# اسم المقرر

## مبادئ الإحصاء

د. سعيد سيف الدين

جامعة الملك فيصل
عمادة التعلم الإلكتروني والتعليم عن بعد

الحمد لله رب العالمين ، والصلاة والسلام على خاتم الأنبياء والمرسلين سيدنا ونبينا محمد بن عبد الله وعلى آله وصحبه أجمعين

# المحاضرة الرابعة عشرة والأخيرة

# عناصر المحاضرة

## تجميع للتعريفات النظرية الخاصة بهذا المقرر

أما التفاصيل وكيفية إجراء العمليات الحسابية والأمثلة فمرجعها المحاضرات المسجلة

# الباب الأول : مفاهيم أساسية

**الإحصاء الوصفي :** هو العلم الذي يهتم بجمع وتبويب وعرض ووصف البيانات وحساب بعض المقاييس الخاصة بها دون الوصول إلى نتائج أو استدلالات خاصة

**الإحصاء الاستقرائي أو الاستدلال الإحصائي أو الإحصاء الاستدلالي :** هو العلم الذي يختص يبحث في استقراء النتائج واتخاذ القرارات

**المتغير النوعي :** هو المتغير الذي لا يمكن التعبير عنه بعدد [مثل لون العين/رأيك في موضوع/لون سيارات بأحد المواقف/....]

**المتغير الكمي :** هو المتغير الذي يُعبر عنه بعدد [مثل عدد الطلاب/الوزن/الدخل/.....]

**المتغير الكمي المتصل :** هو المتغير الذي يمكن أن يأخذ أي قيمة بين قيمتين ، أي يمكن أن يُقاس ولا يُعد [الوزن/الدخل/.....]

**المتغير الكمي المتقطع :** هو المتغير الذي يمكن أن يأخذ قيمتين لكن لا يأخذ أي قيمة بينهما ، أي يمكن أن يُعد ولا يُقاس [عدد الطلاب/عدد أيام شهر ما/.....]

**البيانات النوعية :** هي البيانات (أو المشاهدات) التي يكون فيها المتغير متغير نوعي .

**البيانات الكمية :** هي البيانات (أو المشاهدات) التي يكون فيها المتغير متغير كمي .

**البيانات الكمية المتصلة :** هي البيانات (أو المشاهدات) التي يكون فيها المتغير متغير كمي متصل .

**البيانات الكمية المتقطعة :** هي البيانات (أو المشاهدات) التي يكون فيها المتغير متغير كمي متقطع .

# الباب الأول : مفاهيم أساسية

**البيانات المنفصلة :** هي بيانات إما نوعية أو كمية متقطعة .

**جمع البيانات :** هي عملية الحصول على القياسات الخاصة بظاهرة معينة وعادةً ما تُسمى البيانات المجمعة **بالبيانات الخام**

**تنظيم وعرض البيانات :** هي عملية وضع البيانات المجمعة في جداول خاصة وعرضها بطرق مناسبة

**تحليل البيانات :** هي عملية إيجاد مقاييس تتحدد قيمها من البيانات وتعطي بعض الدلالات عن الظاهرة تحت الدراسة

**استقراء النتائج واتخاذ القرارات :** هي الاستنتاجات التي يتوصل إليها الباحث من خلال تحليله للبيانات وعادةً ما تكون على شكل تقديرات أو تنبوءات أو تعميمات أو قرارات بالرفض أو القبول

# الباب الثاني : التوزيعات التكرارية

**المدى $R$ :** وهو "الفرق بين أكبر قيمة وأقل قيمة" في البيانات المعروضة

**الجدول (التوزيع)التكراري :** هو جدول يوضح قيم المتغير وتكراراتها

**الجدول (التوزيع)التكراري النسبي :** هو جدول يوضح قيم المتغير وتكراراتها وأيضاً تكراراتها النسبية

**التكراري النسبي :** هو خارج قسمة التكرار على مجموع التكرارات [أو عبارة مكافئة] ويمكن أن يُعبر عنه كنسبة عادية أو نسبة مئوية

## العرض البياني للبيانات المنفصلة :

(١) **طريقة الأعمدة البسيطة :** حيث تُمثل كل قيمة من قيم المتغير بعمود (خط رأسي) طوله يُعبر عن تكرار تلك القيمة

(٢) **طريقة القضبان البسيطة :** حيث تُمثل كل قيمة من قيم المتغير بقضيب (خط أفقي) طوله يُعبر عن تكرار تلك القيمة

(٣) **طريقة المضلع التكراري :** حيث تُمثل كل قيمة من قيم المتغير وتكرارها بنقطة ثم نقوم بتوصيل هذه النقاط بخط منكسر (بواسطة المسطرة)

(٤) **طريقة المنحنى التكراري :** حيث تُمثل كل قيمة من قيم المتغير وتكرارها بنقطة ثم نقوم بتوصيل هذه النقاط بخط ممهد (باليد)

(٥) **طريقة الدائرة :** حيث تُمثل كل قيمة من قيم المتغير بقطاع من دائرة وذلك طبقاً لتكرارها ، وتتحدد الزاوية المركزية لكل قطاع من :

$$\text{الزاوية المركزية لقيمة ما} = \frac{\text{تكرار القيمة}}{\text{مجموع التكرارات}} \times ٣٦٠$$

أو

**الزاوية المركزية لقيمة ما = التكرار النسبي للقيمة × ٣٦٠**

# الباب الثاني : التوزيعات التكرارية

## العرض البياني للبيانات المنفصلة لظاهرتين :

(١) طريقة الأعمدة المزدوجة : حيث يُمثل كل زوج من قيم الظاهرتين (أي قيمة من الظاهرة الأولى وقيمة من الظاهرة الثانية) بعمود مزدوج

(٢) طريقة الأعمدة المجزأة : حيث يُمثل كل زوج من قيم الظاهرتين بعمود واحد يتم تجزئته طبقاً لقيم الظاهرتين .

## العرض الجدولي للبيانات المتصلة :

وفي هذا النوع من البيانات يُعبر عن المتغير على صورة فئات ، ويكون لكل فئة حدان : أدنى وأعلى ويكون الحد الأدنى لكل فئة (عدا الأولى) مساوياً للحد الأعلى للفئة السابقة ، والحد الأعلى لكل فئة هو الحد الأدنى للفئة التالية . ويكون :

- طول أي فئة = حدها الأعلى – حدها الأدنى
- مركز أي فئة = مجموع حديها الأدنى والأعلى مقسوماً على 2
- كثافة تكرار أي فئة = تكرار الفئة مقسوماً على طولها

**الجدول (التوزيع) التكراري** :هو جدول يوضح فئات المتغير وتكراراتها

**الجدول (التوزيع) التكراري النسبي** : هو جدول يوضح فئات المتغير وتكراراتها وأيضاً تكراراتها النسبية

وإذا كان الحد الأدنى للفئة الأولى غير معروف يُسمى الجدول (التوزيع) مفتوح من أسفل

وإذا كان الحد الأعلى للفئة الأخيرة غير معروف يُسمى الجدول (التوزيع) مفتوح من أعلى

وإذا كان الحدان : الأدنى للفئة الأولى والأعلى للفئة الأخيرة غير معروفين يُسمى الجدول (التوزيع) مفتوح من الطرفين

# الباب الثاني : التوزيعات التكرارية

## الجدول (التوزيع) التكراري الصاعد :

التكرار المتجمع الصاعد المناظر لقيمة معية $a$ لمتغير $x$ هو مجموع تكرارات جميع قيم المتغير الأقل من $a$

## الجدول (التوزيع) التكراري الهابط :

التكرار المتجمع الهابط المناظر لقيمة معية $a$ لمتغير $x$ هو مجموع تكرارات جميع قيم المتغير الأكبر من أو تساوي $a$

## العرض البياني للبيانات المتصلة :

(١) طريقة الدائرة : حيث تُمثل كل فئة من فئات المتغير بقطاع من دائرة وذلك طبقاً لتكرارها [مشابهة تماماً لتمثيل البيانات المنفصلة]

(٢) طريقة المدرج التكراري : حيث تُمثل الفئات بمستطيلات متلاصقة بحيث يمثل كل مستطيل إحدى الفئات ، بحيث تقع قاعدة المستطيل (الممثل لفئة ما) على المحور الأفقي [محور المتغير] وممتدة بين الحد الأدنى للفئة وحدها الأعلى [أي طول قاعدة المستطيل يساوي طول الفئة] وارتفاعه هو كثافة تكرار الفئة ومساحته هي تكرار الفئة .

(٣) طريقة المضلع التكراري : حيث تُمثل كل فئة من الفئات بنقطة إحداثياتها هي مركز الفئة وكثافة تكرارها ثم نقوم بتوصيل هذه النقاط بخط منكسر (بواسطة المسطرة) .

(٤) طريقة المنحنى التكراري : حيث تُمثل كل فئة من الفئات بنقطة إحداثياتها هي مركز الفئة وكثافة تكرارها ثم نقوم بتوصيل هذه النقاط بخط ممهد (باليد) .

# الباب الثاني : التوزيعات التكرارية

## المضلع التكراري المتجمع الصاعد :

حيث تُمثل كل فئة من الفئات بنقطة إحداثياتها هي الحد الأدنى للفئة والتكرار المتجمع الصاعد المناظر ثم نقوم بتوصيل هذه النقاط بخط منكسر (بواسطة المسطرة) ، وهنا لابد من تكوين الجدول التكراري المتجمع الصاعد . وأحياناً يُسمى هذا المضلع بـــ "مضلع الأقل من" .

## المضلع التكراري المتجمع الهابط :

حيث تُمثل كل فئة من الفئات بنقطة إحداثياتها هي الحد الأدنى للفئة والتكرار المتجمع الهابط المناظر ثم نقوم بتوصيل هذه النقاط بخط منكسر (بواسطة المسطرة) ، وهنا لابد من تكوين الجدول التكـــراري المتجمع الصاعد . وأحياناً يُسمى هذا المضلع بـــ "مضلع الأكبر من أو يساوي" .

# الباب الثالث : مقاييس النزعة المركزية

**الوسط الحسابي :** هو مجموع القيم مقسوماً على عددها

مزاياه :سهولة حسابه – يأخذ في الاعتبار جميع القيم – لا يحتاج لترتيب القيم

عيوبه : يتأثر بشدة بالقيم المتطرفة – لا يمكن إيجاده بالرسم – لا يمكن حسابه للتوزيعات المفتوحة

**الوسيط :** هو القيمة التي تقسم مجموعة من القيم (المرتبة تصاعدياً أو تنازلياً) إلى مجموعتين متساويتين في العدد

مزاياه : سهولة حسابه – لا يتأثر بالقيم المتطرفة – يمكن حسابه بالرسم – يمكن حسابه للتوزيعات المفتوحة

عيوبه : يحتاج لترتيب القيم تصاعدياً أو تنازلياً – لا يأخذ في الاعتبار جميع البيانات

**المنوال (الشائع) :** هو القيمة الأكثر تكراراً (أو شيوعاً) .

مزاياه :سهولة حسابه – لا يتأثر كثيراً بالقيم المتطرفة – لا يحتاج لترتيب البيانات – يمكن تحديده في حالة البيانات النوعية

عيوبه : قد لا يتواجد – قد يكون هناك أكثر من منوال

وللمنحنيات التكرارية وحيدة المنوال وبسيطة الالتواء هناك علاقة اعتبارية أساسية هي : **الوسط – المنوال = ٣(الوسط – الوسيط)**

- فإذا كان المنحنى متماثلاً يكون : الوسط = الوسيط = المنوال
- وإذا كان المنحنى ملتوياً لليمين (التواء موجب) يكون : الوسط أكبر من الوسيط أكبر من المنوال
- وإذا كان المنحنى ملتوياً لليسار (التواء سالب) يكون : الوسط أقل من الوسيط أقل من المنوال

# الباب الرابع : مقاييس التشتت

**التشتت** : هو الدرجة التي تتجه بها البيانات الكمية للانتشار حول قيمة متوسطة (أحد مقاييس النزعة المركزية) ، ومن مقاييسه :

(١) المدى : مدى مجموعة من البيانات الكمية هو الفرق بين أكبر قيمة في البيانات وأقل قيمة فيها

(٢) الانحراف المتوسط (أو متوسط الانحرافات) : هو متوسط القيم المطلقة للانحرافات عن الوسط الحسابي

(٣) الانحراف المعياري : هو الجذر التربيعي لمتوسط مربعات الانحرافات عن الوسط الحسابي

متوسط مجموع مربعات الانحرافات عن الوسط الحسابي يُسمى التباين ، أي أن الانحراف المعياري هو الجذر التربيعي للتباين [أو التباين هو مربع الانحراف المعياري]

(٤) المدى الربيعي : هو الفرق بين الربيع الثالث والربيع الأول

(٥) الانحراف الربيعي : هو نصف المدى الربيعي [أي نصف الفرق بين الربيع الثالث والربيع الأول]

$Q_1$ [الربيع الأول] هي تلك القيمة التي يقع تحتها 25% من القيم [وبالطبع وفوقها 75% من القيم]

$Q_3$ [الربيع الثالث] هي تلك القيمة التي يقع تحتها 75% من القيم [وبالطبع وفوقها 25% من القيم]

(٤) المدى المئيني : هو الفرق بين المئين التسعين والمئين العاشر

$P_{10}$ [المئين العاشر] هي تلك القيمة التي يقع تحتها 10% من القيم [وبالطبع وفوقها 90% من القيم]

$P_{90}$ [المئين التسعون] هي تلك القيمة التي يقع تحتها 90% من القيم [وبالطبع وفوقها 10% من القيم]

# الباب الرابع : مقاييس التشتت

وفي حالة التوزيعات متوسطة الالتواء هناك علاقات اعتبارية (تقريبية) بين مقاييس التشتت السابقة تتلخص في العلاقتين :

الانحراف المتوسط = $\frac{٤}{٥}$ × الانحراف المعياري

الانحراف الربيعي = $\frac{٢}{٣}$ × الانحراف المعياري

**التشتت النسبي :** هو : $\frac{\text{التشتت المطلق}}{\text{المتوسط}} \times 100$ ومقاييسه :

(١) معامل الاختلاف (معامل التشتت) ويساوي :

$$\frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}} \times 100$$

(٢) معامل الاختلاف الربيعي ويساوي :

$$\frac{\text{الربيع الثالث} - \text{الربيع الأول}}{\text{الربيع الثالث} + \text{الربيع الأول}} \times 100$$

---

**الدرجة المعيارية لقيمة ما :** تساوي $\frac{\text{القيمة} - \text{الوسط الحسابي}}{\text{الانحراف المعياري}}$ ولها أهمية كبيرة في مقارنة نتائج بيانات مختلفة ببعضها

# الباب الخامس : الالتواء والتفرطح

**الالتواء :** هو درجة تماثل أو البعد عن التماثل لتوزيع ما ، ومقاييسه :

(١) معامل بيرسون الأول للالتواء = $\frac{\text{الوسط} - \text{المنوال}}{\text{الانحراف المعياري}}$

(٢) معامل بيرسون الثاني للالتواء = $\frac{٣(\text{الوسط} - \text{الوسيط})}{\text{الانحراف المعياري}}$

(٣) معامل الالتواء الربيعي= $\frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$

(٤) معامل الالتواء المئيني= $\frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$

معامل التواء موجب يعني التواء لليمين ، ومعامل التواء سالب يعني التواء لليسار

**التفرطح :** هو درجة تدبب (ارتفاع أو انخفاض) قمة المنحنى مقارنة بقمة منحنى التوزيع الطبيعي ، ومقياسه : :

معامل التفرطح المئيني = $\frac{\text{نصف المدى الربيعي}}{\text{المدى المئيني}} = \frac{\text{الانحراف الربيعي}}{\text{المدى المئيني}}$

معامل تفرطح أكبر من 0.26 يعني مدبب ، وأقل يعني مفرطح

# الباب السادس : تحليل الارتباط

يمكن تحديد ما إذا كان هناك ارتباط بين متغيرين أم لا ونوع هذا الارتباط (إن وُجِد) وقوته وذلك عن طريق "شكل الانتشار" أو عن طريق "معامل ارتباط الرتب *r*" حيث :

$$r = 1 - \frac{6\sum D^2}{n(n^2 - 1)}$$

حيث $D$ تمثل الفرق في الرتب بين قيم $y$ , $x$ والتي تنحصر قيمته بين +1 , -1، فإذا كانت قيمته :

- موجبة ، دل ذلك على أن هناك ارتباط طردي بين المتغيرين $y$ , $x$
- سالبة ، دل ذلك على أن هناك ارتباط عكسي بين المتغيرين $y$ , $x$
- صفراً ، دل ذلك على أنه ليس هناك ارتباط بين المتغيرين $y$ , $x$

أما قوة الارتباط فتحددها القيمة المطلقة لمعامل الارتباط طبقاً للجدول المبين

| القيمة المطلقة لمعامل الارتباط | قوة الارتباط |
|---|---|
| 0 | لا يوجد ارتباط |
| $0 < r \leq 0.4$ | ارتباط ضعيف |
| $0.4 < r \leq 0.6$ | ارتباط متوسط |
| $0.6 < r < 1$ | ارتباط قوي |
| 1 | ارتباط تام |
| > 1 | خطأ في الحسابات |

# تمت بحمد الله