35
عدد العلب	4	10	14	12	7	3

المطلوب: حساب مدى أوزان العلب.

الحل:

$$R = 40 - 10 = 30 \text{ كلغ}$$

مزايا وعيوب المدى

يمكن تلخيص أهم ما يميز المدى وأبرز حدوده في الجدول التالي:

(1) مزايا المدى

- سهولة حسابه وبساطة فكرته
- يستخدم كثيراً في الإعلانات الجوية (الحرارة، الرطوبة، الضغط) وفي مراقبة الجودة.
- يعطي فكرة سريعة وأولية عن درجة تشتت البيانات.

(2) عيوب المدى

- يعتمد على قيمتين متطرفتين فقط، ويهمل باقي القيم.

- يتأثر بشدة بالقيم الشاذة.
- يصعب حسابه من الجداول المفتوحة أو من البيانات الوصفية.

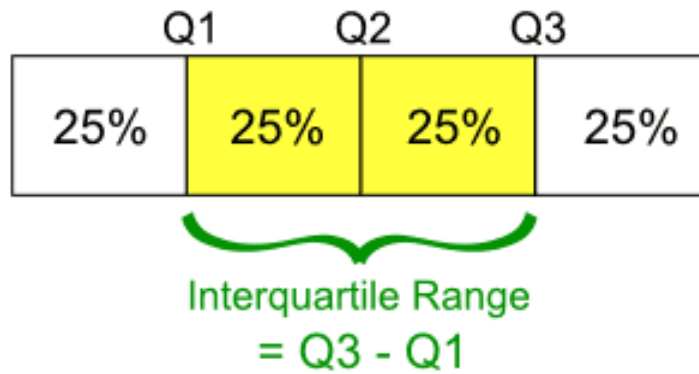
ثانياً: المدى الربيعي والانحراف الربيعي

(a) المدى الربيعي (Inter-Quartile Range)

يُعرّف المدى الربيعي بأنه الفرق بين الربع الثالث والربع الأول¹، أي:

$$RQ = Q_3 - Q_1$$

وقد سبق بيان طريقة حساب الربيعات، سواء في حالة البيانات غير المبوبة أو المبوبة (متغير كمي منفصل أو متصل)، في الفصل السابق من هذا المطبوعة.



الشكل 4 - 1: فكرة المدى الربيعي

¹ ترجمة : محمود علي أبو النصر، الاحتمالات و الإحصاء "ملخصات ايزي شوم"، الطبعة العربية الأولى، الدار الدولية للاستثمارات الثقافية، مصر، 2004، ص 27.

الانحراف الربيعي (Quartile Deviation)

لما كان المدى يعتمد على قيمتين متطرفتين فقط، فإن أي قيمة شاذة بينهما تجعل الاعتماد عليه كمقياس تشتت مضللاً. لذلك لجأ الإحصائيون إلى مقياس يعتمد على نصف القيم الوسطى ويهمل نصف القيم المتطرفة، وهو ما يجعله غير متأثر بالقيم الشاذة، ويُعرف بالانحراف الربيعي¹، ويُحسب كما يلي:

$$QD = (Q_3 - Q_1)/2$$

حيث

$$QD = QR/2$$

ويتضح من العلاقة أعلاه أن الانحراف الربيعي يساوي نصف المدى الربيعي.

مثال رقم (04-04): إذا كانت قيمتا الربع الأول والربع الثالث لتوزيع معين هما $Q_1 = 5.2$ و $Q_3 = 9.6$ ، احسب المدى الربيعي والانحراف الربيعي.

الحل:

$$RQ = 9.6 - 5.2 = 4.4$$

$$QD = 4.4 / 2 = 2.2$$

مزايا وعيوب الانحراف الربيعي

(1) مزايا الانحراف الربيعي

- لا يتأثر بالقيم الشاذة.
- يمكن حسابه من الجداول المفتوحة.

(2) عيوب الانحراف الربيعي

¹ هدى برهان سيف الدين، الإحصاء في علم النفس-مقاييس التشتت-، مطبوعة جامعية، جامعة الملك عبد العزيز، السعودية، 2014، ص50.

- لا يأخذ جميع قيم التوزيع في الحسبان.
- يصعب التعامل معه جبرياً، ولا يمكن حسابه من البيانات الوصفية.

ثالثاً: الانحراف المتوسط (Mean Deviation)

يتجاوز الانحراف المتوسط بعض القصور الملازم للمدى والانحراف الربيعي، إذ يعتمد على جميع قيم التوزيع دون استثناء، ويقوم على حساب متوسط الانحرافات المطلقة لقيم المتغير عن أحد مقاييس النزعة المركزية، وغالباً ما يُعتمد الوسط الحسابي أساساً لذلك.

(a) الانحراف المتوسط في حالة البيانات غير المبوبة

وهو متوسط الانحرافات المطلقة عن الوسط الحسابي ويتم حسابه وفق الآتي¹

$$MD = \sum [x_i - \bar{X}] / n$$

مثال رقم (04-05): جد الانحراف المتوسط للقيم التالية: 12، 15، 9، 18، 6

الحل: يجب أولاً حساب الوسط الحسابي:

$$\bar{X} = (12+15+9+18+6) / 5 = 60 / 5 = 12$$

ثم نطبق قانون الانحراف المتوسط:

$$MD = (|12-12|+|15-12|+|9-12|+|18-12|+|6-12|) / 5 = (0+3+3+6+6)/5 = 18/5 = 3.6$$

¹ Clemens . Reimann , Peter . Filzmoser ,statistical data analysis explained ,John wiley and sons, new York,2008,p57.

الانحراف المتوسط في حالة البيانات المبوبة

الانحراف المتوسط لمتغير كمي منفصل:

يُحسب الانحراف المتوسط للقيم $x_1, x_2, \dots, x_n$ المرجحة بتكراراتها $f_1, f_2, \dots, f_n$ وفق العلاقة:

$$MD = \sum f_i [x_i - \bar{X}] / \sum f_i$$

مثال رقم (04-06): يمثل الجدول التالي عدد الزبائن الذين استقبلهم 20 موظف استقبال يومياً في أحد البنوك، والمطلوب حساب الانحراف المتوسط لهذا التوزيع:

عدد الزبائن	10	15	20	25	المجموع
عدد الموظفين	4	6	8	2	20

الحل:

f_i	x_i	$f_i \cdot x_i$	$ x_i - \bar{X} $	$f_i x_i - \bar{X} $
4	10	40	7	28
6	15	90	2	12
8	20	160	3	24
2	25	50	8	16
20	المجموع	340	/	80

$$\bar{X} = 340 / 20 = 17$$

$$MD = 80 / 20 = 4$$

الانحراف المتوسط لمتغير كمي متصل:

يُحسب الانحراف المتوسط لمراكز الفئات $x_1, x_2, \dots, x_n$ المرجحة بتكراراتها المناظرة وفق نفس العلاقة السابقة¹:

$$MD = \sum f_i [x_i - \bar{X}] / \sum f_i$$

مثال رقم (04-07): يمثل الجدول التالي توزيع الإنتاج اليومي (بالوحدة) لدى 60 عاملاً في مصنع للأحذية، والمطلوب حساب الانحراف المتوسط لهذا التوزيع:

الإنتاج	120-100	140-120	160-140	180-160	200-180	220-200	240-220
عدد العمال	5	8	12	15	10	6	4

الحل:

الفئات	f_i	x_i	$f_i \cdot x_i$	$ x_i - \bar{X} $	$f_i x_i - \bar{X} $
100-120	5	110	550	57	285
120-140	8	130	1040	37	296
140-160	12	150	1800	17	204
160-180	15	170	2550	3	45
180-200	10	190	1900	23	230
200-220	6	210	1260	43	258
220-240	4	230	920	63	252
المجموع	60	/	10020	/	1570

$$\bar{X} = 10020 / 60 = 167$$

$$MD = 1570 / 60 \approx 26.17$$

¹ Isabel. Willemse , statistical methods and calculation skills, third edition, Juta ,new York,2009,p76.

مزايا وعيوب الانحراف المتوسط

(1) مزايا الانحراف المتوسط

- يأخذ جميع قيم التوزيع في الحسبان
- أدق من المدى والانحراف الربيعي كونه يعتمد على متوسط الانحرافات المطلقة عن الوسط الحسابي

(2) عيوب الانحراف الربيعي

- يتأثر بالقيم الشاذة
- يصعب التعامل معه جبرياً، ولا يمكن حسابه من البيانات الوصفية

رابعاً: التباين والانحراف المعياري (Variance and Standard Deviation)

يُعدّان من أهم مقاييس التشتت وأكثرها استخداماً، ويمثل الانحراف المعياري الجذر التربيعي للتباين. ولما كان التباين يُقاس بوحدات مربعة لا تتفق مع وحدة قياس المتغير الأصلي، كونه يعتمد على مجموع مربعات الانحرافات، فقد لجأ الإحصائيون إلى أخذ جذره التربيعي للحصول على مقياس يقاس بنفس وحدة المتغير محل الدراسة، وهو الانحراف المعياري.

(a) التباين والانحراف المعياري في حالة البيانات غير المبوبة

يُعرّف التباين بأنه متوسط مربعات انحرافات القيم عن وسطها الحسابي، ويُحسب بإحدى العلاقتين التاليتين:

(1) علاقة التعريف

$$V(x) = \sum (x_i - \bar{X})^2 / n$$

(2) العلاقة الموسعة

$$V(x) = \left(\sum x_i^2 / n \right) - \bar{X}^2$$

ويُحسب الانحراف المعياري بأخذ الجذر التربيعي للتباين:

$$S = \sqrt{V(x)}$$

ملاحظة: التباين ليس له وحدة قياس

مثال رقم (04-08): سُجِّلَت أزمان توصيل 8 طلبيات عبر تطبيق للتوصيل السريع (بالدقائق) كما يلي: 22، 15، 30، 18، 25، 12، 28، 20. احسب التباين والانحراف المعياري بطريقتين مختلفتين.

الحل: يجب أولاً حساب الوسط الحسابي:

$$\bar{X} = (22+15+30+18+25+12+28+20) / 8 = 170 / 8 = 21.25 \text{ دقيقة}$$

❖ بطريقة التعريف:

مجموع مربعات الانحرافات =

$$= 1.5625 + 45.5625 + 85.5625 + 14.0625 + 10.5625 + 76.5625 + 39.0625 + 0.5625 = 273.5$$

$$V(X) = 273.5 / 8 = 34.19$$

❖ بالطريقة الموسعة:

$$\sum x_i^2 = 3886 \text{ مجموع مربعات القيم}$$

$$V(X) = 3886/8 - (21.25)^2 = 485.75 - 451.56 = 34.19$$

ومنه: الانحراف المعياري يساوي

$$S = \sqrt{34.19} \approx 5.85 \text{ دقيقة}$$

التباين والانحراف المعياري في حالة البيانات المبوبة

في حالة متغير كمي منفصل: توجد طريقتان لحساب التباين والانحراف المعياري:

(1) الطريقة الموسعة

$$V(X) = (\sum f_i x_i^2 / \sum f_i) - \bar{X}^2$$

(2) طريقة التعريف

$$V(X) = \sum f_i (x_i - \bar{X})^2 / \sum f_i$$

ويُحسب الانحراف المعياري بأخذ الجذر التربيعي للتباين:

$$S = \sqrt{V(x)}$$

علما أن: x_i هي قيم المتغير

مثال رقم (04-09): احسب التباين والانحراف المعياري بطريقتين مختلفتين لبيانات المثال رقم (04-06).

الحل:

$f_i(x_i - \bar{X})^2$	$(x_i - \bar{X})^2$	$f_i \cdot x_i^2$	$f_i \cdot x_i$	f_i	x_i
196	49	400	40	4	10
24	4	1350	90	6	15
72	9	3200	160	8	20
128	64	1250	50	2	25
420	/	6200	340	20	المجموع

$$\bar{X} = \frac{\sum f_i \cdot x_i}{\sum f_i} = 340/20 = 17$$

- طريقة التعريف: $V(X) = 420/20 = 21$

- الطريقة الموسعة: $V(X) = 6200/20 - 17^2 = 310 - 289 = 21$

$$S = \sqrt{21} \approx 4.58$$

في حالة متغير كمي متصل: تُطبَّق نفس العلاقتين السابقتين على مراكز الفئات:

(1) الطريقة الموسعة

$$V(X) = (\sum f_i x_i^2 / \sum f_i) - \bar{X}^2$$

(2) طريقة التعريف

$$V(X) = \sum f_i (x_i - \bar{X})^2 / \sum f_i$$

علما أن: x_i هي مراكز الفئات

مثال رقم (04-10): احسب التباين والانحراف المعياري بطريقتين مختلفتين لبيانات المثال رقم (04-07).

الحل:

الفئات	f_i	x_i	$f_i \cdot x_i$	$f_i \cdot x_i^2$	$(x_i - \bar{X})^2$	$f_i(x_i - \bar{X})^2$
100-120	5	110	550	60500	3249	16245
120-140	8	130	1040	135200	1369	10952
140-160	12	150	1800	270000	289	3468
160-180	15	170	2550	433500	9	135
180-200	10	190	1900	361000	529	5290
200-220	6	210	1260	264600	1849	11094
220-240	4	230	920	211600	3969	15876
المجموع	60	/	10020	1736400	/	63060

$$\bar{X} = 10020/60 = 167$$

- طريقة التعريف: $V(X) = 63060/60 = 1051$

- الطريقة الموسعة: $V(X) = 1736400/60 - 167^2 = 28940 - 27889 = 1051$

$$S = \sqrt{1051} \approx 32.42 \text{ وحدة}$$

خصائص الانحراف المعياري

يتميز الانحراف المعياري بجملة من الخصائص، أبرزها:

- ❖ قيمة الانحراف المعياري موجبة دوماً أو تساوي الصفر، وتبلغ الصفر فقط عندما تتساوى جميع قيم التوزيع، إذ لا يوجد عندئذ أي تباين بين المفردات.
- ❖ كلما زاد تشتت القيم حول الوسط الحسابي زادت قيمة الانحراف المعياري، والعكس صحيح.
- ❖ إذا أضيف أو طُرح مقدار ثابت من جميع قيم المتغير، فإن الانحراف المعياري والتباين لا يتغيران.
- ❖ إذا ضربت كل قيمة من قيم المتغير x في مقدار ثابت a لتصبح القيم الجديدة $y = ax$ ، فإن انحرافها المعياري الجديد يساوي: $S_y = a \cdot S_x$ ، ويصبح التباين الجديد: $V(y) = a^2 \cdot V(x)$.
- ❖ إذا كانت لدينا التوليفة الخطية $y = ax + b$ ، فإن الانحراف المعياري للمتغير y يبقى: $S_y = a \cdot S_x$ ، ويصبح التباين: $V(y) = a^2 \cdot V(x)$.

مزايا وعيوب الانحراف المعياري

(1) مزايا الانحراف المعياري

- يأخذ جميع قيم التوزيع في الحسبان.
- يسهل التعامل معه جبرياً.
- أشهر مقاييس التشتت وأكثرها استخداماً.

(2) عيوب الانحراف المعياري

- يتأثر بالقيم الشاذة.
- يصعب حسابه من الجداول المفتوحة.
- لا يمكن حسابه من البيانات الوصفية.

خامساً: معامل الاختلاف ومعامل الاختلاف الربيعي

(a) معامل الاختلاف (Coefficient of Variation)

تتم مقارنة تشتت مجموعتين مباشرة من خلال مقارنة قيم مقاييس التشتت المحسوبة لكل منهما، شريطة أن تشترك المجموعتان في نفس وحدة القياس؛ فيمكن على سبيل المثال مقارنة تشتت علامات فوج من الطلبة بتشتت علامات فوج آخر. أما إذا اختلفت وحدة القياس بين المجموعتين، كأن تكون إحداها مقيسة بالأيام والأخرى بالوحدات النقدية، فلا يصح عندئذ مقارنة مقياس التشتت لإحداها بمقياس التشتت للأخرى مباشرة.

وفي مثل هذه الحالة، يُلجأ إلى معامل الاختلاف الذي يتيح الكشف عن المجموعة الأكثر تشتتاً من بين مجموعات تختلف وحدات قياسها، إذ تُعتبر المجموعة ذات معامل الاختلاف الأكبر هي الأكثر تشتتاً. ويُحسب معامل الاختلاف اعتماداً على الوسط الحسابي والانحراف المعياري كما يلي:

$$CV = \left(\frac{S}{\bar{X}} \right) * 100$$

ويُلاحظ أن معامل الاختلاف قيمة بلا وحدة قياس، إذ يُعبّر عنه كنسبة مئوية.

مثال رقم (11-04): ترغب إدارة أحد المستشفيات في معرفة أي العاملين أكثر تبايناً من حيث التكلفة: مدة انتظار المرضى (بالدقائق) أم تكلفة فحص طبي معيّن (بالوحدة النقدية)، وذلك بهدف توجيه جهود الترشيح نحو العامل الأكثر تشتتاً. وبعد دراسة عيّنة من الحالات السابقة، تم التوصل إلى النتائج التالية:

$$\text{دقيقة } S = 2.5, \bar{X} = 12.5 \text{ مدة الانتظار}$$

$$\text{وحدة نقدية } S = 85, \bar{Y} = 850 \text{ تكلفة الفحص}$$

الحل: بما أن وحدتي القياس مختلفتان، نحسب معامل الاختلاف لكل متغير:

$$CV_x = (2.5 / 12.5) \times 100 = 20\%$$

$$CV_y = (85 / 850) \times 100 = 10\%$$

يتضح من النتيجة أن مدة انتظار المرضى هي المتغير الأكثر تشتتاً، مما يستدعي توجيه الاهتمام نحو ضبط هذه المدة وتقليل التفاوت فيها، باعتبارها مصدر التذبذب الأكبر في أداء الخدمة.

معامل الاختلاف الربيعي (Quartile Coefficient of Variation)

هو نسبة مئوية للفرق بين الربع الثالث والربع الأول إلى مجموعهما، ويُلبأ إليه في الحالات التي يتعذر فيها حساب الوسط الحسابي والانحراف المعياري، كما هو الشأن في الجداول المفتوحة من طرف واحد أو من الطرفين، ويُحسب وفق العلاقة التالية:

$$QCV = [(Q_3 - Q_1) / (Q_3 + Q_1)] * 100$$

مثال رقم (04-12): إذا كانت لدينا المعطيات التالية: $Q_1 = 6$ و $Q_3 = 14$ ، احسب معامل الاختلاف الربيعي.

الحل:

$$QCV = [(14 - 6) / (14 + 6)] \times 100 = (8/20) \times 100 = 40\%$$

ملاحظة:

- إذا كان معامل الاختلاف ومعامل الاختلاف الربيعي أكبر من 50% نقول إن التشتت قوي.
- إذا كان معامل الاختلاف ومعامل الاختلاف الربيعي أصغر من 50% نقول إن التشتت ضعيف.
- إذا كان معامل الاختلاف ومعامل الاختلاف الربيعي يساوي 50% نقول إن التشتت متوسط.

تمارين الفصل الرابع

التمرين الأول:

يبيّن الجدول التالي توزيع 50 مزرعة حسب إنتاجيتها من القمح (قنطار/هكتار):

الإنتاجية	20-10	30-20	40-30	50-40	60-50	70-60	المجموع
عدد المزارع	4	8	15	12	7	4	50

المطلوب: احسب المدى، الانحراف الربيعي، التباين بطريقتين، الانحراف المعياري، والانحراف المتوسط.

الحل:

الفئات	f_i	x_i	$f_i \cdot x_i$	$F \uparrow$	$f_i \cdot x_i^2$	$ x_i - \bar{X} $	$f_i x_i - \bar{X} $
20-10	4	15	60	4	900	24.4	97.6
30-20	8	25	200	12	5000	14.4	115.2
40-30	15	35	525	27	18375	4.4	66
50-40	12	45	540	39	24300	5.6	67.2
60-50	7	55	385	46	21175	15.6	109.2
70-60	4	65	260	50	16900	25.6	102.4
المجموع	50	/	1970	/	86650	/	557.6

1. المدى:

$$R = 70 - 10 = 60 \text{ قنطار/هكتار}$$

2. الانحراف الربيعي:

أولاً نستخرج قيم التكرار المتجمع الصاعد (مبينة في العمود $F \uparrow$ أعلاه).

الربع الأول Q1	الربع الثالث Q3
رتبة $Q1 = f/4 = 12.5$ ، فئة الربع الأول: $Q1 = 30 + 40 - 30$ $[(12.5 - 12)/15] \times 10 = 30.33$	رتبة $Q3 = 3f/4 = 37.5$ ، فئة الربع الثالث: $Q3 = 40 + 50 - 40$ $[(37.5 - 27)/12] \times 10 = 48.75$

$$QD = (48.75 - 30.33) / 2 = 9.21$$

3. التباين والانحراف المعياري بطريقتين: يجب أولاً حساب الوسط الحسابي:

$$\bar{X} = 1970 / 50 = 39.4$$

طريقة التعريف	الطريقة الموسعة
$V(X) = \Sigma f_i(x_i - \bar{X})^2 / \Sigma f_i = 9032 / 50 = 180.64$	$V(X) = \Sigma f_i x_i^2 / \Sigma f_i - \bar{X}^2 = 86650 / 50 - 39.4^2 = 180.64$

$$S = \sqrt{180.64} \approx 13.44$$

4. الانحراف المتوسط:

$$MD = 557.6 / 50 = 11.15$$

التمرين الثاني:

عند دراسة الرواتب الشهرية لموظفي قسمين في إحدى الشركات، حصلنا على النتائج التالية:

قسم المبيعات	قسم الإنتاج
$\bar{X} = 45000$ ، $S = 6000$	$\bar{X} = 28000$ ، $S = 5000$

المطلوب: أي القسمين أكثر تشتتاً في الرواتب؟

الحل: بما أن وحدة القياس واحدة (رواتب بنفس العملة) لكن الوسطين الحسابيين مختلفان، فإن المقارنة السليمة تتم عبر معامل الاختلاف:

$$CV\% 13.33 = 100 \times (45000/6000) = (\text{المبيعات})$$

$$CV\% 17.86 = 100 \times (28000/5000) = (\text{الإنتاج})$$

نلاحظ أن معامل الاختلاف في قسم الإنتاج أكبر منه في قسم المبيعات، ومنه فإن رواتب قسم الإنتاج هي الأكثر تشتتاً.

التمرين الثالث:

يبين الجدول التالي توزيع 40 مريضاً حسب عدد أيام الإقامة في المستشفى للتعافي من عملية جراحية معينة:

المجموع	30-25	25-20	20-15	15-10	أقل من 10	مدة الإقامة (أيام)
40	6	10	12	8	4	عدد المرضى

المطلوب:

1. ما مقياس النزعة المركزية المناسب لهذا الجدول؟ احسبه.
2. ما مقياس التشتت المناسب؟ احسبه.
3. احسب معامل الاختلاف المناسب، وحدد درجة التشتت.

الحل:

1. بما أن الجدول مفتوح من الطرف الأول (أقل من 10)، فإن مقياس النزعة المركزية المناسب هو الوسيط. نستخرج أولاً التكرار المتجمع الصاعد:

30-25	25-20	20-15	15-10	أقل من 10	مدة الإقامة
6	10	12	8	4	عدد المرضى
40	34	24	12	4	F↑

رتبة الوسيط: $f/2 = 20$ ، الفئة الوسيطة: 20-15

$$Me = 15 + [(20 - 12) / 12] \times 5 = 18.33 \text{ يوماً}$$

2. بما أن الجدول مفتوح، فإن مقياس التشتت المناسب هو الانحراف الربيعي، ويستلزم حساب Q1 و3Q:

الربيع الأول Q1	الربيع الثالث Q3
رتبة $Q1 = f/4 = 10$ ، فئة الربيع الأول: -10 15	رتبة $Q3 = 3f/4 = 30$ ، فئة الربيع الثالث: 25-20
$Q1 = 10 + [(10-4)/8] \times 5 = 13.75$	$Q3 = 20 + [(30-24)/10] \times 5 = 23$

$$QD = (23 - 13.75) / 2 = 4.625 \text{ يوماً}$$

3. نظراً لتعذر حساب معامل الاختلاف العادي في الجداول المفتوحة، نحسب معامل الاختلاف الربيعي:

$$QCV = [(23 - 13.75) / (23 + 13.75)] \times 100 = (9.25/36.75) \times 100 \approx 25.17 \%$$

بما أن معامل الاختلاف الربيعي أقل من 50%، فإن درجة التشتت تُعدّ ضعيفة، أي أن مدد الإقامة متقاربة نسبياً حول الوسيط.

الفصل الخامس

مقاييس الشك

الفصل الخامس: مقاييس الشكل

تمهيد

لا تكفي مقاييس النزعة المركزية ومقاييس التشتت وحدهما لوصف البيانات الإحصائية وإجراء المقارنات بين التوزيعات التكرارية المختلفة؛ فقد يتساوى توزيعان في وسطهما الحسابي وفي انحرافهما المعياري، ومع ذلك نجد أنهما مختلفين تماماً من حيث الشكل، خاصة فيما يخص طريقة انتشار البيانات على المنحنى البياني الممثل لهما من حيث التواءه وتفرطحه عن الوضع الطبيعي (الاعتدالي). لهذا السبب بالذات، دعت الحاجة إلى استخدام مقاييس إضافية لتحقيق هذا الغرض بالذات، وهي ما اصطلح على تسميته بمقاييس الشكل.

ولتوضيح أهمية هذه المقاييس بمثال بسيط، لنختل توزيعين لرواتب موظفين في مؤسستين مختلفتين، لهما نفس الوسط الحسابي (50000 دج) ونفس الانحراف المعياري (10000 دج)؛ ومع ذلك فإن التوزيع الأول قد يكون متماثلاً (أغلب الرواتب قريبة من المتوسط بشكل متوازن)، بينما قد يكون التوزيع الثاني ملتوياً بشدة نحو اليمين (أغلب الموظفين برواتب منخفضة نسبياً، مع عدد قليل جداً من الإطارات العليا برواتب مرتفعة جداً ترفع المتوسط). فالوسط الحسابي والانحراف المعياري وحدهما لا يكشفان هذا الفارق الجوهرى بين التوزيعين، بينما تكشفه مقاييس الشكل بوضوح.

وتنقسم مقاييس الشكل أساساً إلى صنفين: مقاييس الالتواء التي تصف مدى ابتعاد التوزيع عن التماثل، ومقاييس التفرطح التي تصف مدى تفرطح أو تدبب قمة المنحنى الممثل للتوزيع. ولدراسة كل من الالتواء والتفرطح، نحتاج بداية إلى مفهوم رياضي أساسي يُعرف بـ"العزوم"، ولذا سنستهل هذا الفصل بشرح العزوم بنوعها قبل الانتقال إلى مقاييس الالتواء ثم مقاييس التفرطح.

أولاً: العزوم (Moments)

العزم مفهوم مأخوذ أصلاً من الفيزياء، حيث يُستعمل لتحديد مدى أثر قوة معينة في جعل جسم ما يدور حول محور ثابت. أما في الإحصاء، فتُستعمل العزوم كمؤشرات شاملة تساعد على تحديد طبيعة توزيع البيانات وتغيراتها، ويمكن تعريف العزم عن أي نقطة أو قيمة مرجعية؛ فإذا كانت هذه القيمة المرجعية تساوي صفراً، سُمي العزم بالعزم البسيط، أما إذا كانت هذه القيمة المرجعية أحد مقاييس النزعة المركزية، كالوسط الحسابي، سُمي عندئذٍ بالعزم المركزي.

العزوم البسيطة (Simple Moments)

يُعرّف العزم البسيط من الدرجة (K) للمتغير (x_i) ، الذي يأخذ القيم $(x_1, x_2, x_3, \dots, x_n)$ ، ويرمز له بالرمز (m_k) ، بأنه متوسط قيم المتغير مرفوعة إلى الأس (K). وتختلف طريقة حساب العزم البسيط باختلاف طبيعة البيانات، سواء كانت بيانات مبوبة أو غير مبوبة، وذلك وفق العلاقتين التاليتين¹:

(a) البيانات المبوبة:

$$m_k = \sum (n_i * x_i^k) / \sum n_i$$

(b) البيانات غير المبوبة :

$$m_k = \sum x_i^k / n$$

حيث تمثل (x_i) قيم المتغير في حالة المتغير الكمي المنقطع، بينما تمثل مراكز الفئات في حالة المتغير الكمي المستمر.

¹ سامية تيلولت، مبادئ في الإحصاء، الطبعة الثانية، دار الحديث للكتاب، الجزائر، 2009، ص 126.

مثال تمهيدي: قبل الانتقال إلى مثال شامل بالبيانات المبوبة، نوضح فكرة العزم البسيط بمثال بسيط على بيانات غير مبوبة. لتكن لدينا القيم التالية: 6 ، 9 ، 14 ، 7 ، 11 ، 5 ، 12. المطلوب: أوجد العزم البسيط الأول والثاني والثالث

$$m_1 = (6+9+14+7+11+5+12) / 7 = 64 / 7 \approx 9.14$$

$$m_2 = (6^2+9^2+14^2+7^2+11^2+5^2+12^2) / 7 = 652 / 7 \approx 93.14$$

$$m_3 = (6^3+9^3+14^3+7^3+11^3+5^3+12^3) / 7 = 7216 / 7 \approx 1030.86$$

ونلاحظ أن العزم البسيط الأول (m_1) هو نفسه الوسط الحسابي لهذه القيم مباشرة (9.14)، وهي خاصية عامة سنعتمد عليها الآن في المثال الشامل التالي بالبيانات المبوبة.

مثال رقم (05-01): يوضح الجدول التالي توزيع 100 عامل حسب أجرهم اليومي، معبراً عنه بمئات الدينارات الجزائرية.

الفئات	15-10	20-15	25-20	30-25	35-30	المجموع
عدد العمال	22	30	25	15	8	100

المطلوب: أوجد العزم البسيط من الدرجة الأولى والعزم البسيط من الدرجة الثالثة للتوزيع الإحصائي الآتي.

الحل:

الفئات	f_i	x_i	$f_i \times x_i$	$f_i \times x_i^3$
15-10	22	12.5	275	42968.75
20-15	30	17.5	525	160781.25

284765.625	562.5	22.5	25	25-20
312304.688	412.5	27.5	15	30-25
274300	260	32.5	8	35-30
1075120.313	2035	/	100	المجموع
$f_i \times x_i^3$	$f_i \times x_i$	x_i	f_i	الفئات

$$m_1 = 2035 / 100 = 20.35$$

$$m_3 = 1075120.313 / 100 \approx 10751.20$$

وتجدر الإشارة إلى أن العزم البسيط من الدرجة الأولى (m_1) يساوي الوسط الحسابي للتوزيع، وهي خاصية عامة سنعتمد عليها لاحقاً في حساب مقاييس الالتواء والتفرطح.

العزوم المركزية

تُعدّ العزوم المركزية مقاييس شبيهة بالتباين من حيث طريقة حسابها، وتُستخدم غالباً لحساب مقاييس الالتواء وتحديد اتجاه التواء توزيع البيانات، كما تُستخدم في حساب مقاييس التفرطح وتحديد شكل التوزيع. وتُحسب العزوم المركزية لدرجات (مراتب) مختلفة.

ويُعرّف العزم المركزي من الدرجة (K) للمتغير (x_i)، الذي يأخذ القيم ($x_1, x_2, \dots, x_n$)، ويرمز له بالرمز (μ_k)، بأنه متوسط انحرافات قيم المتغير عن وسطها الحسابي، مرفوعة إلى الأس (K). وتختلف طريقة حسابه باختلاف طبيعة البيانات، سواء كانت بيانات مبوبة أو غير مبوبة، وذلك وفق العلاقتين التاليتين:

تُعدّ العزوم المركزية مقاييس شبيهة بالتباين من حيث طريقة حسابها، وتُستخدم غالباً لحساب مقاييس الالتواء وتحديد اتجاه التواء توزيع البيانات، كما تُستخدم في حساب مقاييس التفرطح وتحديد شكل التوزيع.

وتُحسب العزوم المركزية من رتب (درجات) مختلفة، ويُعرّف العزم المركزي للمتغير (x_i) ، الذي يأخذ القيم $(x_1, x_2, \dots, x_n)$ ، ويُرمز له بالرمز (μ_k) ، بأنه متوسط انحرافات القيم عن وسطها الحسابي، مرفوعةً إلى الأس (K) . وتختلف طريقة حسابه باختلاف طبيعة البيانات، وذلك وفق العلاقتين التاليتين:

(a) البيانات المبوبة :

$$\mu_k = \sum n_i (x_i - \bar{X})^k$$

(b) البيانات غير المبوبة :

$$\mu_k = \sum (x_i - \bar{X})^k / n$$

ملاحظة: تتراوح مراتب العزوم المركزية نظرياً من 0 إلى ما لا نهاية، إلا أن العزوم الأربعة الأولى فقط هي الشائعة الاستعمال في الإحصاء التطبيقي. ومن أهم الحالات الخاصة لهذه العزوم:

$$\mu_0 = 1 \text{ (دائماً)}$$

$$\mu_1 = 0 \text{ (دائماً، بما أن مجموع انحرافات القيم عن وسطها الحسابي معدوم)}$$

$$\mu_2 = V(X) \text{ (التباين)}$$

أي أن العزم المركزي الثاني ليس سوى التباين الذي سبقت دراسته في الفصل الرابع، وهذا ما يفسر التشابه الكبير بين طريقة حساب العزوم المركزية وطريقة حساب التباين.

مثال رقم (05-02): أوجد العزم المركزي الثالث والرابع من التوزيع التكراري لبيانات المثال السابق رقم (05-01).

الحل: الوسط الحسابي هو نفسه العزم البسيط الأول الذي سبق حسابه في المثال (05-01) :

$$m_1 = \bar{X} = 20.35$$

الفئات	f_i	x_i	$(x_i - \bar{X})^3$	$f_i \times (x_i - \bar{X})^3$	$(x_i - \bar{X})^4$	$f_i \times (x_i - \bar{X})^4$
15-10	22	12.5	-599.695	-13193.3	4667.626	102707.8
20-15	30	17.5	-22.196	-665.88	62.575	1877.25
25-20	25	22.5	9.938	248.45	21.394	534.85
30-25	15	27.5	365.526	5482.89	2004.72	30070.8
35-30	8	32.5	1793.613	14348.9	12261.72	98093.8
المجموع	100	/	/	6220.97	/	233284.5

ملاحظة حسابية: القيم المبينة أعلاه حُسبت بدقة أكبر باستخدام البرامج الحسابية. وبعد تدقيق مجموع العمود الخامس $(f_i * (x_i - \bar{X})^3)$ ، نلاحظ أن قيمة الإشارة السالبة الأولى تقترب من مجموع القيم الموجبة اللاحقة، ومن ثم تكون النتيجة النهائية كما يلي:

$$\mu_3 = \sum f_i (x_i - \bar{X})^3 / \sum f_i = \frac{8743.57}{100} = 87.44$$

$$\mu_4 = \sum f_i (x_i - \bar{X})^4 / \sum f_i = \frac{299596.62}{100} = 2995.97$$

العلاقة بين العزوم المركزية والبسيطة

قد يكون حساب العزوم المركزية مباشرة (ب طرح الوسط الحسابي من كل قيمة، ثم رفع الفارق للأس المطلوب) عملية مرهقة، خاصة عند التعامل مع بيانات كثيرة. ولتقادي ذلك، يمكن التعبير عن العزوم المركزية بدلالة العزوم البسيطة فقط (m_1, m_2, m_3, m_4) ، وهو ما يُختصر العمليات الحسابية كثيراً. وأهم هذه العلاقات المختصرة للعزوم الأربعة الأولى هي:

$$\mu_2 = m_2 - m_1^2$$

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3$$

$$\mu_4 = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4$$

ونلاحظ أن العلاقة الأولى ($\mu_2 = m_2 - m_1^2$) هي نفسها العلاقة الموسعة لحساب التباين التي سبق استعمالها في الفصل الرابع، مما يؤكد مجدداً أن العزم المركزي الثاني هو التباين بعينه.

مثال رقم (03-05): للتحقق من هذه العلاقة، لنحسب العزم المركزي الثالث لتوزيع جديد يمثل عدد ساعات التكوين التي تلقاها 24 موظفاً خلال شهر معين:

الفئات	5-0	10-5	15-10	20-15	المجموع
عدد الموظفين	5	9	7	3	24

الحل: نحسب أولاً العزوم البسيطة الثلاثة الأولى:

الفئات	f_i	x_i	$f_i \times x_i$	$f_i \times x_i^2$	$f_i \times x_i^3$
0-5	5	2.5	12.5	31.25	78.125
5-10	9	7.5	67.5	506.25	3796.875
10-15	7	12.5	87.5	1093.75	13671.875
15-20	3	17.5	52.5	918.75	16078.125
المجموع	24	/	220	2550	33625

$$m_1 = 220 / 24 \approx 9.17, m_2 = 2550 / 24 = 106.25, m_3 = 33625 / 24 \approx 1401.04$$

ثم نطبق العلاقة المختصرة مباشرة:

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3 = 1401.04 - 3 \times 9.17 \times 106.25 + 2 \times 9.17^3 \approx 19.68$$

وهذه هي نفس النتيجة التي كنا سنحصل عليها بالضبط لو طبقنا العلاقة المباشرة (بطرح الوسط الحسابي من كل مركز فئة ورفعته للأس الثالث)، مما يؤكد صحة العلاقة المختصرة وفائدتها العملية في تقليل الحسابات، خاصة عند استعمال الآلة الحاسبة.

ثانيا : الالتواء (Skewness)

يُقصد بالالتواء مدى ابتعاد توزيع البيانات عن التماثل، أي مدى تركيز مفرداته على أحد جانبي الوسط الحسابي بدرجة أكبر من الجانب الآخر. ويمكن أن يأخذ التوزيع التكراري، وفقاً لطريقة توزع مفرداته حول قيمة مركزية معينة، غالباً ما تكون الوسط الحسابي، أحد الأشكال الثلاثة التالية:

- **موجب الالتواء:** يكون توزيع البيانات موجب الالتواء عندما يمتد ذيله بصورة أكبر نحو

اليمين (Skewed to the right)، ويكون ترتيب مقاييس النزعة المركزية على النحو الآتي:

$$\bar{X} > M_e > M_o$$

حيث تمثل ($\bar{X}$) الوسط الحسابي، و(M_e) الوسيط، و(M_o) المنوال.

- **سالب الالتواء:** يكون توزيع البيانات سالب الالتواء عندما يمتد ذيله بصورة أكبر نحو اليسار

(Skewed to the left)، ويكون ترتيب مقاييس النزعة المركزية على النحو الآتي:

$$\bar{X} < M_e < M_o$$

حيث تمثل (M_o) المنوال، و(M_e) الوسيط، و($\bar{X}$) الوسط الحسابي.

- **متماثل (معتدل الالتواء):** يكون التوزيع متماثلاً أو معتدلاً الالتواء إذا تساوت مقاييس النزعة

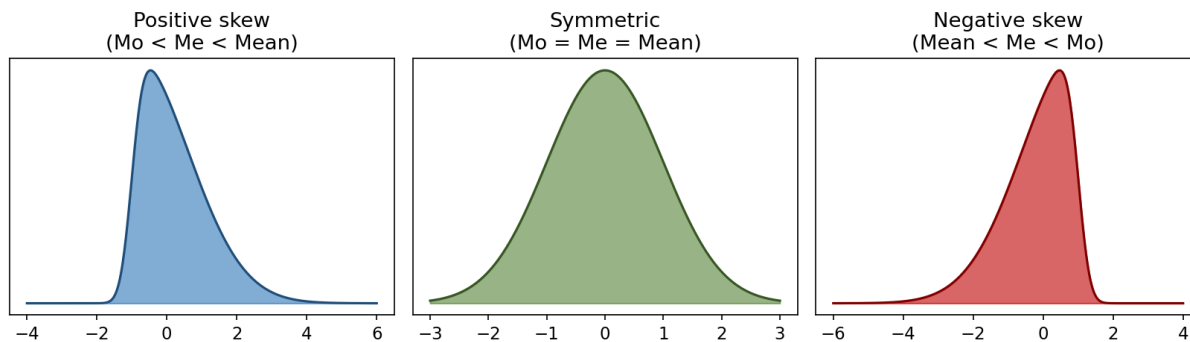
المركزية الثلاثة، أي:

$$\bar{X} = M_e = M_o$$

وفي هذه الحالة، تقع 50% من القيم على يمين الوسط الحسابي و50% منها على يساره. ويتميز هذا التوزيع بكون منحناه متماثلاً حول محور يمر بقيمته المركزية، بحيث ينطبق نصفاه تماماً عند طيه حول هذا المحور. ويُعدّ هذا الشكل توزيعاً معيارياً تُقاس بالنسبة إليه بقية أشكال التوزيعات.

ويوضح الشكل التالي الحالات الثلاث للالتواء، والمتمثلة في: التوزيع موجب الالتواء، والتوزيع سالب

الالتواء، والتوزيع المتماثل (معتدل الالتواء).



الشكل 5- 1: الأشكال الثلاثة لالتواء (موجب، متماثل، سالب)

ملاحظة: لا يكون شكل المنحنى التكراري واضحاً دائماً في الحالات العملية، مما يصعب الحكم بالعين المجردة على نوع التوزيع (متماثل أم ملتو، وإلى أي جهة). لهذا السبب، لا بد من الاعتماد على مقاييس علمية حسابية دقيقة للحكم على شكل المنحنى التكراري، وهو ما نستعرضه فيما يلي.

يمكن قياس الالتواء باستخدام عدة طرق، من أهمها: معامل بيرسون، ومعامل يول وكيندال (الالتواء الربيعي)، ومعامل فيشر. وفيما يلي نعرض هذه الطرق الثلاث تباعاً.

(a) معامل بيرسون لالتواء (Pearson's Coefficient of Skewness)

يُستخدم معامل بيرسون لالتواء عندما يكون التوزيع قريباً من التماثل وغير شديد الالتواء، ويمكن حسابه باستخدام إحدى المعادلتين التاليتين:

(a) المعادلة الأولى (بالمنوال)

$$SK_p = (\bar{X} - M_o) / \delta_x$$

(b) المعادلة الثانية (بالوسيط)

$$SK_p = 3(\bar{X} - M_e) / \delta_x$$

حيث تمثل $(\bar{X})$ الوسط الحسابي، و (M_o) المنوال، و (M_e) الوسيط، و δ_x الانحراف المعياري.

ويتراوح معامل بيرسون للالتواء بين (-1) و $(+1)$. فإذا كانت قيمته موجبة، دلّ ذلك على أن التوزيع يمتد نحو اليمين، وإذا كانت سالبة، دلّ ذلك على أن التوزيع يمتد نحو اليسار، أما إذا كانت مساوية للصفر، فهذا يعني أن التوزيع متماثل.

وتُستخدم القسمة على الانحراف المعياري لإزالة تأثير وحدات القياس، مما يسمح بمقارنة درجة الالتواء بين توزيعات أو ظواهر مختلفة.

مثال رقم (05-04): أوجد معامل الالتواء لبيرسون للتوزيع التكراري لبيانات المثال رقم (05-01)، باستخدام المعادلتين معًا.

الحل: نحسب أولاً الانحراف المعياري بالطريقة المعتادة، وذلك بالاعتماد على مجموع $(n_i * x_i^2)$ ، الذي يساوي: (45412.5)

$$\delta x = \sqrt{(45412.5/100 - 20.35^2)} = \sqrt{(454.125 - 414.1225)} = \sqrt{40.0025} \approx 6.32$$

❖ باستخدام المنوال

الفئة المنوالية هي $[15-20]$ ، لكونها الفئة ذات أكبر تكرار، حيث يبلغ تكرارها (30). ومن ثم، نطبق قانون بيرسون لحساب المنوال:

$$Mo = 15 + [8 / (8+5)] \times 5 \approx 18.08$$

$$SKp1 = (20.35 - 18.08) / 6.32 \approx 0.36$$

بما أن معامل الالتواء لبيرسون المحسوب باستخدام المنوال موجب، فإن التوزيع يكون ملتويًا نحو اليمين.

❖ باستخدام الوسيط

رتبة الوسيط هي:

$$\frac{N}{2} = \frac{100}{2} = 50$$

وبما أن التكرار المتجمع للفئة [15–20] يبلغ 52، وهو أول تكرار متجمع يتجاوز رتبة الوسيط 50، فإن الفئة [15–20] تمثل الفئة الوسيطة.

$$Me = 15 + [(50 - 22) / 30] \times 5 \approx 19.67$$

$$SKp2 = 3 \times (20.35 - 19.67) / 6.32 \approx 0.32$$

وتؤكد المعادلة الثانية النتيجة نفسها، إذ إن التوزيع ملتوٍ نحو اليمين. ويُلاحظ أن قيمة الالتواء تختلف قليلاً بين المعادلتين (0.36 مقابل 0.32)، وهو أمر طبيعي راجع لاختلاف خصائص المنوال والوسيط، إلا أن الاتجاه (الإشارة) العام يبقى موحدًا في الحالتين، وهو ما يهمنا فعلياً في تفسير شكل التوزيع.

مثال رقم (05-05): للتدريب على حالة الالتواء السالب (نحو اليسار)، لنأخذ توزيعاً جديداً يمثل أعمار 80 مشتركاً في نادٍ رياضي:

الفئات	25-20	30-25	35-30	40-35	45-40	50-45	المجموع
عدد المشتركين	4	8	14	22	20	12	80

المطلوب: احسب معامل الالتواء بيرسون بالمعادلتين معاً.

الحل: الوسط الحسابي لهذا التوزيع $\bar{X} = 37.63$ سنة، والانحراف المعياري $\delta x \approx 6.85$ سنة (بنفس أسلوب الحساب المعتاد).

أولاً: باستخدام المنوال. الفئة المنوالية هي [35–40] (أكبر تكرار 22):

$$Mo = 35 + [8 / (8+2)] \times 5 = 39$$

$$SKp1 = (37.63 - 39) / 6.85 \approx -0.20$$

ثانياً: باستخدام الوسيط. رتبة الوسيط $F/2 = 40$ ، والفئة الوسيطة [35-40] (التكرار المتجمع 48 يفوق 40 مباشرة):

$$Me = 35 + [(40 - 26) / 22] \times 5 \approx 38.18$$

$$SKp2 = 3 \times (37.63 - 38.18) / 6.85 \approx -0.24$$

بما أن قيمتي معامل بيرسون سالتان بالمعادلتين معاً، فإن هذا التوزيع ملتوٍ نحو اليسار: أغلب المشتركين في الفئات العمرية الأكبر نسبياً، مع أقلية من المشتركين الأصغر سناً تسحب الوسط الحسابي نحو الأسفل

معامل الالتواء يول وكيندال (معامل الالتواء الربيعي) (Yule & Kendall Coefficient)

يُفضَّل حساب هذا المعامل في الجداول المفتوحة من الطرفين أو من طرف واحد، كما يُفضَّل في حالة وجود قيم متطرفة ضمن البيانات غير المبوبة، حيث يأخذ بعين الاعتبار العلاقة بين الربع الأول والربع الثالث والوسيط معاً، وفق العلاقة التالية:

$$SK_q = (Q_3 + Q_1 - 2M_e) / (Q_3 - Q_1)$$

ويتراوح مجال تغيره بين -1 و $+1$ ؛ فإذا كانت قيمته سالبة فالالتواء نحو اليسار، وإذا كانت موجبة فالالتواء نحو اليمين، أما إذا كانت معدومة فالتوزيع متماثل.

مثال رقم (05-06): احسب معامل الالتواء الربيعي لمعطيات المثال السابق رقم (05-01).

الحل: سبق حساب قيمة الوسيط في المثال (05-03) ($Me = 19.67$)؛ نحسب الآن الربع الأول والربع الثالث:

الربع الأول Q1	الربع الثالث Q3
رتبة $Q1 = F/4 = 25$ ، فئة الربع الأول	رتبة $Q3 = 3F/4 = 75$ ، فئة الربع الثالث
$Q1 = 15 + [(25-22)/30] \times [20-15]$ $5 = 15.5$	$Q3 = 20 + [(75-52)/25] \times [25-20]$ $5 = 24.6$

$$SKq = (24.6 + 15.5 - 2 \times 19.67) / (24.6 - 15.5) = (40.1 - 39.34) / 9.1 \approx 0.08$$

بما أن معامل الالتواء يول وكيندال موجب، فإن التوزيع ملتوٍ نحو اليمين، وهو ما يتوافق مع نتيجة معامل بيرسون أعلاه، وإن كانت درجة الالتواء المقاسة بهذه الطريقة أضعف نسبياً.

مثال رقم (05-07): احسب معامل الالتواء الربيعي لتوزيع أعمار المشتركين في النادي الرياضي (مثال 05-03).

الحل: سبق حساب الوسيط أعلاه ($Me = 38.18$)؛ نحسب الآن الربع الأول والربع الثالث:

$$Q1 = F/4 = 20، \text{ فئة } [35-30] \text{ رتبة}$$

$$Q3 = 3F/4 = 60، \text{ فئة } [45-40] \text{ ورتبة}$$

$$Q1 = 30 + [(20 - 12) / 14] \times 5 \approx 32.86$$

$$Q3 = 40 + [(60 - 48) / 20] \times 5 = 43$$

$$SKq = (43 + 32.86 - 2 \times 38.18) / (43 - 32.86) = -0.50 / 10.14 \approx -0.05$$

بما أن معامل الالتواء يول وكيندال سالب، فإن التوزيع ملتوٍ نحو اليسار، وهو ما يتوافق تماماً مع نتيجة معامل بيرسون بالمعادلتين أعلاه.

معامل الالتواء فيشر (Fisher Skewness Coefficient)

هو نسبة العزم المركزي الثالث إلى مكعب الانحراف المعياري، ويُعطى بالعلاقة التالية:

$$SK_f = \mu_3 / \delta_x^3$$

ويتراوح مجال تغيره بين -1 و $+1$ كذلك؛ فإذا كانت قيمته سالبة فالالتواء نحو اليسار، وإذا كانت موجبة فالالتواء نحو اليمين، أما إذا كانت معدومة فالتوزيع متماثل.

مثال رقم (05-08): احسب معامل الالتواء فيشر لمعطيات المثال السابق رقم (05-01).

الحل: سبق حساب العزم المركزي الثالث في المثال (05-02) $(\mu_3 \approx 87.44)$ ، والانحراف المعياري في المثال (05-03) $(\delta_x \approx 6.32)$ ؛ نطبق القانون مباشرة:

$$SK_f = 87.44 / 6.32^3 = 87.44 / 252.44 \approx 0.35$$

بما أن معامل الالتواء فيشر موجب، فإن التوزيع ملتوٍ نحو اليمين، وهي نفس النتيجة والاتجاه اللذين توصلنا إليهما بالطريقتين السابقتين، مما يعزز الثقة في هذا الاستنتاج.

مثال رقم (05-09): احسب معامل الالتواء فيشر لتوزيع أعمار المشتركين في النادي الرياضي (مثال 05-04).

الحل: العزم المركزي الثالث لهذا التوزيع $\mu_3 \approx -126.95$ ، والانحراف المعياري $\delta_x \approx 6.85$:

$$SK_f = -126.95 / 6.85^3 = -126.95 / 321.42 \approx -0.40$$

بما أن معامل الالتواء فيشر سالب، فإن التوزيع ملتوٍ نحو اليسار، وهي نفس النتيجة التي توصلنا إليها بالطريقتين السابقتين (بيرسون ويول وكيندال)، مما يؤكد اتساق الاستنتاج عبر مختلف طرق القياس، سواء كان الالتواء موجباً كما في المثال الأول أو سالباً كما في هذا المثال الثاني.

تطبيقات عملية لمقاييس الالتواء: يُستخدم الالتواء في مجالات عديدة، منها التمويل (لدراسة توزيع عوائد الأسهم، حيث يفضّل المستثمرون عادة الالتواء الموجب الذي يعني احتمال تحقيق عوائد استثنائية مرتفعة)، والتأمين (لتقدير احتمال الخسائر الكبيرة النادرة)، وضبط الجودة الصناعية (لرصد انحراف الإنتاج عن المواصفات في اتجاه معين). فمعرفة اتجاه الالتواء لا تكتفي بوصف الشكل، بل تساعد أيضاً في اتخاذ قرارات عملية دقيقة.

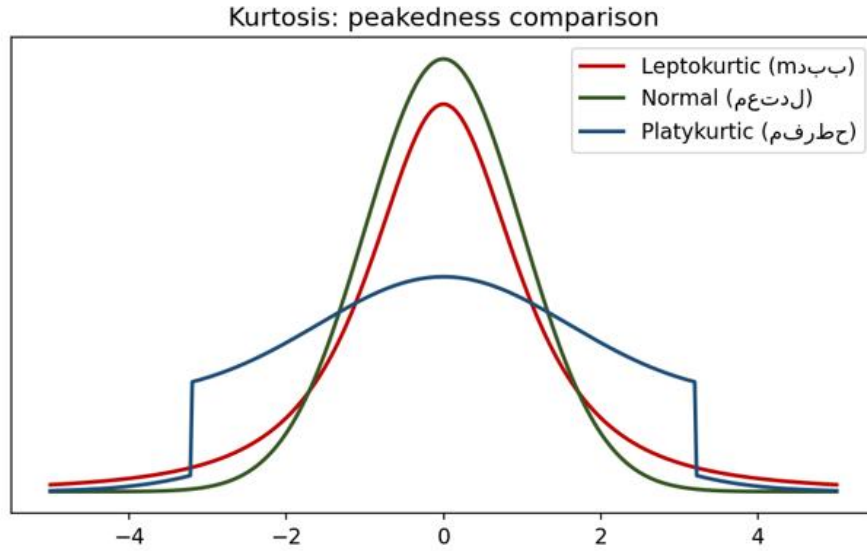
ثالثاً: التفرطح (Kurtosis)

التفرطح مقياس يصف ارتفاع قمة المنحنى من حيث الاعتدال أو التدبب أو التفرطح؛ فعند تمثيل التوزيع التكراري بمنحنى تكراري، قد يكون هذا المنحنى مدبباً أو منبسّطاً (مفرطحاً): فعندما يتركز عدد أكبر من القيم بالقرب من منتصف المنحنى ويقل في طرفيه، يكون المنحنى مدبباً؛ وعندما يتركز عدد أكبر من القيم على طرفي المنحنى ويقل بالقرب من المنتصف، يكون المنحنى مفرطحاً أو منبسّطاً¹.

ويمكن الفرق الجوهرى بين هذه الأشكال في درجة تشتت البيانات حول الوسط الحسابي: فالمنحنى المدبب يعكس تشتتاً قوياً (تركز القيم قرب الوسط مع بعض القيم البعيدة)، بينما يعكس المنحنى المفرطح تشتتاً أضعف نسبياً (انتشاراً أوسع وأكثر انتظاماً للقيم)، أما المنحنى المعتدل (الطبيعي) فيمثل الحالة الوسطى بينهما.

والشكل التالي يقارن بين منحنى مدبب ومنحنى مفرطح، بالنسبة إلى المنحنى الطبيعي المعتدل:

¹ نداء محمد الصوص، مبادئ الإحصاء، الطبعة الأولى، مكتبة المجتمع العربي للنشر والتوزيع، عمان، 2007، ص65



الشكل 5-2: مقارنة بين منحنى مدبب ومنحنى مفرطح ومنحنى معتدل

ويمكن قياس التفرطح باستخدام عدد من الطرق، من أهمها: معامل بيرسون، معامل فيشر، ومعامل كيلي.

(a) معامل التفرطح بيرسون (Pearson Kurtosis Coefficient)

يُحسب بالعلاقة التالية:

$$K_p = \mu_4 / \delta_x^4$$

حيث μ_4 العزم المركزي الرابع، و δ_x^4 الانحراف المعياري مرفوعاً للقوة الرابعة. فإذا كان معامل التفرطح بيرسون أكبر من "3" يكون المنحنى مدبباً، وإذا كان أقل من "3" يكون المنحنى مفرطحاً، أما إذا كان يساوي "3" تماماً فيكون المنحنى معتدلاً (كالتوزيع الطبيعي).

مثال رقم (05-10): احسب معامل التفرطح بيرسون لمعطيات المثال السابق رقم (05-01).

الحل: سبق حساب العزم المركزي الرابع في المثال (05-02) ($\mu_4 \approx 2995.97$)، والانحراف المعياري في المثال (05-03) ($\delta_x \approx 6.32$)، وبالتالي ($\delta_x^4 \approx 1341.85$):

$$K_p = 2995.97 / 1341.85 \approx 2.23$$

بما أن معامل التفرطح بيرسون (2.23) أقل من 3، فإن منحنى التوزيع مفطح.

معامل التفرطح فيشر (Fisher Kurtosis Coefficient)

استعمل فيشر مؤشر بيرسون مطروحاً منه العدد ثلاثة، وفق العلاقة التالية:

$$k_f = (\mu_4 / \delta_x^4) - 3$$

فإذا كان معامل التفرطح فيشر أكبر من "0" يكون المنحنى مدبباً، وإذا كان أقل من "0" يكون المنحنى مفطحاً، أما إذا كان يساوي "0" فيكون المنحنى معتدلاً.

مثال رقم (05-11): احسب معامل التفرطح فيشر لمعطيات المثال السابق رقم (05-01).

الحل: بتطبيق نفس نتيجة معامل بيرسون المحسوب أعلاه مباشرة:

$$K_f = 2.23 - 3 = -0.77$$

بما أن معامل التفرطح فيشر (-0.77) أقل من صفر، فإن منحنى التوزيع مفطح، وهي نفس النتيجة التي توصلنا إليها بمعامل بيرسون.

معامل التفرطح كيلي (Kelly Kurtosis Coefficient)

يُحسب بإحدى العلاقتين التاليتين، إما باستعمال المئينات أو باستعمال العشيريات:

(a) باستعمال العشيريات

$$K = (Q_3 - Q_1) / [2 * (D_9 - D_1)]$$

(b) باستعمال المئينات

$$K = (Q_3 - Q_1) / [2 * (P_{90} - P_{10})]$$

حيث Q_1 الربع الأول، Q_3 الربع الثالث، P_{10} المئين العاشر، P_{90} المئين التسعون، D_1 العشير الأول، و D_9 العشير التاسع؛ مع العلم أن: $D_1 = P_{10}$ و $D_9 = P_{90}$.

ويُعرف شكل التوزيع حسب قيمة معامل كيلي (K) كما يلي: إذا كانت $K = 0.263$ فالتوزيع طبيعي (معتدل أو متماثل)؛ وإذا كانت $K > 0.263$ فالتوزيع مدبب؛ وإذا كانت $K < 0.263$ فالتوزيع مفرطح.

مثال رقم (05-12): احسب معامل التفرطح كيلي لمعطيات المثال السابق رقم (05-01).

الحل: سبق حساب قيمتي الربع الأول والثالث في المثال (05-04): $Q_1 = 15.5$ ، $Q_3 = 24.6$. نحسب الآن المئين العاشر والمئين التسعون:

المئين التسعون P_{90}	المئين العاشر P_{10}
رتبة $P_{90} = 90F/100 = 90$ ، فئة P_{90} $P_{90} = 25 + [(90-77)/15] \times 5 \approx 29.33$ $[25-30]$	رتبة $P_{10} = 10F/100 = 10$ ، فئة P_{10} $P_{10} = 10 + [(10-0)/22] \times 5 \approx 12.27$ $[10-15]$

$$K = (24.6 - 15.5) / [2 \times (29.33 - 12.27)] = 9.1 / 34.12 \approx 0.267$$

بما أن قيمة معامل كيلي (0.267) قريبة جداً من القيمة المرجعية (0.263) وأكبر منها بفارق ضئيل، فإن التوزيع يُصنّف نظرياً مدبباً بدرجة طفيفة جداً حسب هذا المعيار بالذات، في حين صنّفه معامل بيرسون وفisher مفرطحاً بوضوح أكبر. ويُفسّر هذا التباين البسيط بين النتائج باختلاف حساسية كل معامل لجزء مختلف من التوزيع (معامل كيلي يركز فقط على المسافة بين الربعين والمئينين المتطرفين، بينما يعتمد معامل بيرسون وفisher على جميع قيم التوزيع)، وهو ما يؤكد أهمية عدم الاكتفاء بمعامل واحد عند دراسة شكل توزيع ما، بل الأجدر مقارنة نتائج عدة معاملات معاً للوصول إلى تشخيص أدق.

مثال رقم (05-13): احسب معامل التفرطح كيلي لتوزيع أعمار المشتركين في النادي الرياضي.

الحل: سبق حساب $Q_1 = 32.86$ و $Q_3 = 43$ أعلاه. نحسب الآن المئين العاشر والمئين التسعون:

(فئة [50-45]: $P90 = 90F/100 = 72$ ؛ فئة [30-25]: $P10 = 10F/100 = 8$ ، ورتبة 8

$$P10 = 25 + [(8 - 4) / 8] \times 5 = 27.5$$

$$P90 = 45 + [(72 - 68) / 12] \times 5 \approx 46.67$$

$$K = (43 - 32.86) / [2 \times (46.67 - 27.5)] = 10.14 / 38.34 \approx 0.265$$

وهذه القيمة (0.265) قريبة جداً من القيمة المرجعية (0.263) أيضاً، بنفس الطريقة التي لاحظناها في المثال الأول، مما يؤكد أن معامل كيلي يميل غالباً إلى إعطاء نتائج قريبة من حد الاعتدال في التوزيعات التي لا يكون تفرطحها أو التواءها شديدين.

خلاصة: يلخص الجدول التالي أهم معاملات الالتواء والتفرطح المدروسة في هذا الفصل، صيغتها،

ومجال دلالتها:

المعامل	الصيغة	متى يُستعمل	دلالة الإشارة/القيمة
بيرسون للالتواء	$(X - Mo) / \delta x$ $(X - Me) / \delta x^3$	توزيع قريب من التماثل	موجب: يمين، سالب: يسار، صفر: تماثل
يول وكيندال	$(Q_3 + Q_1 - 2Me) / (Q_3 - Q_1)$	جداول مفتوحة أو قيم شاذة	موجب: يمين، سالب: يسار، صفر: تماثل
فيشر للالتواء	$\mu_3 / \delta x^3$	الحساب الدقيق الشامل	موجب: يمين، سالب: يسار، صفر: تماثل
بيرسون للتفرطح	$\mu_4 / \delta x^4$	الحساب الدقيق الشامل	3: > 3: < 3: = 3: معتدل
فيشر للتفرطح	الصيغة	متى يُستعمل	دلالة الإشارة/القيمة
كيلي للتفرطح	$(X - Mo) / \delta x$ $(X - Me) / \delta x^3$	توزيع قريب من التماثل	موجب: يمين، سالب: يسار، صفر: تماثل

خطوات عملية لدراسة شكل توزيع تكراري: يمكن تلخيص منهجية دراسة شكل أي توزيع تكراري في الخطوات التالية: أولاً، حساب الوسط الحسابي والانحراف المعياري (وهما ضروريان لمعظم المعاملات). ثانياً، تحديد المنوال والوسيط والربيعيات إذا لزم الأمر. ثالثاً، حساب معامل أو أكثر من معاملات الالتواء لتحديد اتجاه ودرجة الابتعاد عن التماثل. رابعاً، حساب معامل أو أكثر من معاملات التقعر لتحديد درجة تركيز القيم حول الوسط. وأخيراً، تفسير النتائج مجتمعة لتكوين صورة شاملة عن شكل التوزيع، مع الانتباه إلى أن اختلاف النتائج الطفيف بين المعاملات المختلفة أمر طبيعي ولا يعني وجود خطأ، بل يعكس زوايا نظر مختلفة لنفس الظاهرة.

تمارين الفصل الخامس

التمرين الأول:

البيانات التالية تمثل توزيع عينة من 100 أسرة حسب عدد أفرادها:

عدد الأفراد	2	3	4	5	6	7	8	9	المجموع
عدد الأسر	5	12	28	32	14	6	2	1	100

المطلوب:

- احسب العزم المركزي الثالث والرابع.
- احسب معامل الالتواء ومعامل التفرطح فيشر، مع التعليق على النتيجة.

الحل:

نحسب أولاً الوسط الحسابي:

x_i	f_i	$f_i \times x_i$	$f_i \times x_i^2$	$(x_i - \bar{X})^3$	$f_i(x_i - \bar{X})^3$	$f_i(x_i - \bar{X})^4$
2	5	10	20	-19.465	-97.33	522.29
3	12	36	108	-4.827	-57.92	148.11
4	28	112	448	-0.329	-9.20	11.75
5	32	160	800	0.030	0.95	0.29
6	14	84	504	2.248	31.47	41.05
7	6	42	294	12.331	73.96	89.90
8	2	16	128	36.238	72.48	260.30
9	1	9	81	80.062	80.06	645.30
المجموع	100	469	2383	/	94.47	1719.29

$$Kf = \mu_4 / \delta x^4 - 3 = 17.19 / 1.35^4 - 3 = 17.19 / 3.32 - 3 \approx 5.18 - 3 = 2.18$$

بما أن معامل التقعر فيشر (2.18) أكبر من صفر بوضوح، فإن منحنى التوزيع مدبب، أي أن القيم تتركز بشكل كبير حول القيمة المنوالية (5 أفراد) مقارنة بالتوزيع الطبيعي.

التمرين الثاني:

الجدول التالي يمثل توزيع الأجر الأسبوعي لعينة من 25 عاملاً في أحد المصانع (الوحدة: 10^3 دينار جزائري):

الأجر الأسبوعي	15-5	25-15	35-25	45-35	المجموع
عدد العمال	4	9	8	4	25

المطلوب: احسب معاملات الالتواء والتقعر الثلاثة (بيرسون، يول وكيندال أو كيلي حسب المتاح، وفيشر).

الحل:

الفئات	f_i	x_i	$f_i \times x_i$	$f_i \times x_i^2$	$F \uparrow$
5-15	4	10	40	400	4
15-25	9	20	180	3600	13
25-35	8	30	240	7200	21
35-45	4	40	160	6400	25
المجموع	25	/	620	17600	/

$$\bar{X} = 620 / 25 = 24.8 (10^3 \text{ دج})$$

$$\delta x = \sqrt{(17600/25 - 24.8^2)} = \sqrt{(704 - 615.04)} = \sqrt{88.96} \approx 9.43$$

- معامل بيرسون للالتواء: الفئة المنوالية هي [15-25] (أكبر تكرار 9):

$$Mo = 15 + [5 / (5+1)] \times 10 \approx 23.33$$

$$SKp = (24.8 - 23.33) / 9.43 \approx 0.16$$

بما أن معامل الالتواء بيرسون موجب، فإن التوزيع ملتوٍ نحو اليمين.

- معامل الالتواء يول وكيندال: رتبة الوسيط $F/2 = 12.5$ ، الفئة الوسيطة [15-25]:

$$Me = 15 + [(12.5 - 4) / 9] \times 10 \approx 24.44$$

$$\text{ورتبة } Q1 = F/4 = 6.25 \text{ (فئة [15-25])،}$$

(فئة [25-35]) $Q3 = 3F/4 = 18.75$ (فئة [15-25])، ورتبة $Q1 = F/4 = 6.25$ ورتبة

$$Q1 = 15 + [(6.25-4)/9] \times 10 \approx 17.5$$

$$Q3 = 25 + [(18.75-13)/8] \times 10 \approx 32.19$$

$$SKq = (32.19 + 17.5 - 2 \times 24.44) / (32.19 - 17.5) = 0.81 / 14.69 \approx 0.055$$

بما أن معامل الالتواء يول وكيندال موجب (وإن كان ضعيفاً جداً)، فإن التوزيع ملتوٍ نحو اليمين، وهو ما يتوافق مع نتيجة معامل بيرسون.

- معامل الالتواء فيشر: بحساب العزم المركزي الثالث (بنفس أسلوب المثال 05-02)

$$\text{نجد } \mu_3 \approx 48.38$$

$$SKf = 48.38 / 9.43^3 = 48.38 / 838.56 \approx 0.058$$

بما أن معامل الالتواء فيشر موجب أيضاً، فإن جميع المعاملات الثلاثة تتفق على أن التوزيع ملتوٍ نحو اليمين، وإن كانت درجة الالتواء ضعيفة بشكل عام (قريبة من التماثل).

- معامل التفرطح بيرسون وفيشر: بحساب العزم المركزي الرابع نجد $\mu_4 \approx 16642.36$:

$$Kp = 16642.36 / 9.43^4 = 16642.36 / 7913.86 \approx 2.10$$

$$Kf = 2.10 - 3 = -0.90$$

بما أن معامل التفرطح بيرسون (2.10) أقل من 3، ومعامل فيشر (-0.90) أقل من صفر، فإن منحنى توزيع الأجور مفرطح، أي أن أجور العمال متوزعة بشكل أكثر انتشاراً حول الوسط الحسابي مقارنة بالتوزيع الطبيعي.

التمرين الثالث:

سُجِّلَت أزمان انتظار 60 زبوناً في فرع بنكي معين (بالدقائق)، فكانت موزعة كما يلي:

زمن الانتظار	10-0	20-10	30-20	40-30	50-40	المجموع
عدد الزبائن	6	16	20	12	6	60

المطلوب:

- 1) احسب الوسط الحسابي والانحراف المعياري لهذا التوزيع.
- 2) احسب العزم المركزي الثالث والرابع باستخدام العلاقة المختصرة بدلالة العزوم البسيطة.
- 3) احسب معامل الالتواء بيرسون (بالمنوال) ومعامل التفرطح فيشر، وعلّق على شكل التوزيع.

الحل:

1- الوسط الحسابي والانحراف المعياري:

الفئات	f_i	x_i	$f_i \times x_i$	$f_i \times x_i^2$	$f_i \times x_i^3$	$f_i \times x_i^4$
0-10	6	5	30	150	750	3750
10-20	16	15	240	3600	54000	810000
20-30	20	25	500	12500	312500	7812500

18007500	514500	14700	420	35	12	30-40
24603750	546750	12150	270	45	6	40-50
51237500	1428500	43100	1460	/	60	المجموع

$$m_1 = \bar{X} = 1460 / 60 \approx 24.33 \text{ دقيقة}$$

$$m_2 = 43100 / 60 \approx 718.33$$

$$\delta x = \sqrt{(718.33 - 24.33^2)} = \sqrt{(718.33 - 592.11)} = \sqrt{126.22} \approx 11.23 \text{ دقيقة}$$

2- العزم المركزي الثالث والرابع بالعلاقة المختصرة:

$$m_3 = 1428500 / 60 \approx 23808.33, m_4 = 51237500 / 60 \approx 853958.33$$

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3 = 23808.33 - 3 \times 24.33 \times 718.33 + 2 \times 24.33^3 \approx 186.07$$

$$\mu_4 = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4 = 853958.33 - 4 \times 24.33 \times 23808.33 + 6 \times 24.33^2 \times 718.33 - 3 \times 24.33^4 \approx 36826.07$$

3- معامل الالتواء بيرسون ومعامل التقطح فيشر: الفئة المنوالية هي [20-30] (أكبر تكرار 20):

$$Mo = 20 + [4 / (4+8)] \times 10 \approx 23.33$$

$$SKp = (24.33 - 23.33) / 11.23 \approx 0.09$$

بما أن معامل الالتواء بيرسون موجب (وإن كان ضعيفاً جداً، قريباً من الصفر)، فإن التوزيع ملتو قليلاً نحو اليمين، أي أنه قريب من التماثل مع ميل طفيف نحو أزمدة الانتظار الأطول.

$$Kf = \mu_4 / \delta x^4 - 3 = 36826.07 / 11.23^4 - 3 = 36826.07 / 15901.98 - 3 \\ \approx 2.32 - 3 = -0.68$$

بما أن معامل التفرطح فيشر (−0.68) أقل من صفر، فإن منحنى توزيع أزمنة الانتظار مفرطح، أي أن أزمنة الانتظار منتشرة بشكل أوسع نسبياً حول الوسط الحسابي مقارنة بالتوزيع الطبيعي، ولا تتركز بشدة حول قيمة واحدة.

التمرين الرابع:

يمثل الجدول التالي توزيع 50 موظفاً حسب عدد سنوات الأقدمية في منصب عملهم الحالي:

سنوات الأقدمية	1	2	3	4	5	6	7	المجموع
عدد الموظفين	10	14	12	7	4	2	1	50

المطلوب:

(1) احسب العزوم البسيطة الأربعة الأولى، ثم استنتج منها العزمين المركزيين الثالث والرابع بالعلاقة المختصرة.

(2) احسب معامل الالتواء بيرسون (بالمنوال) ومعامل الالتواء فيشر.

(3) احسب معامل التفرطح فيشر، وعلّق على النتيجة معاً (الالتواء والتفرطح).

الحل:

1- العزوم البسيطة والمركزية:

$$m_4 = 209.82, m_3 = 43.38, m_2 = 10.14, m_1 = X = 2.82$$

$$\delta x = \sqrt{(m_2 - m_1^2)} = \sqrt{(10.14 - 7.9524)} = \sqrt{2.1876} \approx 1.48$$

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3 \approx 2.45$$

$$\mu_4 = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4 \approx 14.60$$

2- معامل الالتواء: المنوال هو القيمة صاحبة أكبر تكرار، أي $Mo = 2$ سنتان (تكرار 14):

$$SKp = (2.82 - 2) / 1.48 \approx 0.55$$

$$SKf = \mu_3 / \delta x^3 = 2.45 / 1.48^3 = 2.45 / 3.24 \approx 0.76$$

بما أن قيمتي معاملي الالتواء موجبتان معاً، فإن التوزيع ملتوٍ بوضوح نحو اليمين: أغلب الموظفين حديثو التعيين نسبياً (أقدمية سنتين أو أقل)، مع أقلية قليلة من ذوي الأقدمية الطويلة تسحب الوسط الحسابي نحو الأعلى.

3- معامل التفرطح فيشر:

$$Kf = \mu_4 / \delta x^4 - 3 = 14.60 / 1.48^4 - 3 = 14.60 / 4.80 - 3 \approx 3.04 - 3 = 0.04$$

بما أن معامل التفرطح فيشر (0.04) قريب جداً من الصفر (وموجب بفارق ضئيل جداً)، فإن منحني هذا التوزيع قريب من الاعتدال في تفرطه، رغم كونه ملتوياً بوضوح نحو اليمين. وهذا مثال جيد يوضح أن الالتواء والتفرطح مقياسان مستقلان تماماً عن بعضهما: فقد يكون التوزيع شديد الالتواء دون أن يكون بالضرورة شديد التدبب أو التفرطح، والعكس صحيح أيضاً.

الفصل السادس

الارتباط والاختلاف الحظي البسيط

الفصل السادس: الارتباط والانحدار الخطي البسيط

تمهيد

اقتصر تحليلنا في الفصول السابقة على دراسة متغير إحصائي واحد في كل مرة (كالوسط الحسابي لأجور موظفي مؤسسة، أو الانحراف المعياري لأعمار عينة من المرضى)؛ غير أن الواقع العملي كثيراً ما يفرض على الباحث دراسة متغيرين إحصائيين معاً، بهدف الكشف عما إذا كانت بينهما علاقة أو رابطة إحصائية، ومدى قوة هذه الرابطة واتجاهها.

فقد يرغب باحث في معرفة ما إذا كان هناك ارتباط بين عدد ساعات المذاكرة ونتائج الطلبة في الامتحانات، أو بين حجم الإنفاق الإعلاني ورقم مبيعات مؤسسة تجارية، أو بين المستوى التعليمي ومستوى الدخل. ولدراسة مثل هذه العلاقات، نحتاج أولاً إلى أدوات لعرض بيانات متغيرين معاً (الجدول التكرارية الثنائية)، ثم إلى مقاييس رقمية تلخص قوة العلاقة واتجاهها (معاملات الارتباط)، وأخيراً إلى نموذج رياضي يسمح بالتنبؤ بقيمة أحد المتغيرين انطلاقاً من الآخر (الانحدار الخطي البسيط)، وهو ما يتناوله هذا الفصل بالتفصيل.

أولاً: جداول التوزيعات الثنائية (Bivariate Distributions)

عندما ندرس متغيرين إحصائيين X و Y في آن واحد لنفس مجموعة المفردات، فإننا نحصل على ما يُسمى بالتوزيع التكراري الثنائي أو المشترك، ويُعرض عادة في جدول ذي مدخلين يبين تكرار كل توليفة ممكنة من قيم المتغيرين معاً.

الجدول التكراري المشترك

يُعرّف الجدول التكراري المشترك بأنه جدول يوضح عدد المفردات (التكرار المشترك) التي تحمل قيمة معينة من المتغير الأول X مقترنة بقيمة معينة من المتغير الثاني Y في آن واحد، ويُرمز للتكرار المشترك للخلية (i, j) بالرمز f_{ij} .

مثال رقم (06-01): يمثل الجدول التالي توزيع 900 طالب في إحدى الجامعات حسب الجنس والتخصص الدراسي (علمي، أدبي، تقني):

الجنس / التخصص	علمي	أدبي	تقني	المجموع
ذكور	210	150	90	450
إناث	180	140	130	450
المجموع	390	290	220	900

يمثل هذا الجدول التوزيع التكراري المشترك للمتغيرين (الجنس والتخصص)؛ فمثلاً يشير الرقم 210 (تقاطع سطر "ذكور" وعمود "علمي") إلى أن 210 طالباً ذكراً ملتحقون بالتخصص العلمي.

الجدول التكرارية الهامشية

التوزيع الهامشي لأحد المتغيرين هو التوزيع التكراري البسيط لهذا المتغير وحده، بغض النظر عن قيم المتغير الآخر؛ ويُحصل عليه بجمع تكرارات كل سطر (للحصول على التوزيع الهامشي للمتغير الأول) أو كل عمود (للحصول على التوزيع الهامشي للمتغير الثاني) في الجدول المشترك¹. (1)

مثال رقم (06-02): استخراج التوزيعين الهامشيين لمتغيري الجنس والتخصص من بيانات المثال السابق (06-01).

الحل: التوزيع الهامشي للجنس هو مجموع كل سطر:

الجنس	ذكور	اناث	المجموع
التكرار	450	450	900

والتوزيع الهامشي للتخصص هو مجموع كل عمود:

التخصص	علمي	ادبي	تقني	المجموع
التكرار	390	290	220	900

ونلاحظ أن كلا التوزيعين الهامشيين يمثلان في الحقيقة توزيعاً تكرارياً بسيطاً لمتغير واحد فقط، تماماً كما سبق دراسته في الفصل الثاني.

¹ عبد الحليم الحمزة، كتاب بيداغوجي في الإحصاء 1، كلية العلوم الاقتصادية والعلوم التجارية وعلوم التسيير، جامعة العربي التبسي، تبسة، 2022، ص ص 100-103.

الجدول التكرارية الشرطية

التوزيع الشرطي لمتغير Y بشرط قيمة معينة من المتغير X هو توزيع تكرارات Y ضمن سطر (أو عمود) واحد فقط من الجدول المشترك المقابل لتلك القيمة من X ، وعادة ما يُعبر عنه بالتكرار النسبي (بقسمة كل تكرار على مجموع السطر أو العمود المعني)¹ (1).

مثال رقم (06-03): استخرج التوزيع الشرطي للتخصص لدى الذكور، والتوزيع الشرطي للتخصص لدى الإناث، من بيانات المثال (06-01)، وقارن بينهما.

الحل: بقسمة تكرار كل تخصص في سطر الذكور على مجموع الذكور (450)، وفي سطر الإناث على مجموع الإناث (450):

المجموع	تقني	أدبي	علمي	التخصص
100	20.00	33.33	46.67	التوزيع الشرطي لدى الذكور (%)
100	28.89	31.11	40.00	التوزيع الشرطي لدى الإناث (%)

نلاحظ أن نسبة الالتحاق بالتخصص العلمي أعلى نسبياً لدى الذكور (46.67%) منها لدى الإناث (40%).

بينما نسبة الالتحاق بالتخصص التقني أعلى لدى الإناث (28.89%) منها لدى الذكور (20%).

وهذا النوع من المقارنة بين التوزيعات الشرطية هو بالضبط ما يسمح لنا باستكشاف وجود علاقة محتملة بين المتغيرين قبل الانتقال إلى القياس الرقمي الدقيق لهذه العلاقة.

¹ عبد الحليم الحمزة، مرجع سابق

التباين المشترك (Covariance)

التباين المشترك مقياس يصف اتجاه العلاقة الخطية بين متغيرين X و Y معاً (موجبة أو سالبة)، ويُحسب كمتوسط حاصل ضرب انحرافات كل من X و Y عن وسطيهما الحسابيين، وفق العلاقة التالية¹: (1)

$$Cov(X, Y) = \sum n_{ij}(x_i - \bar{X})(y_i - \bar{Y})/N$$

فإذا كانت قيمة التباين المشترك موجبة، دلّ ذلك على أن المتغيرين يتغيران في نفس الاتجاه (كلما زاد أحدهما زاد الآخر)؛ وإذا كانت سالبة، دلّ ذلك على أنهما يتغيران في اتجاهين متعاكسين؛ أما إذا كانت معدومة أو قريبة من الصفر، فهذا يعني غياب أي علاقة خطية بينهما.

مثال رقم (06-04): يمثل الجدول التالي توزيع 16 موظفاً حسب عدد سنوات الخبرة (X) والراتب الشهري بالآلاف من الدينار (Y):

$X \setminus Y$	3	5	المجموع
2	5	3	8
4	2	6	8
المجموع	7	9	16

المطلوب: احسب التباين المشترك بين عدد سنوات الخبرة والراتب الشهري.

الحل: نحسب أولاً الوسطين الحسابيين:

$$\bar{X} = (2 \times 8 + 4 \times 8) / 16 = 48 / 16 = 3$$

$$\bar{Y} = (3 \times 7 + 5 \times 9) / 16 = 66 / 16 = 4.125$$

¹ عبد الحليم الحمزة، مرجع سابق

ثم نحسب مجموع حاصل ضرب الانحرافات مرجحة بتكرارها المشترك لكل خلية:

$$\sum f_{ij}(x_i - \bar{X})(y_j - \bar{Y}) = 5(2-3)(3-4.125) + 3(2-3)(5-4.125) + 2(4-3)(3-4.125) \\ + 6(4-3)(5-4.125) = 5.625 + (-2.625) + (-2.25) + 5.25 = 6$$

$$\text{Cov}(X, Y) = 6 / 16 = 0.375$$

بما أن قيمة التباين المشترك موجبة (0.375)، فإن هناك علاقة موجبة (طردية) بين عدد سنوات الخبرة والراتب الشهري: كلما زادت سنوات الخبرة، زاد الراتب الشهري في الغالب، وهي نتيجة منطقية ومتوقعة.

ملاحظة: رغم أن التباين المشترك يعطينا فكرة عن اتجاه العلاقة، إلا أنه لا يخبرنا شيئاً عن قوة هذه العلاقة، كما أن قيمته تتأثر بوحدة قياس كل من المتغيرين، مما يجعل مقارنة قوة الارتباط بين أزواج مختلفة من المتغيرات أمراً صعباً باستعماله وحده. ولتجاوز هذا القصور، طُوّر معامل الارتباط الخطي (بيرسون)، الذي يُعد بمثابة تباين مشترك "موحد" أو "مُطَبَّع"، كما سيرد تفصيله في القسم الموالي.

ثانياً: معامل الارتباط الخطي بيرسون (Pearson's Correlation Coefficient)

معامل الارتباط الخطي، المعروف أيضاً بمعامل ارتباط بيرسون، هو أكثر مقاييس الارتباط استعمالاً، ويقيس درجة واتجاه العلاقة الخطية (المستقيمة) بين متغيرين كميين X و Y ، ويُرمز له بالرمز (r) .

ويُعطى معامل الارتباط بيرسون بنسبة التباين المشترك إلى حاصل ضرب الانحرافين المعياريين للمتغيرين، وهو ما يجعله مقياساً موحداً (بلا وحدة قياس)، محصوراً دوماً بين -1 و $+1$:

$$r = \text{Cov}(X, Y) / (\delta_x * \delta_y)$$

ويمكن تفسير قيمة معامل الارتباط بيرسون كما يلي:

- إذا كانت $r = +1$: ارتباط خطي موجب تام (كل الأزواج تقع بالضبط على مستقيم صاعد).

- إذا كانت $r = -1$: ارتباط خطي سالب تام (كل الأزواج تقع بالضبط على مستقيم نازل).
- إذا كانت $r = 0$: انعدام تام للعلاقة الخطية بين المتغيرين (قد توجد علاقة غير خطية رغم ذلك).
- وكلما اقتربت $|r|$ من الواحد الصحيح، كانت العلاقة الخطية أقوى، وكلما اقتربت من الصفر كانت أضعف.

(a) معامل الارتباط بيرسون للبيانات غير المبوبة

في حالة البيانات غير المبوبة، يُحسب معامل الارتباط بيرسون مباشرة من الصيغة العملية التالية، الأكثر استعمالاً لتفادي حساب الانحرافات عن الوسط لكل قيمة على حدة:

$$r = \frac{[n \sum x_i y_i - \sum x_i \sum y_i]}{\sqrt{\{[n \sum x_i^2 - (\sum x_i)^2][n \sum y_i^2 - (\sum y_i)^2]\}}}$$

مثال رقم (05-06): سجّلت إدارة إحدى المؤسسات التكوينية عدد ساعات المذاكرة الأسبوعية (X) وعلامة الاختبار النهائي من 100 نقطة (Y) لعينة من 10 متدربين:

المتدرب	1	2	3	4	5	6	7	8	9	10
ساعات المذاكرة (X)	2	3	4	4	5	6	7	7	8	9
علامة الاختبار (Y)	40	45	50	55	58	65	70	75	80	90

المطلوب: احسب معامل الارتباط بيرسون بين عدد ساعات المذاكرة وعلامة الاختبار، وعلّق على النتيجة.

الحل: نحسب الأعمدة المساعدة اللازمة:

الإحصائية	Σx_i	Σy_i	$\Sigma x_i y_i$	Σx_i^2	Σy_i^2
القيمة	55	628	3780	349	41764

$$r = [10 \times 3780 - 55 \times 628] / \sqrt{\{[10 \times 349 - 55^2][10 \times 41764 - 628^2]\}}$$

$$r = [37800 - 34540] / \sqrt{\{[3490 - 3025][417640 - 394384]\}} = 3260 / \sqrt{(465 \times 23256)}$$

$$r = 3260 / \sqrt{10814040} \approx 3260 / 3288.5 \approx 0.991$$

بما أن معامل الارتباط بيرسون (0.991) قريب جداً من الواحد الصحيح وموجب الإشارة، فإن هناك ارتباطاً خطياً موجباً قوياً جداً بين عدد ساعات المذاكرة الأسبوعية وعلامة الاختبار النهائي: فكلما زاد عدد ساعات المذاكرة، ارتفعت العلامة المتحصل عليها بشكل شبه منتظم.

معامل الارتباط بيرسون للبيانات المبوبة (طريقة الانحرافات المختصرة)

عندما يكون حجم البيانات كبيراً، يُلجأ إلى تبويب كل من المتغيرين X و Y في فئات، وعرض تكراراتهما المشتركة في جدول ثنائي المدخل؛ ويُحسب معامل الارتباط بيرسون في هذه الحالة بطريقة الانحرافات المختصرة (شبيهة بطريقة الترميز المستعملة لحساب الوسط الحسابي)، وذلك باختيار وسط فرضي X_0 و Y_0 (عادة مركزا الفئتين الأكثر تكراراً)، ثم حساب الانحرافات المرمزة:

$$dx_i = \frac{(x_i - X_0)}{K_x}, dy_i = \frac{(y_i - Y_0)}{K_y}$$

حيث K_x و K_y طولاً فئتي X و Y على التوالي. ثم يُطبَّق معامل الارتباط بيرسون على القيم المرمزة مباشرة، إذ إن الترميز الخطي لا يغيّر قيمة معامل الارتباط:

$$r = \frac{[N \sum n_{ij} dx_i dy_j - \sum n_i dx_i \sum n_j dy_j]}{\sqrt{\{[N \sum n_i dx_i^2 - (\sum n_i dx_i)^2][N \sum n_j dy_j^2 - (\sum n_j dy_j)^2]\}}}$$

مثال رقم (06-06): يمثل الجدول التالي توزيع 40 طالباً حسب علامتيهما في مادتي الرياضيات (X) والفيزياء (Y) (من 20 نقطة لكل مادة):

الرياضيات الفيزياء	4-0	8-4	12-8	16-12	20-16	المجموع
4-0	4	1	0	0	0	5
8-4	2	5	2	0	0	9
12-8	0	2	7	2	0	11
16-12	0	0	2	6	2	10
20-16	0	0	0	1	4	5
المجموع	6	8	11	9	6	40

المطلوب: احسب معامل الارتباط بيرسون بين علامتي الرياضيات والفيزياء.

الحل: نختار الفئة [12-8] كوسط فرضي لكل من المتغيرين ($X_0 = Y_0 = 10$)، وهي مركز الفئة الأكثر تكراراً تقريباً، وطول الفئة $K_x = K_y = 4$ ، فنحصل على الانحرافات المرمزة التالية لمراكز الفئات: 2-، 1-، 0، 1، 2.

نحسب أولاً الأعمدة المساعدة الهامشية لكل متغير:

مراكز فئات الرياضيات	2	6	10	14	18	المجموع
التكرار الهامشي f_i	5	9	11	10	5	40

dx_i	2	1	0	-1	-2	
مراكز فئات الفيزياء	18	14	10	6	2	
التكرار الهامشي f_{ij}	6	9	11	8	6	40
dy_j	2	1	0	-1	-2	

ثم نحسب المجاميع اللازمة:

$$\sum f_i dx_i = 5(-2) + 9(-1) + 11(0) + 10(1) + 5(2) = -10 - 9 + 0 + 10 + 10 = 1$$

$$\sum f_i dx_i^2 = 5(4) + 9(1) + 11(0) + 10(1) + 5(4) = 20 + 9 + 0 + 10 + 20 = 59$$

$$\sum f_j dy_j = 6(-2) + 8(-1) + 11(0) + 9(1) + 6(2) = -12 - 8 + 0 + 9 + 12 = 1$$

$$\sum f_j dy_j^2 = 6(4) + 8(1) + 11(0) + 9(1) + 6(4) = 24 + 8 + 0 + 9 + 24 = 65$$

وأخيراً نحسب مجموع حاصل ضرب التكرار المشترك في الانحرافين المرمّزين، خلية بخلية (نتجاهل الخلايا ذات التكرار المعدوم أو ذات dx أو dy معدومة):

الخلية (رياضيات، فيزياء)	f_{ij}	dx_i	dy_j	$f_{ij} \times dx_i \times dy_j$
[0-4]، [0-4]	4	-2	-2	16
[4-8]، [0-4]	1	-2	-1	2
[0-4]، [4-8]	2	-1	-2	4
[4-8]، [4-8]	5	-1	-1	5
[12-16]، [12-16]	6	1	1	6
[16-20]، [12-16]	2	1	2	4
[12-16]، [16-20]	1	2	1	2

16	2	2	4	[16-20],[16-20]
55	/	/	/	المجموع

$$r = [40 \times 55 - 1 \times 1] / \sqrt{\{[40 \times 59 - 1^2][40 \times 65 - 1^2]\}} = [2200 - 1] /$$

$$\sqrt{\{[2360 - 1][2600 - 1]\}}$$

$$r = 2199 / \sqrt{(2359 \times 2599)} = 2199 / \sqrt{6130441} \approx 2199 / 2476.0 \approx 0.888$$

بما أن معامل الارتباط بيرسون (0.888) موجب وقريب من الواحد الصحيح، فإن هناك ارتباطاً خطياً موجباً قوياً بين علامتي الرياضيات والفيزياء لدى هؤلاء الطلبة: الطالب المتفوق في الرياضيات يميل غالباً إلى التفوق في الفيزياء أيضاً، وهي نتيجة متوقعة نظراً للطبيعة المنطقية والكمية المشتركة بين المادتين.

ثالثاً: معامل الارتباط الرتبي سبيرمان (Spearman's Rank Correlation)

يُستعمل معامل الارتباط الرتبي سبيرمان عندما يتعذر قياس المتغيرين قياساً كمياً دقيقاً، أو عندما تكون البيانات في الأصل عبارة عن رتب أو ترتيب (كترتيب المتسابقين، أو تقييمات وصفية مثل ممتاز/جيد جداً/جيد...)، وذلك بدلاً من معامل الارتباط بيرسون الذي يتطلب بيانات كمية بمستوى قياس فاصلي أو نسبي.

ويُحسب معامل الارتباط الرتبي سبيرمان، ويُرمز له بالرمز (rs)، وفق العلاقة التالية:

$$rs = 1 - [6 \sum d_i^2] / [n(n^2 - 1)]$$

حيث d_i هو الفرق بين رتبة المفردة (i) في المتغير الأول ورتبتها في المتغير الثاني، و f عدد المفردات. ويتراوح مجال تغير هذا المعامل بين -1 و $+1$ كذلك، بنفس تفسير معامل بيرسون من حيث الاتجاه والقوة.

مثال رقم (06-07): قيم 8 نزلاء في فندق معين جودة خدمة الاستقبال (X) ونظافة الغرف (Y)، حسب سلم تقييم من ست درجات: ممتاز (رتبة 1)، جيد جداً (رتبة 2)، جيد (رتبة 3)، مقبول (رتبة 4)، ضعيف (رتبة 5)، ضعيف جداً (رتبة 6). وكانت النتائج كما يلي:

النزيل	1	2	3	4	5	6	7	8
تقييم الاستقبال	ممتاز	جيد	مقبول	ضعيف	جيد جداً	مقبول	جيد	ضعيف جداً
تقييم النظافة	جيد جداً	جيد	ضعيف	ضعيف	ممتاز	جيد جداً	جيد	ضعيف

المطلوب: احسب معامل الارتباط الرتبي سبيرمان بين تقيمي الاستقبال والنظافة.

الحل: نستبدل التقييمات الوصفية برتبها العددية المقابلة، ثم نحسب الفروق ومربعاتها:

النزيل	رتبة الاستقبال	رتبة النظافة	d_i	d_i^2
1	1	2	-1	1
2	3	3	0	0
3	4	5	-1	1
4	5	5	0	0
5	2	1	1	1
6	4	2	2	4
7	3	3	0	0
8	6	5	1	1
المجموع	/	/	/	8

$$r_s = 1 - [6 \times 8] / [8 \times (64 - 1)] = 1 - 48/504 \approx 1 - 0.095 = 0.905$$

بما أن معامل الارتباط الرتبي (0.905) موجب وقريب جداً من الواحد الصحيح، فإن هناك ارتباطاً قوياً جداً بين تقييم النزلاء لجودة الاستقبال وتقييمهم لنظافة الغرف: النزيل الذي يقيم الاستقبال إيجابياً يميل غالباً إلى تقييم النظافة إيجابياً أيضاً، مما يوحي بأن رضا النزلاء ينعكس بشكل عام على مختلف جوانب الخدمة الفندقية معاً.

مثال رقم (06-08): يمثل الجدول التالي الإنفاق الإعلاني الشهري (X) ورقم المبيعات الشهري (Y) لثمانية متاجر تابعة لسلسلة تجارية واحدة (بآلاف الدينار):

المتجر	1	2	3	4	5	6	7	8
الإنفاق الإعلاني	200	180	160	150	140	130	120	100
رقم المبيعات	92	95	70	60	58	78	68	50

المطلوب: احسب معامل الارتباط الرتبي سبيرمان بين الإنفاق الإعلاني ورقم المبيعات، مستعملاً أسلوب الرتب مباشرة.

الحل: نرتب قيم كل متغير تصاعدياً ونعطي الرتبة 1 لأصغر قيمة، ثم نحسب الفروق ومربعاتها:

المتجر	رتبة الإنفاق	رتبة المبيعات	d_i	d_i^2
1	8	7	1	1
2	7	8	-1	1
3	6	5	1	1
4	5	3	2	4
5	4	2	2	4
6	3	6	-3	9
7	2	4	-2	4

0	0	1	1	8
24	/	/	/	المجموع

$$rs = 1 - [6 \times 24] / [8 \times (64 - 1)] = 1 - 144/504 \approx 1 - 0.286 = 0.714$$

بما أن معامل الارتباط الرتبي (0.714) موجب ومتوسط إلى قوي، فإن هناك ارتباطاً إيجابياً ملحوظاً بين الإنفاق الإعلاني ورقم المبيعات، وإن لم يكن تاماً؛ إذ توجد عوامل أخرى غير الإعلان (كالموقع الجغرافي للمتجر، أو جودة الخدمة) تؤثر أيضاً في رقم المبيعات، مما يفسر عدم وصول الارتباط إلى قيمته القصوى.

رابعاً: معامل الاقتران فاي (Coefficient de Contingence Phi)

يُستعمل معامل الاقتران فاي لقياس درجة الارتباط بين متغيرين وصفيين (كيفيين) كل منهما يأخذ قيمتين فقط (متغيران ثنائيان)، ويُعرض عادة في جدول توافقي (Contingency Table) بحجم 2×2. ويُرمز لهذا المعامل بالرمز (φ)، ويُحسب وفق العلاقة التالية، إذا كان الجدول التوافقي على الشكل:

Y = لا	Y = نعم	
b	a	X = نعم
d	c	X = لا

$$\phi = (ad - bc) / \sqrt{[(a + b)(c + d)(a + c)(b + d)]}$$

ويتراوح مجال تغير هذا المعامل بين 1- و 1+، بنفس منطق التفسير المعتاد لمعاملات الارتباط.

مثال رقم (06-09): استطلعت آراء 100 موظف حول حضورهم دورة تدريبية معينة (نعم/لا) ونجاحهم في شهادة اعتماد مهني لاحقة (نعم/لا)، فكانت النتائج كما يلي:

المجموع	لم ينجح في الشهادة	نجح في الشهادة	
50	10	40	حضر الدورة

50	35	15	لم يحضر الدورة
100	45	55	المجموع

المطلوب: احسب معامل الاقتران فاي بين حضور الدورة التدريبية والنجاح في شهادة الاعتماد، وعلّق على النتيجة.

الحل:

$$\varphi = (40 \times 35 - 10 \times 15) / \sqrt{(50 \times 50 \times 55 \times 45)} = (1400 - 150) / \sqrt{6187500}$$

$$\varphi = 1250 / 2487.5 \approx 0.503$$

بما أن معامل الاقتران (0.503) موجب ومتوسط القوة، فإن هناك علاقة إيجابية بين حضور الدورة التدريبية والنجاح في شهادة الاعتماد: الموظفون الذين حضروا الدورة كانت نسبة نجاحهم في الشهادة أعلى نسبياً (80%) مقارنة بمن لم يحضروها (30%)، مما يدعم فرضية فعالية هذه الدورة التدريبية في تحسين فرص النجاح.

مثال رقم (10-06): استطلعت آراء 100 مستهلك حول امتلاكهم لهاتف ذكي (نعم/لا) واستعمالهم لتطبيق دفع إلكتروني (نعم/لا)، فكانت النتائج كما يلي:

المجموع	لا يستعمله	يستعمل الدفع الإلكتروني	
70	15	55	يمتلك هاتفاً ذكياً
30	20	10	لا يمتلك هاتفاً ذكياً
100	35	65	المجموع

المطلوب: احسب معامل الاقتران فاي بين امتلاك الهاتف الذكي واستعمال تطبيق الدفع الإلكتروني.

الحل:

$$\varphi = (55 \times 20 - 15 \times 10) / \sqrt{(70 \times 30 \times 65 \times 35)} = (1100 - 150) / \sqrt{4778250}$$

$$\phi = 950 / 2185.9 \approx 0.435$$

بما أن معامل الاقتران (0.435) موجب ومتوسط القوة، فإن امتلاك الهاتف الذكي مرتبط إيجابياً باستعمال تطبيقات الدفع الإلكتروني، وهي نتيجة منطقية تماماً بما أن أغلب تطبيقات الدفع الإلكتروني الحديثة تتطلب أصلاً امتلاك هاتف ذكي لتثبيتها واستعمالها.

خامساً: معامل التوافق (Coefficient de Contingence)

يُستعمل معامل التوافق لقياس درجة الارتباط بين متغيرين وصفيين (كيفيين) عندما يأخذ أحدهما أو كلاهما أكثر من قيمتين (جدول توافقي أكبر من 2×2)، ويعتمد حسابه على قيمة إحصائية مربع كاي (χ^2) التي تقيس الفرق بين التكرارات المشاهدة والتكرارات المتوقعة نظرياً في حال استقلالية المتغيرين. ويُحسب معامل التوافق، ويُرمز له بالرمز (C)، وفق العلاقة التالية:

$$C = \sqrt{[\chi^2 / (\chi^2 + N)]}$$

حيث يُحسب χ^2 وفق العلاقة المعروفة:

$$\chi^2 = \sum \sum [(n_{ij} - e_{ij})^2 / e_{ij}]$$

حيث f_{ij} التكرار المشاهد للخلية (i, j)، و e_{ij} التكرار المتوقع لها في حال استقلالية المتغيرين، ويُحسب بالعلاقة: $e_{ij} = (\text{مجموع سطر } i \times \text{مجموع عمود } j) / F$.

ويتراوح معامل التوافق بين 0 (استقلالية تامة، أي انعدام العلاقة) وقيمة قصوى تقترب من 1 دون بلوغها أبداً (كلما زاد عدد فئات الجدول)، وكلما اقتربت قيمته من الواحد دلّ ذلك على ارتباط أقوى بين المتغيرين.

مثال رقم (06-11): استطلعت آراء 200 فرد حول المستوى التعليمي (ابتدائي، ثانوي، جامعي) ومستوى الدخل (منخفض، متوسط، مرتفع)، فكانت النتائج كما يلي:

المجموع	مرتفع	متوسط	منخفض	المستوى التعليمي \ الدخل
50	5	15	30	ابتدائي
70	15	35	20	ثانوي
80	45	25	10	جامعي
200	65	75	60	المجموع

المطلوب: احسب معامل التوافق بين المستوى التعليمي ومستوى الدخل، وعلّق على النتيجة.

الحل: نحسب أولاً التكرارات المتوقعة لكل خلية e_{ij} = مجموع السطر $\times$ مجموع العمود / 200):

مرتفع	متوسط	منخفض	المستوى التعليمي \ الدخل
16.25	18.75	15	ابتدائي
22.75	26.25	21	ثانوي
26	30	24	جامعي

ثم نحسب قيمة χ^2 بجمع $[(f_{ij}-e_{ij})^2/e_{ij}]$ لجميع الخلايا التسع:

$$\chi^2 \approx 15 + 0.729 + 8.019 + 0.048 + 3.512 + 2.622 + 8.167 + 0.833 + 13.885 \approx 52.03$$

$$C = \sqrt{[52.03 / (52.03 + 200)]} = \sqrt{(52.03/252.03)} = \sqrt{0.2065} \approx 0.454$$

بما أن معامل التوافق (0.454) متوسط القوة وأكبر بوضوح من الصفر، فإن هناك ارتباطاً إيجابياً ملموساً بين المستوى التعليمي ومستوى الدخل: كلما ارتفع المستوى التعليمي، ارتفعت احتمالية الانتماء إلى فئة الدخل المرتفع (نلاحظ ذلك مباشرة من الجدول: 45 من أصل 80 حاملي مستوى جامعي ينتمون لفئة الدخل المرتفع، مقابل 5 فقط من أصل 50 من ذوي المستوى الابتدائي).

سادساً: الانحدار الخطي البسيط (Régression Linéaire Simple)

بعد أن اكتفينا في الأقسام السابقة بالكشف عن وجود علاقة إحصائية بين متغيرين وقياس قوتها واتجاهها بواسطة معاملات الارتباط، ينتقل بنا التحليل الإحصائي خطوة أبعد: فبمجرد التأكد من وجود ارتباط قوي بين متغيرين، يصبح بإمكاننا صياغة نموذج رياضي يسمح بالتنبؤ بقيم أحد المتغيرين (المتغير التابع Y) انطلاقاً من قيم المتغير الآخر (المتغير المستقل أو المفسر X). وهذا بالضبط هو موضوع الانحدار الخطي البسيط.

(a) مفهوم الانحدار الخطي ومعادلة خط الانحدار

يهدف الانحدار الخطي البسيط إلى تمثيل العلاقة بين متغيرين كميين X و Y بمعادلة خط مستقيم يُقَرَّب العلاقة الحقيقية بينهما بأفضل شكل ممكن، بحيث يمكن استعمال هذه المعادلة لاحقاً للتنبؤ بقيمة Y عند معرفة قيمة X . وتكتب معادلة خط الانحدار (المستقيم المقدّر) على الشكل التالي:

$$\hat{Y} = a + bX$$

حيث $\hat{Y}$ القيمة المقدّرة (المتنبأ بها) للمتغير التابع، a الحد الثابت (تقاطع المستقيم مع محور الترتيب)، و b ميل خط الانحدار (معامل الانحدار)، الذي يقيس مقدار التغير في Y الناتج عن زيادة X بوحدة واحدة.

ولإيجاد أفضل مستقيم ممثل لسحابة النقاط (X_i, Y_i) ، يُعتمد على طريقة المربعات الصغرى (Méthode des Moindres Carrés)، التي تقوم على اختيار قيمتي a و b اللتين تجعلان مجموع مربعات الانحرافات بين القيم الحقيقية Y_i والقيم المقدّرة $\hat{Y}_i$ أصغر ما يمكن. وتؤدي هذه الطريقة إلى العلاقتين التاليتين لحساب معلمتي النموذج:

$$b = [n \sum x_i y_i - \sum x_i \sum y_i] / [n \sum x_i^2 - (\sum x_i)^2]$$

$$a = \bar{Y} - b\bar{X}$$

حيث X و $\bar{Y}$ الوسطان الحسابيان للمتغيرين X و Y على التوالي.

شكل العلاقة بين X و Y حسب إشارة معامل الانحدار

تحدد طبيعة العلاقة بين المتغيرين من خلال إشارة معامل الانحدار (b):

- إذا كانت $b > 0$: العلاقة طردية (موجبة)، أي أن Y يزداد كلما ازداد X ، ويكون خط الانحدار صاعداً.
- إذا كانت $b < 0$: العلاقة عكسية (سالبة)، أي أن Y يتناقص كلما ازداد X ، ويكون خط الانحدار نازلاً.
- إذا كانت $b = 0$: انعدام العلاقة الخطية بين المتغيرين، ويكون خط الانحدار أفقياً (موازيًا لمحور الفواصل).

مثال تطبيقي شامل على الانحدار الخطي البسيط

مثال رقم (06-12): سجّلت إدارة إحدى المؤسسات التربوية عدد ساعات المذاكرة اليومية (X) والعلامة المتحصّل عليها في اختبار قصير من 20 نقطة (Y) لعينة من 10 طلبة:

الطالب	1	2	3	4	5	6	7	8	9	10
ساعات المذاكرة اليومية (X)	1	2	2	3	4	4	5	6	6	7
العلامة من 20 (Y)	8	9	11	10	13	14	15	16	17	19

المطلوب: أوجد معادلة خط الانحدار الخطي لـ Y بدلالة X . - تتبأ بعلامة طالب يذاكر 9 ساعات يومياً.

الحل: نحسب أولاً الأعمدة المساعدة:

Σx_i^2	$\Sigma x_i y_i$	Σy_i	Σx_i	الإحصائية القيمة
196	592	132	40	

$$\bar{X} = 40/10 = 4$$

$$\bar{Y} = 132/10 = 13.2$$

(1) حساب معامل الانحدار b والحد الثابت a:

$$b = [10 \times 592 - 40 \times 132] / [10 \times 196 - 40^2] = [5920 - 5280] / [1960 -$$

$$1600] = 640 / 360 \approx 1.778$$

$$a = 13.2 - 1.778 \times 4 = 13.2 - 7.111 \approx 6.089$$

ومنه تصبح معادلة خط الانحدار المقدرة:

$$\hat{Y} = 6.089 + 1.778X$$

ونلاحظ أن معامل الانحدار موجب ($b \approx 1.778$)، مما يدل على وجود علاقة طردية بين عدد ساعات المذاكرة والعلامة المتحصّل عليها: فكل ساعة إضافية من المذاكرة اليومية تقابلها زيادة متوقعة قدرها 1.778 نقطة تقريباً في علامة الاختبار، بافتراض ثبات العوامل الأخرى.

(2) التنبؤ بعلامة طالب يذاكر 9 ساعات يومياً: بتعويض $X = 9$ في معادلة خط الانحدار:

$$\hat{Y} = 6.089 + 1.778 \times 9 = 6.089 + 16.0 \approx 22.09$$

ملاحظة مهمة: القيمة المقدرة (22.09) تتجاوز السقف الأقصى للعلامة (20 نقطة)، وهو ما يُعرف بمشكلة "التنبؤ خارج مجال المعاينة" (Extrapolation)؛ إذ إن قيمة $X=9$ تقع خارج المجال الذي بُني

عليه النموذج (1 إلى 7 ساعات)، ولذا يجب التعامل مع مثل هذه التنبؤات بحذر شديد، والاكتفاء عملياً بالتنبؤ ضمن المجال الذي استُمدت منه بيانات العينة أو بالقرب منه.

الطريقة البيانية لتقدير معلمتي نموذج الانحدار

يمكن كذلك تقدير معلمتي النموذج (a و b) تقديراً تقريبياً بيانياً، دون اللجوء إلى الحساب الجبري، وذلك برسم سحابة النقاط (X_i, Y_i) على معلم متعامد، ثم رسم الخط المستقيم الذي يمر وسط هذه النقاط بأفضل شكل ممكن (بحيث تتوزع النقاط فوقه وتحتة بالتساوي تقريباً). وعندئذ:

يُمثل الحد الثابت (a) نقطة تقاطع هذا المستقيم مع محور الترتيب (Y) عند $X = 0$.

يُمثل معامل الانحدار (b) ميل هذا المستقيم، ويمكن تقديره بيانياً بقسمة الفرق بين ترانتيب نقطتين على هذا المستقيم على الفرق بين فاصلتيهما.

وتبقى هذه الطريقة البيانية تقريبية وأقل دقة من طريقة المربعات الصغرى المشروحة أعلاه، إلا أنها مفيدة جداً كخطوة استكشافية أولى، إذ تسمح بالتحقق البصري السريع من معقولية النموذج المقدر جبرياً، ومن مدى ملاءمة الافتراض الخطي أصلاً لوصف العلاقة بين المتغيرين (فقد تكون سحابة النقاط منحنية الشكل مثلاً، مما يستدعي نموذجاً غير خطي بدلاً من الانحدار الخطي البسيط).

قياس قوة الانحدار ودراسة صلاحية النموذج

لا يكفي إيجاد معادلة خط الانحدار وحدها، بل ينبغي أيضاً تقييم مدى جودة هذا النموذج في تفسير التغيرات الحاصلة في المتغير التابع، وذلك بحساب معامل الارتباط ومعامل التحديد، وكذا الأخطاء المعيارية لمعلمتي النموذج.

أ. **معامل الارتباط ومعامل التحديد:** يُستعمل معامل الارتباط بيرسون (r)، الذي سبق شرح طريقة حسابه في القسم الثاني من هذا الفصل، لقياس قوة العلاقة الخطية بين X و Y ؛ أما مربع هذا المعامل (r^2)، فيُعرف بمعامل التحديد (Coefficient de détermination)،

ويقيس نسبة التغيرات في المتغير التابع Y التي يُفسّرها النموذج (أي التي تُعزى إلى تغيرات المتغير المستقل X).

بتطبيق نفس معطيات المثال (06-12) على علاقة معامل الارتباط بيرسون:

$$r = [10 \times 592 - 40 \times 132] / \sqrt{\{[10 \times 196 - 40^2][10 \times 1862 - 132^2]\}} = 640 / \sqrt{(360 \times 1796)}$$

$$r = 640 / \sqrt{646560} \approx 640 / 804.1 \approx 0.796$$

بما أن معامل الارتباط (0.796) موجب وقوي نسبياً، فإن النموذج جيد إلى حد بعيد؛ ومعامل التحديد المقابل له يساوي:

$$r^2 \approx 0.796^2 \approx 0.634$$

أي أن نموذج الانحدار الخطي يُفسّر حوالي 63.4% من التغيرات الحاصلة في العلامة المتحصّل عليها، بينما تُعزى نسبة 36.6% المتبقية إلى عوامل أخرى غير مدرجة في النموذج (كالقدرات الذهنية الفردية للطالب، أو جودة أسلوب المذاكرة، أو ظروف الامتحان).

ب. **الأخطاء المعيارية لمعلمتي النموذج:** نظراً لأن معلمتي النموذج (a و b) مقدّرتان من عينة محدودة الحجم، فإنهما عرضة لهامش خطأ معيّن؛ ويُقاس هذا الهامش بالخطأ المعياري لكل معلمة، والذي يُستعمل لاحقاً لبناء فترات ثقة أو اختبار معنوية المعلمة إحصائياً. ويُحسب الخطأ المعياري لتقدير النموذج (المتعلق بالبواقي) أولاً بالعلاقة التالية:

$$SU = \sqrt{\left[\sum (y_i - \hat{y}_i)^2 / (n - 2) \right]}$$

ثم يُحسب الخطأ المعياري لمعامل الانحدار (b) بالعلاقة:

$$Sb = SU / \sqrt{\left[\sum x_i^2 - n\bar{X}^2 \right]}$$

مثال (تكملة المثال 06-12): احسب الخطأ المعياري لتقدير النموذج والخطأ المعياري لمعامل الانحدار b.

الحل: نحسب أولاً القيم المقدرة ($\hat{Y}_i$) والبقايا ($Y_i - \hat{Y}_i$) لكل طالب:

الطالب	1	2	3	4	5	6	7	8	9	10
Y_i	8	9	11	10	13	14	15	16	17	19
$\hat{Y}_i$	7.87	9.64	9.64	11.42	13.20	13.20	14.98	16.76	16.76	18.53
$Y_i - \hat{Y}_i$	0.13	-0.64	1.36	-1.42	-0.20	0.80	0.02	-0.76	0.24	0.47

ومجموع مربعات هذه البقايا يساوي تقريباً 5.82، ومنه:

$$SU = \sqrt{[5.82 / (10-2)]} = \sqrt{0.7275} \approx 0.853$$

$$Sb = 0.853 / \sqrt{(196 - 10 \times 4^2)} = 0.853 / \sqrt{(196-160)} = 0.853 / \sqrt{36} = 0.853 / 6 \approx 0.142$$

ونلاحظ أن الخطأ المعياري لمعامل الانحدار (0.142) صغير نسبياً مقارنة بقيمة المعامل نفسه (1.778)، مما يعزز الثقة في دقة تقدير هذا المعامل ويوحي بمعنويته الإحصائية؛ ويُستعمل هذا الخطأ المعياري لاحقاً في مقرر الإحصاء الاستدلالي لبناء فترة ثقة حول b أو لاختبار الفرضية القائلة بانعدام العلاقة الخطية بين X و Y (أي اختبار الفرضية $b = 0$).

تمارين الفصل السادس

التمرين الأول:

يمثل الجدول التالي عدد سنوات الخبرة (X) والإنتاجية اليومية بالوحدة (Y) لعينة من 8 عمال في ورشة صناعية:

العامل	1	2	3	4	5	6	7	8
سنوات الخبرة (X)	1	2	3	4	5	6	7	8
الإنتاجية اليومية (Y)	20	24	23	28	30	33	35	40

المطلوب: احسب معامل الارتباط بيرسون بين سنوات الخبرة والإنتاجية اليومية، وعلّق على النتيجة.

الحل: نحسب الأعمدة المساعدة:

$$\Sigma y^2=7043, \Sigma x^2=204, \Sigma xy=1145, \Sigma y=233, \Sigma x=36, f=8$$

$$r = [8 \times 1145 - 36 \times 233] / \sqrt{[8 \times 204 - 36^2][8 \times 7043 - 233^2]}$$

$$r = [9160 - 8388] / \sqrt{[1632 - 1296][56344 - 54289]} = 772 / \sqrt{(336 \times 2055)}$$

$$r = 772 / \sqrt{690480} \approx 772 / 831.0 \approx 0.929$$

بما أن معامل الارتباط (0.929) موجب وقوي جداً، فإن الإنتاجية اليومية ترتفع بشكل شبه منتظم مع ازدياد سنوات الخبرة، وهو ما يبرر اعتماد سنوات الخبرة كمعيار مهم عند تقييم أداء العمال أو تحديد الأجور.

التمرين الثاني:

رتب 8 مزارعين حسب كمية استعمال الأسمدة الكيميائية (X) وحسب المردود الزراعي المحصل عليه (Y)، فكانت النتائج كما يلي (الرتبة 1 تمثل أعلى قيمة):

المزارع	1	2	3	4	5	6	7	8
رتبة استعمال الأسمدة	1	3	2	5	4	7	6	8
رتبة المردود الزراعي	2	1	3	4	6	5	8	7

المطلوب: احسب معامل الارتباط الرتبي سبيرمان بين استعمال الأسمدة والمردود الزراعي، وعلق على النتيجة.

الحل:

المزارع	1	2	3	4	5	6	7	8	المجموع
d_i	-1	2	-1	1	-2	2	-2	1	/
d_i^2	1	4	1	1	4	4	4	1	20

$$rs = 1 - [6 \times 20] / [8 \times (64 - 1)] = 1 - 120/504 \approx 1 - 0.238 = 0.762$$

بما أن معامل الارتباط الرتبي (0.762) موجب وقوي نسبياً، فإن هناك علاقة إيجابية واضحة بين كمية استعمال الأسمدة الكيميائية والمردود الزراعي المحصل عليه، وإن كانت دون الارتباط التام، مما يوحي بتدخل عوامل زراعية أخرى (كنوعية التربة أو كمية الري) في تحديد المردود النهائي.

التمرين الثالث:

استُطلعت آراء 120 مريضاً حول التزامهم بالنظام الغذائي الموصوف (نعم/لا) وتحسّن حالتهم الصحية بعد ثلاثة أشهر (نعم/لا)، فكانت النتائج كما يلي:

المجموع	لم تتحسن حالته	تحسنت حالته	
60	12	48	التزم بالنظام الغذائي
60	42	18	لم يلتزم بالنظام الغذائي
120	54	66	المجموع

المطلوب: احسب معامل الاقتران فاي بين الالتزام بالنظام الغذائي وتحسّن الحالة الصحية، وعلّق على النتيجة من منظور صحي.

الحل:

$$\varphi = (48 \times 42 - 12 \times 18) / \sqrt{(60 \times 60 \times 66 \times 54)} = (2016 - 216) / \sqrt{12830400}$$

$$\varphi = 1800 / 3582.0 \approx 0.503$$

بما أن معامل الاقتران (0.503) موجب ومتوسط إلى قوي، فإن هناك علاقة إيجابية واضحة بين الالتزام بالنظام الغذائي الموصوف وتحسّن الحالة الصحية للمرضى (80% من الملتزمين تحسنت حالتهم مقابل 30% فقط من غير الملتزمين)، وهي نتيجة تدعم أهمية برامج التوعية الصحية بضرورة الالتزام بالتوصيات الطبية الغذائية.

التمرين الرابع:

يمثل الجدول التالي توزيع 150 موظفاً حسب المستوى الإداري (تنفيذي، إشرافي، إداري عليا) ومستوى الرضا الوظيفي (منخفض، متوسط، مرتفع):

المجموع	مرتفع	متوسط	منخفض	المستوى الإداري \ الرضا
50	5	20	25	تنفيذي
50	15	25	10	إشرافي
50	30	15	5	إدارة عليا
150	50	60	40	المجموع

المطلوب: احسب معامل التوافق بين المستوى الإداري ومستوى الرضا الوظيفي، وعلق على النتيجة من منظور إداري.

الحل: نحسب أولاً التكرارات المتوقعة لكل خلية:

مرتفع	متوسط	منخفض	المستوى الإداري \ الرضا
16.67	20	13.33	تنفيذي
16.67	20	13.33	إشرافي
16.67	20	13.33	إدارة عليا

$$\chi^2 \approx 9.11 + 0 + 8.17 + 0.83 + 1.25 + 0.27 + 5.21 + 1.25 + 11.11 \approx 37.20$$

$$C = \sqrt{[37.20 / (37.20 + 150)]} = \sqrt{(37.20/187.20)} = \sqrt{0.1987} \approx 0.446$$

بما أن معامل التوافق (0.446) متوسط القوة، فإن هناك ارتباطاً إيجابياً ملموساً بين المستوى الإداري ومستوى الرضا الوظيفي: كلما ارتفع المستوى الإداري، ارتفعت نسبة الرضا الوظيفي المرتفع (30 من أصل 50 في الإدارة العليا مقابل 5 فقط من أصل 50 في المستوى التنفيذي)، وهي نتيجة قد تدفع إدارة الموارد البشرية إلى دراسة أسباب انخفاض الرضا لدى الفئات التنفيذية واتخاذ إجراءات تحفيزية مناسبة لها.

التمرين الخامس:

سُجِّلَت مصلحة الصيانة في مصنع معين مبلغ الإنفاق الشهري على الصيانة الوقائية للآلات (X)،
بآلاف الدنانير) وعدد الأعطال المسجلة في نفس الشهر (Y)، لعشرة أشهر متتالية:

الشهر	1	2	3	4	5	6	7	8	9	10
الإنفاق على الصيانة (X)	2	3	4	5	6	7	8	9	10	11
عدد الأعطال (Y)	14	13	11	10	9	8	7	6	5	4

المطلوب: أوجد معادلة خط الانحدار لعدد الأعطال بدلالة الإنفاق على الصيانة، واحسب معامل الارتباط، وعلّق على النتيجة من منظور إداري.

الحل:

$$\Sigma y^2=857, \Sigma x^2=505, \Sigma xy=475, \Sigma y=87, \Sigma x=65, f=10$$

$$\bar{Y} = 8.7, \bar{X} = 6.5$$

$$b = [10 \times 475 - 65 \times 87] / [10 \times 505 - 65^2] = [4750 - 5655] / [5050 - 4225] = -905 / 825 \approx -1.097$$

$$a = 8.7 - (-1.097) \times 6.5 \approx 8.7 + 7.13 \approx 15.83$$

$$\hat{Y} = 15.83 - 1.097X$$

$$r = -905 / \sqrt{[5050-4225][8570-7569]} = -905 / \sqrt{(825 \times 1001)} \approx -905 / 908.6 \approx -0.996$$

بما أن معامل الانحدار سالب ($b \approx -1.097$) ومعامل الارتباط سالب وقريب جداً من -1 ($r \approx -0.996$)، فإن هناك علاقة عكسية قوية جداً بين الإنفاق على الصيانة الوقائية وعدد الأعطال: كل زيادة بـ 1000 دينار في الإنفاق على الصيانة تقابلها انخفاض متوقع قدره 1.097 عطل تقريباً في الشهر. وهذه نتيجة مهمة تدعم استمرار الاستثمار في برامج الصيانة الوقائية لتقليل الأعطال وتكاليف التوقف غير المخطط له.

التمرين السادس:

سجلت وكالة عقارية مساحة عشر قطع أرضية معروضة للبيع (X ، بمئات الأمتار المربعة) وسعر بيعها الفعلي (Y ، بملايين الدنانير):

القطعة	1	2	3	4	5	6	7	8	9	10
المساحة (X)	2	3	3	4	5	6	6	7	8	9
سعر البيع (Y)	3	4	5	5	7	8	9	9	11	12

المطلوب: أوجد معادلة خط الانحدار لسعر البيع بدلالة المساحة، وتنبأ بسعر قطعة أرض مساحتها 500 م²، واحسب معامل التحديد وفسر معناه.

الحل:

$$\Sigma y^2=615, \Sigma x^2=329, \Sigma xy=449, \Sigma y=73, \Sigma x=53, f=10$$

$$\bar{Y} = 7.3, \bar{X} = 5.3$$

$$b = [10 \times 449 - 53 \times 73] / [10 \times 329 - 53^2] = [4490 - 3869] / [3290 - 2809] = 621 / 481 \approx 1.291$$

$$a = 7.3 - 1.291 \times 5.3 \approx 7.3 - 6.84 \approx 0.457$$

$$\hat{Y} = 0.457 + 1.291X$$

التنبؤ بسعر قطعة مساحتها 500 م² (أي $X = 5$):

$$\hat{Y} = 0.457 + 1.291 \times 5 \approx 0.457 + 6.455 \approx 6.91 \text{ مليون دينار}$$

$$r = 621 / \sqrt{\{[3290 - 2809][6150 - 53 \times 29]\}} = 621 / \sqrt{(481 \times 821)} \approx 621 / 628.4 \approx 0.988$$

$$r^2 \approx 0.988^2 \approx 0.976$$

بما أن معامل التحديد ($r^2 \approx 0.976$) قريب جداً من الواحد الصحيح، فإن نموذج الانحدار الخطي يُفسّر حوالي 97.6% من التغيرات الحاصلة في سعر البيع، بينما لا تُعزى سوى نسبة ضئيلة جداً (2.4%) إلى عوامل أخرى غير المساحة (كالموقع الجغرافي الدقيق أو مدى قربها من الخدمات)؛ وهو ما يجعل المساحة وحدها مؤشراً تنبؤياً قوياً جداً لسعر القطع الأرضية في هذه العينة.

الفصل السابع

الأرقام القياسية

الفصل السابع: الأرقام القياسية

مفهوم الأرقام القياسية

تُعد الأرقام القياسية من أهم الأدوات الإحصائية المستخدمة لقياس التغير الحاصل في الظواهر عبر الزمن، فقد يرغب الباحث مثلاً في قياس التغير في أسعار مجموعة من السلع مجتمعة، أو قياس التغير المتوسط في الكميات المنتجة أو المستهلكة من عدة سلع في آن واحد¹.

فالرقم القياسي أداة إحصائية تهدف إلى قياس وتصوير التغير النسبي الذي يطرأ على ظاهرة معينة من فترة زمنية إلى أخرى، أو من مكان إلى آخر. وبعبارة أخرى، هو مؤشر إحصائي يُستعمل لقياس تطور سعر مادة أو كمية مادة واحدة أو عدة مواد بين زمنين مختلفين، أو بين مجموعتين، أو بين مكانين مختلفين.

ومن الأمثلة الشائعة على الأرقام القياسية: الرقم القياسي لأسعار الاستهلاك، الذي يلخص تطور أسعار جميع السلع والخدمات المستهلكة ويُستعان به لدراسة الحالة العامة للاقتصاد الوطني ولحساب معدل التضخم؛ وكذلك الرقم القياسي لتكاليف المعيشة ومستوى المعيشة، والرقم القياسي للأجور، والرقم القياسي لأسعار الجملة، والرقم القياسي للإنتاج الصناعي، وغيرها من التطبيقات الواسعة في الاقتصاد والإدارة.

ولإنشاء رقم قياسي سليم منهجياً، ينبغي مراعاة النقاط التالية:

- تحديد الغرض من الرقم القياسي بوضوح قبل الشروع في تكوينه.
- تحديد شمولية (مجال) الرقم القياسي، أي تحديد السلع أو الظواهر التي تدخل في نطاقه وتلك التي تُستبعد منه.
- اختيار الفترة أو المكان الذي يُعتمد أساساً لقياس التغير (فترة الأساس).
- اختيار المصادر التي تُجمع منها البيانات اللازمة، مع التأكد من موثوقيتها.
- اختيار الصيغة الرياضية المناسبة لحساب الرقم القياسي.

¹ تيلولت سامية، مبادئ في الإحصاء، دار الحديث للكتاب، الطبعة الثانية، ص ص 186-207، الجزائر، 2009.

- ترجيح السلع بأوزان مناسبة تعكس أهميتها النسبية، إن اقتضى الأمر ذلك.

وبعد استيفاء هذه الخطوات، يجب تحديد فترة زمنية معينة تُعتبر أساساً تُقارن بها بقية الفترات؛ إذ نختار قبل كل شيء، من بين السلسلة الزمنية المدروسة، سنة أو فترة تُسمى "فترة الأساس"، بينما تمثل بقية السنوات أو الفترات "فترات المقارنة" أو "الفترات الحالية".

وتُصنّف الأرقام القياسية إلى ثلاثة أنواع رئيسية: الأرقام القياسية البسيطة، والأرقام القياسية التجميعية، والأرقام القياسية الترجيحية. وسنتناول فيما يلي هذه الأنواع الثلاثة بالتفصيل، مطبقة على الأسعار والكميات وغيرها من الظواهر القابلة للقياس.

أولاً: الرقم القياسي البسيط (I)

يقيس الرقم القياسي البسيط تطور سعر أو كمية مادة واحدة فقط بين فترتين زمنيتين مختلفتين، أو بين مكانين مختلفين، ويُحسب كنسبة بين سعر أو كمية الفترة الحالية (أو فترة المقارنة) وسعر أو كمية فترة الأساس.

يُرمز للفترة الحالية أو فترة المقارنة بالرمز (1)، ويُرمز لفترة الأساس بالرمز (0).

ويُحسب الرقم القياسي البسيط وفق العلاقة العامة التالية:

$$I = (X_1/X_0) * 100$$

حيث X_1 قيمة المتغير (سعراً كان أو كمية) في فترة المقارنة، و X_0 قيمته في فترة الأساس.

الرقم القياسي البسيط للأسعار

إذا كانت لدينا مجموعة من السلع $a, b, c, \dots, z$ ، فإن الرقم القياسي البسيط لسعر كل سلعة يُحسب على حدة وفق نفس العلاقة العامة أعلاه، بتعويض X بالسعر P :

$$IP = (P_1/P_0) * 100$$

مثال رقم (07-01): تمثل البيانات التالية أسعار خمس مواد أساسية لمواد البناء في سنتي 2015 و2023 (بالوحدة النقدية):

المادة	إسمنت	حديد تسليح	رمل	طوب	خشب
السعر سنة 2015 (P_0)	800	500	300	200	1000
السعر سنة 2023 (P_1)	1200	1150	285	206	1000

المطلوب: حساب الرقم القياسي البسيط لأسعار هذه المواد.

الحل: نطبق العلاقة $IP = (P_1/P_0) \times 100$ لكل مادة على حدة:

الإسمنت: $I = (1200/800) \times 100 = 150$ → ارتفاع في السعر بنسبة 50%

حديد التسليح: $I = (1150/500) \times 100 = 230$ → ارتفاع في السعر بنسبة 130%

الرمل: $I = (285/300) \times 100 = 95$ → انخفاض في السعر بنسبة 5% (لأن الرقم القياسي أقل من 100)

الطوب: $I = (206/200) \times 100 = 103$ → ارتفاع طفيف في السعر بنسبة 3%

الخشب: $I = (1000/1000) \times 100 = 100$ → ثبات تام في السعر (لا زيادة ولا نقصان)

ويمكن تمييز ثلاث حالات لقيم الرقم القياسي البسيط:

الحالة الأولى: ثبات في السعر، وفيها يكون الرقم القياسي مساوياً لـ 100 تماماً، كما هو الحال بالنسبة للخشب.

الحالة الثانية: انخفاض في السعر، وفيها تكون قيمة الرقم القياسي أقل من 100، كما هو الحال بالنسبة للرمل، إذ يُحسب مقدار الانخفاض بطرح القيمة المحصلة من 100.

الحالة الثالثة: ارتفاع في السعر، وفيها تكون قيمة الرقم القياسي أكبر من 100، كما هو الحال بالنسبة للإسمنت وحديد التسليح والطوب، ويُحسب مقدار الارتفاع بطرح 100 من القيمة المحصلة.

الرقم القياسي البسيط للكميات

يُحسب الرقم القياسي البسيط للكميات بنفس المنطق، بتعويض X بالكمية Q :

$$IQ = (Q_1/Q_0) * 100$$

مثال رقم (07-02): يبين الجدول التالي أسعار وكميات إنتاج ثلاثة محاصيل زراعية خلال موسمين

متتاليين:

المحصول	الكمية Q_0	الكمية Q_1	السعر P_0	السعر P_1
قمح	100	150	25	20
شعير	200	160	18	27
ذرة	50	155	30	51.6

المطلوب:

(1) حساب الرقم القياسي البسيط للكميات للموسم الثاني مقارنة بالموسم الأول.

(2) حساب الرقم القياسي البسيط للأسعار لنفس الفترة.

الحل:

(1) الرقم القياسي البسيط للكميات:

القمح: $IQ = (150/100) \times 100 = 150$ → ارتفاع في الكمية المنتجة بنسبة 50%

الشعير: $IQ = (160/200) \times 100 = 80$ → انخفاض في الكمية المنتجة بنسبة 20%

الذرة: $IQ = (155/50) \times 100 = 310$ → ارتفاع كبير في الكمية المنتجة بنسبة 210%

2) الرقم القياسي البسيط للأسعار:

القمح: $IP = (20/25) \times 100 = 80$ → انخفاض في السعر بنسبة 20%

الشعير: $IP = (27/18) \times 100 = 150$ → ارتفاع في السعر بنسبة 50%

الذرة: $IP = (51.6/30) \times 100 = 172$ → ارتفاع في السعر بنسبة 72%

ونلاحظ من هذا المثال أن الكمية والسعر قد يتحركان في اتجاهين متعاكسين لنفس المحصول (كما هو حال القمح: زيادة في الكمية مع انخفاض في السعر)، وهو أمر شائع اقتصادياً يعكس قانون العرض والطلب.

مثال رقم (07-03): سُجِّلَت أسعار أسطوانة غاز البوتان المنزلية خلال ثماني سنوات متتالية كما يلي (بالوحدة النقدية):

السنة	2016	2017	2018	2019	2020	2021	2022	2023
السعر	72	95	110	130	130	158	210	275

المطلوب: احسب الأرقام القياسية لهذه السلسلة وفسر النتائج، علماً أن سنة الأساس هي سنة 2019.

الحل: نحسب الرقم القياسي لكل سنة بالنسبة لسعر سنة الأساس (130):

السنة	2016	2017	2018	2019	2020	2021	2022	2023
السعر	72	95	110	130	130	158	210	275
الرقم القياسي	55.4	73.1	84.6	100	100	121.5	161.5	211.5

نلاحظ أن الرقم القياسي لسنة الأساس (2019) يساوي 100 دائماً، لأن سعر سنة الأساس يُقَارَن بنفسه؛ أما السنوات السابقة لها فتحمل أرقاماً قياسية أقل من 100 (تعكس أسعاراً أدنى)، والسنوات اللاحقة

لها فتحمل أرقاماً قياسية أعلى من 100 (تعكس أسعاراً أعلى)، مما يوضح ارتفاعاً مستمراً وحاداً في سعر هذه المادة عبر الفترة المدروسة، خاصة بعد سنة 2020.

ثانياً: الرقم القياسي التجميعي (ΣI)

نظراً لكثرة العمليات الحسابية وتعقيدها عند استعمال الرقم القياسي البسيط لكل سلعة على حدة، خاصة إذا كان عدد السلع كبيراً، فإنه يمكن اللجوء إلى تجميع أسعار (أو كميات) جميع السلع في فترة المقارنة من جهة، وفي فترة الأساس من جهة أخرى، ثم حساب النسبة بين المجموعين وضربها في 100. وتُعطى العلاقة الإحصائية للرقم القياسي التجميعي للأسعار، بافتراض أن f هي عدد السلع أو المواد الداخلة في التركيب، كما يلي:

$$I \sum P = \left(\sum P_1 / \sum P_0 \right) * 100$$

وللكميات بنفس المنطق:

$$I \sum Q = \left(\sum Q_1 / \sum Q_0 \right) * 100$$

مثال رقم (07-04): يبين الجدول التالي أسعار ثلاث مواد استهلاكية (الشاي، البن، السكر) خلال أربع سنوات متتالية (بالوحدة النقدية):

المادة	2018	2019	2020	2021
شاي	3	8	8.5	12
بن	3	5.2	5.4	9.2
سكر	7.5	12	13.7	18

المطلوب: حساب الرقم القياسي التجميعي لهذه المواد الثلاث، علماً أن سنة الأساس هي 2018.

الحل: نحسب أولاً مجموع أسعار المواد الثلاث في كل سنة:

السنة	2018	2019	2020	2021
مجموع الأسعار (ΣP)	13.5	25.2	27.6	39.2

الرقم القياسي التجميعي لسنة 2019 مقارنة بسنة 2018: $I = (25.2/13.5) \times 100 \approx 186.67$

الرقم القياسي التجميعي لسنة 2020: $I = (27.6/13.5) \times 100 \approx 204.44$

الرقم القياسي التجميعي لسنة 2021: $I = (39.2/13.5) \times 100 \approx 290.37$

نستنتج من هذه النتائج أن أسعار هذه المواد الثلاث مجتمعة ارتفعت بشكل مطرد وحاد: بنسبة 86.67% سنة 2019، لتصل الزيادة إلى 104.44% سنة 2020، ثم إلى 190.37% سنة 2021 مقارنة بسنة الأساس 2018.

ملاحظة: على الرغم من بساطة وسهولة حساب الرقم القياسي التجميعي، إلا أن استعماله يتم على حساب الدقة، إذ يُعامل مختلف السلع نفس المعاملة، فتُعطى لها جميعاً نفس الأهمية بغض النظر عن حجم استهلاكها الفعلي أو وزنها الحقيقي في ميزانية المستهلك؛ وهذا اعتبار غير دقيق، إذ لكل سلعة أهميتها ووحدة قياسها وتأثيرها الخاص المتناسب مع تلك الأهمية. لذلك يُستحسن تكوين أرقام قياسية بترجيح مناسب لأسعار السلع تبعاً لأهميتها النسبية، وهذا بالضبط ما ستتناوله الأرقام القياسية الترجيحية المشروحة لاحقاً في هذا الفصل.

ثالثاً: خصائص الأرقام القياسية

حتى نتمكن من اختبار واختيار أفضل صيغة لرقم قياسي معين، تُستعمل عدة معايير رياضية، أهمها معيار الانعكاس ومعيار التحويل (الدوران).

المعيار الأول: خاصية الانعكاس (Propriété de réversibilité)

تتمثل خاصية الانعكاس في تحقق العلاقة التالية بين الرقم القياسي لفترة معينة بالنسبة لفترة أخرى، والرقم القياسي العكسي لهما:

في حالة الأسعار:

$$I_{1/0} * I_{\frac{0}{1}} = 100^2$$

أو في حالة الكميات، بنفس المنطق.

أي أن حاصل ضرب الرقم القياسي الأول (وهو الرقم القياسي للفترة 1 مقارنة بفترة الأساس 0) في الرقم القياسي الثاني (وهو الرقم القياسي العكسي، أي رقم فترة الأساس 0 مقارنة بالفترة 1) يجب أن يساوي 100 تربيع، أي 10000.

مثال رقم (07-05): إذا كان الرقم القياسي لأسعار مادة معينة لسنة 2022 مقارنة بسنة الأساس 2018 يساوي 125، احسب الرقم القياسي العكسي (أي رقم سنة 2018 مقارنة بسنة 2022)، وتحقق من خاصية الانعكاس.

الحل: بتطبيق خاصية الانعكاس:

$$I(2018/2022) = 100^2 / I(2022/2018) = 10000 / 125 = 80$$

وللتحقق: $I(2022/2018) \times I(2018/2022) = 125 \times 80 = 10000 = 100^2$ وهو ما يؤكد تحقق خاصية الانعكاس.

المعيار الثاني: خاصية التحويل أو الدوران (Propriété de circularité)

تتمثل خاصية التحويل أو الدوران في إمكانية ربط عدة أرقام قياسية متتالية للحصول على رقم قياسي مباشر بين أول فترة وآخر فترة، دون الحاجة لإعادة الحساب من البيانات الأصلية. فمثلاً إذا كانت لدينا ثلاثة أرقام قياسية متتالية $I(t1/t0)$ ، و $I(t2/t1)$ ، و $I(t3/t2)$ ، فإن خاصية التحويل تسمح بحساب $I(t3/t0)$ مباشرة وفق العلاقة التالية:

$$I_{t3/t0} = [I_{t1/t0} * I_{t2/t1} * I_{t3/t2}] / 100^2$$

وبصفة عامة، إذا استعملنا f من الأرقام القياسية المتتالية، فإن خاصية التحويل أو الدوران تصبح:

$$I_{tn/t0} = [I_{t1/t0} * I_{t2/t1} * \dots * I_{tn/tn-1}] / 100^{n-1}$$

ملاحظة: يُحسب المعيار الثاني للكميات بنفس الطريقة.

مثال رقم (06-07): انخفض سعر سلعة معينة في سنة 2015 بنسبة 20% مقارنة بسنة 2010، بينما ارتفع سعرها بنسبة 60% في سنة 2010 مقارنة بسنة 2005. المطلوب: حساب الرقم القياسي لسعر هذه السلعة في سنة 2015 مقارنة بسنة 2005، باستعمال خاصية التحويل أو الدوران.

الحل: بما أن السعر انخفض بنسبة 20% في 2015 مقارنة بـ 2010، فإن:

$$I(2015/2010) = 100 - 20 = 80$$

وبما أن السعر ارتفع بنسبة 60% في 2010 مقارنة بـ 2005، فإن:

$$I(2010/2005) = 100 + 60 = 160$$

وبتطبيق خاصية التحويل، هناك رقمان قياسيان متتاليان فقط ($f=2$)، فتصبح العلاقة:

$$I(2015/2005) = [I(2015/2010) \times I(2010/2005)] / 100 = (80 \times 160) / 100 = 128$$

أي أن هناك ارتفاعاً في سعر هذه السلعة بنسبة 28% في سنة 2015 مقارنة بسنة 2005.

مثال رقم (07-07): إذا كانت لدينا الأرقام القياسية التالية التي تبين تطور سعر الفضة عبر فترات متتالية: الرقم القياسي للفترة 1 مقارنة بالفترة 0 يساوي 130، الرقم القياسي للفترة 2 مقارنة بالفترة 1 يساوي 140، والرقم القياسي للفترة 3 مقارنة بالفترة 2 يساوي 125. المطلوب: تحديد الرقم القياسي للفترة 3 مقارنة بالفترة 0 باستعمال خاصية التحويل أو الدوران.

الحل: لدينا ثلاثة أرقام قياسية متتالية ($f=3$)، فتصبح العلاقة:

$$I(3/0) = [I(1/0) \times I(2/1) \times I(3/2)] / 100^2 = (130 \times 140 \times 125) / 10000 = 227.5$$

أي أن سعر الفضة ارتفع في الفترة 3 بنسبة 127.5% مقارنة بالفترة 0.

رابعاً: الأرقام القياسية الترجيحية (Indice pondéré)

يأخذ الرقم القياسي المرجح بعين الاعتبار الأهمية النسبية لكل سلعة من السلع الداخلة في تركيبه، غير أن هذه الأهمية النسبية تتغير باستمرار عبر الزمن؛ ومن هنا يواجه الباحث إشكالية اختيار فترة الأوزان (الترجيح): هل تُستعمل كميات فترة الأساس أم كميات فترة المقارنة؟ ويتوقف اختيار الصيغة المرجحة المناسبة على عدة عوامل، أهمها:

- مدى توفر المعطيات الخاصة بالظاهرة المدروسة.
- درجة السهولة في الحساب.
- مدى تحقيق الصيغة المختارة للخصائص الرياضية المرغوبة (كخاصية الانعكاس والتحويل).
- الهدف والغرض من وضع الرقم القياسي أصلاً.

ومن أهم الأرقام القياسية الترجيحية المستعملة في الإحصاء التطبيقي: رقم لاسبير، رقم باش، رقم فيشر، ورقم مارشال-إدجورث.

(a) الرقم القياسي لاسبير (Laspeyres)

هو الرقم القياسي الذي يعتمد في ترجيحه على كميات فترة الأساس، أي أن الرقم القياسي للأسعار يُرجَّح بكميات فترة الأساس (Q_0)، والرقم القياسي للكميات يُرجَّح بأسعار فترة الأساس (P_0).

❖ لاسبير للأسعار:

$$L_P = \left(\sum P_1 Q_0 / \sum P_0 Q_0 \right) * 100$$

❖ لاسبير للكميات:

$$L_Q = \left(\sum Q_1 P_0 / \sum Q_0 P_0 \right) * 100$$

(b) الرقم القياسي باش (Paasche)

هو الرقم القياسي الذي يعتمد في ترجيحه على كميات فترة المقارنة، أي أن الرقم القياسي للأسعار يُرجَّح بكميات فترة المقارنة (Q_1)، والرقم القياسي للكميات يُرجَّح بأسعار فترة المقارنة (P_1).

❖ باش للأسعار:

$$Pa_p = \left(\sum P_1 Q_1 / \sum P_0 Q_1 \right) * 100$$

❖ باش للكميات:

$$Pa_q = \left(\sum Q_1 P_1 / Q_0 P_1 \right) * 100$$

الرقم القياسي فيشر (Fisher)

هو عبارة عن الوسط الهندسي لكل من الرقمين القياسيين لاسبير وباش، وسُمِّي بالرقم الأمثل (Ideal Index) لأنه يحقق معظم الخصائص الرياضية المطلوبة (كخاصية الانعكاس) التي لا يحققها لاسبير أو باش كل على حدة:

❖ فيشر للأسعار:

$$F_p = \sqrt{(L_p * Pa_p)}$$

❖ فيشر للكميات:

$$F_q = \sqrt{(L_q * Pa_q)}$$

الرقم القياسي مارشال-إدجورث (Marshall-Edgeworth)

يعتمد هذا الرقم القياسي في ترجيحه على متوسط كميات الفترتين معاً (فترة الأساس وفترة المقارنة)، في محاولة للتوفيق بين منطقي لاسبير وباش:

❖ مارشال-إدجورث للأسعار:

$$M_P = \frac{[\sum(P_1(Q_0 + Q_1))]}{[\sum(P_0(Q_0 + Q_1))]} * 100$$

❖ مارشال-إدجورث للكميات:

تُحسب بنفس المنطق، بتبديل أدوار الأسعار والكميات.

مثال رقم (07-08): يمثل الجدول التالي أسعار وكميات أربع سلع (A، B، C، D) خلال سنتي

2018 و 2024:

السلعة	A	B	C	D
السعر (2018) P_0	30	55	80	110
السعر (2024) P_1	42	66	108	330
الكمية (2018) Q_0	5	7	3	1
الكمية (2024) Q_1	8	11	5	1

المطلوب: احسب الرقم القياسي لاسبير وباش وفيشر للأسعار.

الحل: نُكمل الجدول بالأعمدة المساعدة تسهيلاً للحسابات:

السلعة	P_1Q_0	P_0Q_0	P_1Q_1	P_0Q_1
A	210	150	336	240
B	462	385	726	605
C	324	240	540	400

110	330	110	330	D
1355	1932	885	1326	المجموع (Σ)

(1) حساب الرقم القياسي لاسبير للأسعار:

$$L_P = (1326 / 885) \times 100 \approx 149.83$$

أي أن هناك ارتفاعاً في الأسعار بنسبة تقارب 49.83% حسب صيغة لاسبير.

(2) حساب الرقم القياسي باش للأسعار:

$$P_{aP} = (1932 / 1355) \times 100 \approx 142.58$$

أي أن هناك ارتفاعاً في الأسعار بنسبة تقارب 42.58% حسب صيغة باش.

(3) حساب الرقم القياسي فيشر للأسعار:

$$F_P = \sqrt{(149.83 \times 142.58)} \approx \sqrt{21362.75} \approx 146.16$$

أي أن هناك ارتفاعاً في الأسعار بنسبة تقارب 46.16% حسب صيغة فيشر، وهي قيمة وسطية بين نتيجتي لاسبير وباش كما هو متوقع نظرياً، بما أن فيشر هو الوسط الهندسي لهما معاً. ونلاحظ أن صيغة لاسبير أعطت أعلى تقدير للارتفاع (لأنها ترجّح بكميات فترة الأساس الأقل)، بينما أعطت صيغة باش تقديراً أقل (لأنها ترجّح بكميات فترة المقارنة الأكبر)، وهو نمط شائع الملاحظة عملياً عند ارتفاع الأسعار مصحوباً بارتفاع الكميات المستهلكة.

تمارين الفصل السابع

التمرين الأول:

يمثل الجدول التالي أسعار بعض اللوازم المدرسية في بداية سنتين دراسيتين متتاليتين (بالوحدة النقدية):

اللازمة	كراس	قلم	محفظة	كتاب مدرسي	مسطرة
السعر (السنة الأولى)	30	15	600	450	25
السعر (السنة الثانية)	42	18	900	450	20

المطلوب: احسب الرقم القياسي البسيط لسعر كل لازمة، وحدد حالة كل واحدة منها (ثبات، ارتفاع، أو انخفاض).

الحل:

$$\text{الكراس: } I = (42/30) \times 100 = 140 \rightarrow \text{ارتفاع بنسبة } 40\%$$

$$\text{القلم: } I = (18/15) \times 100 = 120 \rightarrow \text{ارتفاع بنسبة } 20\%$$

$$\text{المحفظة: } I = (900/600) \times 100 = 150 \rightarrow \text{ارتفاع بنسبة } 50\%$$

$$\text{الكتاب المدرسي: } I = (450/450) \times 100 = 100 \rightarrow \text{ثبات تام}$$

$$\text{المسطرة: } I = (20/25) \times 100 = 80 \rightarrow \text{انخفاض بنسبة } 20\%$$

نلاحظ أن أغلب اللوازم المدرسية شهدت ارتفاعاً في أسعارها من سنة لأخرى، باستثناء المسطرة التي انخفض سعرها، والكتاب المدرسي الذي بقي سعره ثابتاً (وهو أمر متوقع كون أسعار الكتب المدرسية غالباً ما تكون مدعومة أو مضبوطة إدارياً).

التمرين الثاني:

يمثل الجدول التالي أسعار ثلاث مواد أولية صناعية (البلاستيك، الألمنيوم، النحاس) خلال أربع سنوات متتالية (بالوحدة النقدية):

المادة	2019	2020	2021	2022
بلاستيك	12	13.5	15	21
ألمنيوم	20	22	26	34
نحاس	18	19	25	30

المطلوب: احسب الرقم القياسي التجميعي لهذه المواد الأولية، علماً أن سنة الأساس هي 2019، وعلّق على النتيجة من وجهة نظر مؤسسة صناعية تعتمد على هذه المواد في الإنتاج.

الحل: نحسب أولاً مجموع أسعار المواد الثلاث في كل سنة:

السنة	2019	2020	2021	2022
مجموع الأسعار (ΣP)	50	54.5	66	85

$$I(2020) = (54.5/50) \times 100 = 109 \rightarrow \text{ارتفاع بنسبة } 9\%$$

$$I(2021) = (66/50) \times 100 = 132 \rightarrow \text{ارتفاع بنسبة } 32\%$$

$$I(2022) = (85/50) \times 100 = 170 \rightarrow \text{ارتفاع بنسبة } 70\%$$

التعليق: يعكس هذا الارتفاع المتسارع في تكلفة المواد الأولية (من 9% إلى 70% خلال ثلاث سنوات فقط) ضغطاً متزايداً على تكاليف الإنتاج بالنسبة لأي مؤسسة صناعية تعتمد على هذه المواد؛ وهو مؤشر يستدعي من إدارة المؤسسة إعادة النظر في سياسة التسعير أو البحث عن موردين بديلين أو مواد بديلة لتخفيف الأثر على هامش الربح.

التمرين الثالث:

ارتفع سعر أحد الأدوية الأساسية بنسبة 40% في سنة 2020 مقارنة بسنة 2018، ثم انخفض سعره بنسبة 10% في سنة 2022 مقارنة بسنة 2020.

المطلوب: باستعمال خاصية التحويل أو الدوران، احسب الرقم القياسي لسعر هذا الدواء في سنة 2022 مقارنة بسنة 2018، وفسر النتيجة.

الحل:

$$I(2020/2018) = 100 + 40 = 140$$

$$I(2022/2020) = 100 - 10 = 90$$

وبتطبيق خاصية التحويل (رقمان قياسيان متتاليان):

$$I(2022/2018) = [I(2022/2020) \times I(2020/2018)] / 100 = (90 \times 140) / 100 = 126$$

أي أن سعر هذا الدواء في سنة 2022 أعلى بنسبة 26% مقارنة بسنة 2018، رغم الانخفاض الجزئي المسجل بين 2020 و2022. وهذا مثال جيد يوضح أهمية خاصية التحويل في تتبع التطور الكلي لسعر مادة صحية حساسة عبر عدة فترات متتالية، وهو ما يهم صناع القرار في قطاع الصحة العمومية عند تقييم القدرة الشرائية للمرضى أو عند التفاوض على أسعار الأدوية.

التمرين الرابع:

يمثل الجدول التالي أسعار وكميات مبيعات ثلاثة محاصيل خضروات في سوق جملة الخضار والفواكه خلال سنتين متتاليتين:

المحصول	طماطم	بطاطا	بصل
السعر (P ₀ السنة الأولى)	25	30	20

30	39	35	السعر (P_1 السنة الثانية)
150	300	200	الكمية (Q_0 بالطن، السنة الأولى)
180	330	260	الكمية (Q_1 بالطن، السنة الثانية)

المطلوب: احسب الرقم القياسي لاسبير والرقم القياسي باش والرقم القياسي فيشر لأسعار هذه المحاصيل، وحدد أيهما أنسب لمدير سوق الجملة عند تقييم التضخم في هذا القطاع.

الحل: نُكمل الجدول بالأعمدة المساعدة:

P_0Q_1	P_1Q_1	P_0Q_0	P_1Q_0	المحصول
6500	9100	5000	7000	طماطم
9900	12870	9000	11700	بطاطا
3600	5400	3000	4500	بصل
20000	27370	17000	23200	المجموع (Σ)

$$L_P = (23200 / 17000) \times 100 \approx 136.47 \text{ لاسبير}$$

$$Pa_P = (27370 / 20000) \times 100 \approx 136.85 \text{ باش}$$

$$FP = \sqrt{(136.47 \times 136.85)} \approx 136.66 \text{ فيشر}$$

التعليق: نلاحظ في هذا المثال تقارباً كبيراً جداً بين نتائج الصيغ الثلاث (136.47، 136.85، 136.66)، وهو ما يدل على أن نمط استهلاك هذه المحاصيل الثلاثة لم يتغير بشكل كبير بين السنتين (الأوزان النسبية للمحاصيل الثلاثة متقاربة في كلتا الفترتين). وينصح عادة باعتماد الرقم القياسي فيشر لكونه "الرقم الأمثل" الذي يوازن بين منطقي لاسبير وباش ويحقق أكبر عدد من الخصائص الرياضية المطلوبة، مما يجعله الخيار الأنسب لمدير سوق الجملة عند إعداد تقارير دورية دقيقة حول تطور أسعار الخضار.

الخاتمة العامة

خاتمة عامة

تناولت هذه المطبوعة، عبر فصولها السبعة، أهم المفاهيم والأدوات المنهجية لمقياس الإحصاء الوصفي، بدءاً بالمفاهيم الأساسية والمصطلحات التي يقوم عليها هذا العلم، مروراً بأساليب عرض البيانات الإحصائية جدولياً وبيانياً، ثم مقاييس الوصف الإحصائي الثلاثة (النزعة المركزية، التشتت، الشكل) التي تمكن الطالب من تكوين صورة رقمية دقيقة وشاملة عن أي ظاهرة يدرسها، وصولاً إلى مقاييس الارتباط والانحدار الخطي البسيط التي تفتح الباب أمام دراسة العلاقة بين متغيرين، وانتهاءً بالأرقام القياسية التي تتيح متابعة تطور الظواهر عبر الزمن.

وقد حرصنا طوال هذه المطبوعة على الموازنة بين الطرح النظري والتطبيق العملي، من خلال إثراء كل فصل بأمثلة وتمارين محلولة ومقترحة، تعين الطالب على استيعاب المفاهيم وترسيخها من جهة، وعلى اكتساب المهارة اللازمة لحل المسائل الإحصائية المختلفة من جهة أخرى.

وإذا كان الإحصاء الوصفي يقتصر على وصف البيانات المتوفرة عن مجتمع الدراسة أو عينته دون تعميم النتائج، فإن ما اكتسبه الطالب من مفاهيم وأدوات في هذه المطبوعة يشكل الأساس المتين الذي سيبنى عليه في مقياس الإحصاء الاستدلالي، حيث يتم الانتقال من وصف العينة إلى تعميم النتائج على المجتمع الإحصائي ككل، باستعمال أدوات نظرية الاحتمالات واختبارات الفرضيات وفترات الثقة.

وفي الختام، نأمل أن تكون هذه المطبوعة قد حققت الغرض المرجو منها، وأسهمت في تيسير مقياس الإحصاء الوصفي على طلبتنا الأعزاء، كما ندعو الله عز وجل أن يجعل هذا العمل خالصاً لوجهه الكريم، نافعاً لكل من يطلع عليه من الطلبة والدارسين، والحمد لله رب العالمين.

قائمة المصادر

أولاً: المراجع باللغة العربية

1. تيلولت سامية، مبادئ في الإحصاء، الطبعة الثانية، دار الحديث للكتاب، الجزائر، 2009.
2. جلاطو جيلالي، الإحصاء مع تمارين ومسائل محلولة، ديوان المطبوعات الجامعية، الجزائر، 2001.
3. جلال الصياد، مبادئ الإحصاء، دار حافظ للنشر والتوزيع، المملكة العربية السعودية، 2003.
4. حسن ياسين طعمة، طرق الإحصاء الوصفي، دار صفاء للنشر والتوزيع، الأردن، 2009.
5. سهيل أحمد سمحان، محمود حسين الوادي، مبادئ الإحصاء للاقتصاد والعلوم الإدارية، الطبعة الأولى، دار صفاء للنشر والتوزيع، عمان، 2010.
6. شرف الدين خليل، الإحصاء الوصفي، مكتبة الأنجلو المصرية، القاهرة، 2008.
7. صالح رشيد بطارسة، الإحصاء والاحتمالات، الطبعة الأولى، دار أسامة للنشر والتوزيع، عمان، 2010.
8. عبد الحليم الحمزة، كتاب بيداغوجي في الإحصاء 1، كلية العلوم الاقتصادية والعلوم التجارية وعلوم التسيير، جامعة العربي التبسي، تبسة، 2022.
9. علي عبد السلام العماري، علي حسين العجيلي، الإحصاء والاحتمالات: النظرية والتطبيق، منشورات ELGA، مالطا، 2000.
10. كامل فليف، فتحي حمدان، الإحصاء، دار المناهج للنشر والتوزيع، الأردن، 2012.
11. محمد بوهزة، محاضرات في الإحصاء الوصفي، دار المحمدية العامة، الجزائر، 2011.
12. محمد عبد العال النعيمي، الإحصاء المتقدم في دعم القرار بالتركيز على منظمات الأعمال الإنتاجية، الطبعة الأولى، مؤسسة الوراق للنشر والتوزيع، عمان، 2007.
13. محمود علي أبو النصر (مترجم)، الاحتمالات والإحصاء "ملخصات إيزي شوم"، الطبعة العربية الأولى، الدار الدولية للاستثمارات الثقافية، مصر، 2004.

14. مهدي محمد القصاص، مبادئ الإحصاء والقياس الاجتماعي، مطبوعة جامعية، جامعة المنصورة، مصر، 2007.
15. نداء محمد الصوص، مبادئ الإحصاء، الطبعة الأولى، مكتبة المجتمع العربي للنشر والتوزيع، عمان، 2007.
16. نور الدين حامد، محاضرات وتمارين محلولة في الإحصاء الوصفي، دار اليازوري العلمية للنشر والتوزيع، الأردن، 2017.
17. هدى برهان سيف الدين، الإحصاء في علم النفس (مقاييس التشتت)، مطبوعة جامعية، جامعة الملك عبد العزيز، السعودية، 2014.
18. هشام بورمة، الإحصاء 1: دروس وأمثلة تطبيقية، كلية العلوم الاقتصادية والتجارية وعلوم التسيير، جامعة محمد الصديق بن يحيى، جيجل، الجزائر، 2018/2017.

ثانياً: المراجع الأجنبية

1. Aws Abueid, Big Data and Cloud Computing: Opportunities and Application Areas, 2024.
2. David Freedman, Robert Pisani, Roger Purves, Statistics, Fourth Edition, W. W. Norton and Company, New York, 2007.
3. Bernard Grais, Techniques statistiques 1: Statistique descriptive, 3e edition, Dunod, Paris, 1994.
4. L. Hunyadi, L. Vita, Statistics: Introduction, History, Subject and Task of Statistics, Central Statistical Office, 2008.
5. Jean Michel Josselin, Benoit Le Maux, Statistical Tools for Program Evaluation, Springer International Publishing, 2017.
6. William M. Mendenhall, Statistics for Engineering and the Sciences, Sixth Edition, CRC Press, Taylor and Francis Group, New York, 2016.

7. Karen A. Randolph, Laura L. Myers, Basic Statistics in Multivariate Analysis, Pocket Guides to Social Work Research Methods, New York, 2013.
8. Clemens Reimann, Peter Filzmoser, Statistical Data Analysis Explained, John Wiley and Sons, New York, 2008.
9. Michel Tenenhaus, Statistique: methodes pour decire, expliquer et prevoir, 2e edition, Dunod, Paris, 2007.
10. Timothy C. Urdan, Statistics in Plain English, Third Edition, Taylor and Francis Group, New York, 2010.
11. Isabel Willemse, Statistical Methods and Calculation Skills, Third Edition, Juta, Cape Town, 2009.
12. Sitong Zhu, The Application of Big Data in Economic Statistics, 2024.