المحاضرة النظرية الثالثة
إحصاء

إعداد

د. حيدر هاشم الحسن

25/10/2020

مقاييس النزعة المركزية

2- المنوال: يسمى أحياناً الشائع أو النمط أو القمّة.

والمنوال ببساطة: "القيمة التي تتكرّر أكثر من غيرها من القيم المعطاة " ويكثر استخدامه في البيانات الوصفية، لمعرفة النمط (المستوى) الشائع .

حيث أنّ البيانات سوف تتركّز حول هذه القيمة المحدَّدة أو النموذجيّة.

يوجد هناك بيانات تختلف باختلاف المنوال، بسبب اختلاف عدد القيم المتكرّرة.

فتصنّف البيانات حسب نوع المنوال إلى ثلاثة أنواع هي:

1- بيانات عديمة المنوال: أي أنه لا يوجد قيمة تتكرر أكثر من غيرها.

2- بيانات وحيدة المنوال: حيث تكون هناك قيمة واحدة فقط تتكرر أكثر من غيرها.

3- بيانات عديدة المنوال: حيث تكون هناك قيمتين أو أكثر تتكرر في البيانات.

سنتعرف على طريقة استخراج المنوال من:

1- البيانات المبوّبة.

2- البيانات غير المبوّبة.

1- البيانات غير المبوّبة (التي لا تحوي على فئات):

مثال1: أوجد المنوال للقيم التالية: 4, 2, 5, 3, 4, 5, 2, 3, 4.

المنوال = 4 لأنّه تكرّر أكثر من غيره.

مثال2: أوجد المنوال للقيم التالية: 15, 11, 9, 8, 7

ليس لهذه القيم منوال لعدم تكرار أي قيمة منها.

ويمكن أن يكون هناك أكثر من منوال في حال تكرار قيمتين بنفس العدد.

2- البيانات المبوّبة (التي تحوي على فئات):

بكل بساطة نرجع للجدول والقيمة التي تملك أكبر عدد في التكرارات هي "المنوال".

المنوال في البيانات المبوبة (طريقة الفروض)

ويمكن حسابها من المعادلة التالية :

$$Mod = A + \frac{d_1}{d_1 + d_2} \times L$$

A : الحد الأدنى لفئة المنوال(الفئة المناظرة لأكبر تكرار)

d_1 : الفرق الأول = (تكرار فئة المنوال – تكرار سابق)

d_2 : الفرق الثاني =تكرار فئة المنوال – تكرار لاحق

L : طول فئة المنوال

فئة المنوال = الفئة المناظرة لأكبر تكرار

مثال :

في الجدول التكراري التالي درجات خمسين طالب اقصى درجة هي 100

التكرار المتجمع الصاعد	التكرار	الحدود الحقيقة	حدود الفئة
3	3	59.5 - 49.5	59-50
8	5	69.5 - 59.5	69-60
26	18	79.5 - 69.5	79-70
42	16	89.5 - 79.5	89-80
50	8	99.5 - 89.5	99-90
	50		المجموع

$d_1 = 18 - 5 = 13$

$d_2 = 18 - 16 = 2$

وبالتالي فإنَّ $d_1 + d_2 = 15$

$A = 70$

$L = 59 - 50 = 9$

وبالتالي نقوم بتطبيق المعادلة

$$Mod = 70 + \frac{13}{13+2} \times 9$$

$$= 70 + 7.8 = 77.8$$

ميزات المنوال :

عدم تأثّره بالقيم المتطرفة وأنَّه يمثّل غالبية المشاهدات وعند احتسابه لا يحتاج لكافة قيم التوزيع .

عيوب المنوال

أبرز عيوبه عندما تكون القيم منتشرة على مدى واسع عندها يصبح اقل تعبيرا كمتوسط .

3- الوسيط :

- يمثّل الوسيط أحد قيم النزعة المركزية، ويتم استخدامه عادةً لتحديد المتوسط التقريبي لمجموعة من البيانات، وهو يمثّل الرقم الأوسط من قائمة الأرقام المرتّبة تصاعدياً أو تنازلياً؛ - حيث يقسم القيم إلى قسمين أو جزأين علوي وسفلي، **ليمثّل القيمة الوسطى فيها**

- كما يمكن استخدامه لتحديد المتوسّط بشكل تقريبي،

- وقد تختلف قيمة الوسيط عن قيمة المتوسط الحسابي بشكل كبير، خاصةً عند وجود قيم متطرفة ومختلفة بشكل كبير في قيمها عن معظم القيم أو المشاهدات، الأمر الذي قد يؤدّي إلى التأثير على قيمة المتوسط

- وفي المقابل عدم تأثّر الوسيط بها بشكل كبير؛ حيث يعتبر الوسيط أقل تأثراً بهذه القيم مقارنة بالمتوسط الحسابي

لإيجاد الوسيط لمجموعة من البيانات العددية بشكل يدويّ، يجب اتّباع الخطوات التالية بالترتيب، وهي:

1- ترتيب القِيم من الأصغر إلى الأكبر، أو من الأكبر إلى الأصغر؛ أي يُمكن ترتيبها تصاعديّاً أو تنازليّاً.

2-عدّ القِيم، فإذا كان عددها فرديّاً، فالوسيط هو العدد الذي يتوسّط هذه القيم بعد ترتيبها، ويمكن تحديد ترتيبه عن طريق تطبيق القانون التالي:

ترتيب الوسيط=2/(عدد المشاهدات+1)

فمثلاً الوسيط لمجموعة الأعداد الآتية بعد ترتيبها: 4,5,6,7,8 هو العدد 6، وهي القيمة الثالثة في الترتيب.

3- إذا كان عدد القيم زوجيّاً، فالوسيط حينها هو المتوسّط الحسابي للعددَين الأوسطَين؛ والتي يتم تحديد ترتيبها عن طريق القانون: **عدد المشاهدات/2**، فيكون الوسيط هو المتوسط الحسابي لهذه القيمة والقيمة التي تليها؛

فمثلاً الوسيط لمجموعة الأعداد الآتية بعد ترتيبها: 3,4,7,9,12,15 هو 2 /(9+7)=8، وهو يمثّل المتوسط الحسابي للقيمتين الثالثة والرابعة في الترتيب.

حساب الوسيط للجداول البيانية:

يتمّ عادةً حساب الوسيط للبيانات المجمّعة ضمن الجداول التكرارية من خلال القانون الآتي:

الوسيط= القيمة الدنيا للفئة الوسيطية+ (((مجموع التكرارات الكلي/2) - قيمة التكرار التراكمي قبل الفئة الوسيطية) / تكرار الفئة الوسيطية) × طول الفئة الوسيطية.

مثال: احسب الوسيط للبيانات الآتية التي تمثّل الوقت المستغرق للذهاب إلى العمل لخمسين شخصاً:

الوقت المستغرق	التكرار	التكرار المتجمع (التراكمي)
10 -1	8	8
20 – 11	14	22
30 – 21	12	34
40- 31	9	43
50- 41	7	50
المجموع	50	-

الحل:

يجب لحساب الوسيط أولاً تحديد الفئة التي يوجد فيها (الفئة الوسيطية)، وهي أول فئة تبلغ قيمة التكرار التراكمي لها القيمة ن أو تزيد؛

حيث ن= رتبة الوسيط= مجموع القيم\2، وفي هذه الحالة ن= 50/2=25،

وأول فئة تبلغ قيمة التكرار التراكمي لها العدد 25 هي الفئة الثالثة (21-30).

التعويض في القانون مباشرةً:

حيث: القيمة الدنيا للفئة الوسيطية=**21**. مجموع التكرارات الكلي=50.

قيمة التكرار التراكمي قبل الفئة الوسيطية=22. تكرار الفئة الوسيطية=12.

طول الفئة الوسيطية=10.

الوسيط= القيمة الدنيا للفئة التي يوجد الوسيط فيها+ (((مجموع التكرارات الكلي/2)-قيمة التكرار التراكمي قبل الفئة الوسيطية) / تكرار الفئة الوسيطية)× طول الفئة الوسيطية= 21 + (((50/2) – 22) / 12) × 10 = 23.5 = 24

يتضح مما سبق أنَّ هناك 25 شخصاً يستغرق وقت الذهاب إلى العمل لديهم مدة تقل عن 24 دقيقة،

أمَّا البقية المتمثّلة بالـ 25 الآخرين فيستغرق الذهاب إلى العمل لديهم مدة تزيد عن 24 دقيقة.

مقاييس التشتّت والاختلاف
MEASURES OF DISPERSION AND VARIATION

- عرضنا سابقاً مقاييس النزعة المركزية ووجدنا أنَّ هذه المقاييس مهمة من حيث أنَّها تعطينا فكرة جيدة ووافية عن عناصر العينة المدروسة ومدى نزعتها للتجمع حول المتوسط؛ فالمتوسط الحسابي قيمة قريبة من كل عناصر العينة تقريباً (باستثناء القيم الشاذة)،
- لكن في حال تساوي قيمة المتوسط لعينتين بالرغم من اختلافهما من حيث قيم العناصر فكيف نفرق بين العينتين والحكم من حيث التجانس وغيره من المؤشرات الإحصائية.
- من هنا كان لابدَّ من وجود مؤشّرات خاصّة لاستخدامها في حال تساوي قيمة المتوسط لعينتين،
- هذه المؤشّرات الأكثر أهميةً واستخداماً هي مقاييس التشتّت والاختلاف

1- المدى (RANGE):

- **تعريف: المدى هو الفرق بين أكبر وأصغر قيمة في العينة المدروسة.**
- في جدول التوزيع التكراري يتم حساب المدى على أنه الفرق بين الحد الأعلى للفئة الأخيرة والحدّ الأدنى للفئة الأولى، أو هو الفرق بين مركز الفئة الأولى ومركز الفئة الأخيرة.
- **مثال: لدينا العينة التالية:**

X={33,15,20,25,35}

الحل: نلاحظ أنَّ أصغر قيمة أي MIN (X)= 15 وأكبر قيمة هي MAX(X)= 35 وبالتالي فإنَّ: المدى هو التالي:

RANGE=MAX(X)-MIN(X) = 35 – 15 = 20

مثال: الجدول التكراري يبين توزيع 60 مزرعة حسب المساحة المزروعة بالذرة بالألف دونم

المساحة	20-15	25-20	30-25	35-30	40-35	45-40
عدد المزارع	3	9	15	18	12	3

والمطلوب: حساب المدى للمساحة المزروعة بالذرة

الحل: المدى = مركز الفئة الأخيرة – مركز الفئة الأولى

مركز الفئة الأخيرة= 2/(45+40) = 85/2= 42.5

مركز الفئة الأولى: 2/(20+15) = 35/2 = 17.5

المدى = 17.5-42.5= 25 دونم

مزايا وعيوب المدى:

- **مزايا المدى:**

1- أنه بسيط وسهل الحساب.

2- يكثر استخدامه عند الإعلان عن حالات الطقس والمناخ الجوي، مثل درجات الحرارة ، والرطوبة والحرارة .

3- يستخدم في مراقبة الجودة.

- **من عيوبه:**

1- أنَّه يعتمد على قيمتين فقط، ولا يأخذ جميع القيم بالحسبان.

2- يتأثَّر بالقيم الشاذّة.

2- الانحراف الربيعي: Q

- لجأ الإحصائيون إلى هذا المقياس الذي يعتمد على نصف عدد القيم الوسطى ويهمل نصف عدد القيم المتطرفة، ولذلك لا يتأثَّر هذا المقياس بوجود قيم شاذّة ويعطى بالعلاقة التالية:

$$Q=\frac{Q3-Q1}{2}$$

مزايا وعيوب الانحراف الربيعي:

من مزايا الانحراف الربيعي: 1- يفضل استخدامه كمقياس للتشتت في حالة وجود قيم شاذة،

2- بسيط وسهل في الحساب.

من عيوبه: أنَّه لا يأخذ كل القيم في الاعتبار.

3– الانحراف المتوسط (AD): يعبّر عنه بمتوسط الانحرافات المطلقة للقيم عن وسطها الحسابي

- للبيانات غير المبوّبة يحسب كالتالي:

$$AD = \frac{\sum |X - \mu|}{N} \quad \text{for populations} \qquad (2.6a)$$

and

$$\mathrm{AD} = \frac{\sum |\mathrm{X} - \bar{\mathrm{X}}|}{\mathrm{n}} \quad \text{for samples} \qquad (2.6b)$$

وللبيانات المبوبة بالشكل التالي:

$$AD = \frac{\sum f\left|X - \mu\right|}{N} \quad \text{for populations} \qquad (2.7a)$$

and

$$\mathrm{AD} = \frac{\sum f\left|\mathrm{X} - \overline{\mathrm{X}}\right|}{\mathrm{n}} \quad \text{for samples} \qquad (2.6b)$$

حيث f ترمز إلى تكرار الفئة و X ترمز إلى مركز الفئة.

مزايا وعيوب الانحراف المتوسط:

من مزاياه: – أنّه يأخذ كل القيم في الاعتبار

يعاب عليه: يتأثّر بالقيم الشاذّة – يصعب التعامل معه رياضياً.

4- متوسط الانحراف المطلق MEAN ABSOLUTE DEFIATION

- تعريف: يعرف متوسط الانحراف المطلق بأنّه متوسط مجموع حاصل طرح كل عنصر من عناصر العينة من متوسط العينة بالقيمة المطلقة. ومتوسط الانحراف المطلق.

- ومن الجدير بالتنويه إلى أنَّ متوسط الانحراف المطلق نادر الاستخدام في مجال الإحصاء التطبيقي في تحليل البيانات التجريبية للاختبارات والتجارب الزراعية.
- لذا اكتفينا بهذا القدر بالحديث عن متوسط الانحراف المطلق.

5- التباين:VARIANCE

- يعتبر التباين VARIANCE أهم مقياس من مقاييس التشتّت والمستخدمة على نطاق واسع جداً خاصة في تحليل البيانات التجريبية للاختبارات والتجارب الزراعية ويعبّر عن متوسط مربعات انحرافات القيم عن وسطها الحسابي.

أولاً: التباين في المجتمع (σ^2)

- يعطى بالعلاقة التالية: $\sigma^2=\frac{\sum(x-u)^2}{N}$ حيث أنَّ u هو الوسط الحسابي في المجتمع، أي أنَّ $u=\Sigma x/N$

ثانياً: التباين في العينة:

$$S^2=\left(\sum_{1}^{n}(X_i-\bar{X})^2\right)/(n-1)$$

- كما يمكن حساب التباين VARIANC بطريقة أخرى يطلق عليها طريقة تربيع القيم والتي تستخدم عندما تكون قيم العينة كسرية وهي كما يلي:

$$S^2=\left(\sum_{1}^{n}X_i^2-\left(\sum_{i}^{n}X_i^n\right)^2/n\right)/(n-1)$$

مثال (2): أوجد التشتت أو التباين والانحراف المعياري للعينة التالية باستخدام الطريقة المناسبة:

$X = \{5, 6, 7, 1, 3, 2\}$

الحــل: يمكن الاستعانة بالعلاقة الخاصة بحساب التباين الأولى الأنفة الذكر أو الثانية:

من أجل ذلك علينا إنشاء الجدول المبسّط التالي:

X	$X-X`$	$(X-X`)^2$	X^2
5	1	1	25
6	2	4	36
7	3	9	49
1	-3	9	1
3	-1	1	9
2	-2	4	4
∑ 24	0	28	124

$S^2 = 28 / (6-1) = 5.6$

6- الانحراف المعياري:

- لجأ الإحصائيون إلى الانحراف المعياري كمقياس منطقي يأخذ بعين الاعتبار الجذر التربيعي للتباين لكي يناسب وحدات قياس المتغيّر للعينة فهو سهل الحساب **باعتباره الجذر الموجب للتباين**، إذن هو يساوي:

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

- ومن نتيجة المثال السابق الانحراف المعياري

$$S = \sqrt{5.6} = 2.37$$

- في حالة البيانات المبوبة في جدول توز يع تكراري فإنه يعطى بالعلاقة التالية: حيث أنَّ f هي تكرار الفئة، x هو مركز الفئة، $\bar{X}$ هو الوسط الحسابي، n هي مجموع التكرارات

$$S = \sqrt{\frac{(\sum (X_i - \bar{X})^2 f}{n-1}}$$

- من الجدير بالذكر أنَّه يمكن الملاحظة بسهولة أنَّ الفرق بسيط بين قانون تباين العينة و قانون تباين المجتمع.
- عند تساوي المتوسطات فإنَّه يستخدم عادةً الانحراف المعياري كأساس للمقارنة بين عينتين أو أكثر شرط أن تكون العينات مأخوذة من مجتمع واحد،
- وبالتالي فالعينة ذات الانحراف المعياري **الأقل** هي التي تكون أكثر استقرارا من غيرها **وعناصرها أقل تبعثرا وأكثر تجانسا.**
- أمَّا إذا لم يتحقق الشرط المذكور أنفا أي عندما لا تكون العينات مأخوذة من مجتمع واحد فإنَّ الانحراف المعياري لا يمكن اتخاذه كأساس للمقارنة بين عينتين أو أكثر
- ويستخدم عادة مقياس آخر يطلق عليه **عامل الاختلاف** COFFICIENT OF VARIATION

عامل الاختلاف COFFICIENT OF VARIATION:

يرمز عادة لعامل الاختلاف بـ C.V

- وهو يعرف بأنه أحد مقاييس التشتت أو التباين الهامة والذي يستخدم بشكل عام للمقارنة بين عينتين أو أكثر عندما لا تكون العينات مأخوذة من مجتمع واحد بل من مجتمعات مختلفة، وهو أي عامل الاختلاف يعطى بالعلاقة التالية:

$$C.V = (S / X`) \times 100$$

حيث أنَّ: C.V : عامل الاختلاف

S : الانحراف المعياري

X` : متوسط العينة

من الجدير بالذكر أنَّه كلما كانت قيمة عامل الاختلاف C.V **أصغر** كلما دلَّ ذلك على ثبات العينة وتجانس أفرادها وعناصرها أقل تبعثرا مقارنة بالعينة الأخرى الأكبر بقيمة عامل الاختلاف

مثـــال: قارن بين العينتين التاليتين علماً أنَّهما من مجتمعين مختلفين:

X	3000	4500	4000	4100
Y	900	1200	1700	2200

الحــل: للمقارنة بين العينتين باعتبار أن العينات ليست مأخوذة من مجتمع واحد بل من مجتمعات مختلفة، فان عامل الاختلاف يعطى بالعلاقة التالية: **C.V =(S / X`) * 100**

وبتطبيق العلاقة السابقة الذكر نصل على قيمة عامل الاختلاف وهي:

C.V (X) = 637.7/3900 * 100 = 16 %

C.V (Y) =439.7/1500 *100 = 29 %

ومن الواضح أنَّ عامل الاختلاف للعينة الأولى أصغر من عامل الاختلاف للعينة الثانية وبالتالي فإنَّ العينة الأولى أكثر تجانساً من العينة الثانية.

الخطأ القياسي STANDARD ERROR

وهو يعرف بأنَّه أحد مقاييس التشتت أو التباين الهامة والذي يستخدم بشكل عام من أجل التعبير عن مدى التطابق ما بين متوسط الذي أخذت منه تلك العينة.

وهذا يعني أنَّه كلما كان الخطأ القياسي **أصغر** للعينة كلما كان الفرق أقل ما بين متوسط العينة ومتوسط المجتمع .

إنَّ الخطأ القياسي يعطى بالعلاقة التالية:

$$S_{\bar{X}} = S/\sqrt{N}$$

حيث أنَّ: $S\bar{X}$: الخطأ القياسي، S : الانحراف المعياري، N : عدد عناصر العينة

من الجدير بالذكر أنَّه يمكن استخدام الخطأ القياسي لتحديد المجال الذي يقع ضمنه متوسط المجتمع حيث أنَّ متوسط المجتمع يقع بشكل عام ضمن المجال:

$$\bar{X} \pm S_{\bar{X}} * t$$

ويطلق على هذا المجال كما هو معروف في الإحصاء مجال الثقة حيث t هنا هي درجات الثقة .

إلى اللقاء في المحاضرة القادمة