STAT – 110

ملخص إحصاء 110
لطلاب السنة التحضيرية – كلية العلوم

CH. 3

شرح مبسط باللغتين العربية والإنجليزية
أمثلة متنوعة مشابهة لأسئلة الاختبارات الماضية

إعـــــداد

أستاذ/ قاسم مضوي

جوال 0502180703

CHAPTER 3: DATA DESCRIPTION:

الباب الثالث: وصف البيانات: ويتكون من أربعة أقسام هي:

1- MEASURES OF CENRTRAL TENDENCY:

١ ـ مقاييس النزعة المركزية:

وهي ثلاث مقاييس (المتوسط والوسيط والمنوال)

1. Mean = average المتوسط الحسابي
2. Median الوسيط
3. Mode المنوال

MEAN: المتوسط الحسابي

sum of values divided by total number of that values.
It is used only with quantitative data.

تعريف المتوسط الحسابي: مجموع المفردات (المشاهدات) مقسوما على عددها.

متوسط العينة يرمز له **(X)** بينما متوسط المجتمع يرمز له **(u)**

EXAMPLE 1:
Find the mean 2, 9, 4
SOL.:

Mean = $\overline{X} = \frac{\sum x}{n} = \frac{2+9+4}{3} = \frac{15}{3} = 5$

EXAMPLE 2:
FIND THE VALUE OF X IF YOU HAVE 2, X, 9, the mean of these value equal to 5.

SOL.:

$\overline{X} = \frac{\sum x}{n}$ $\quad 5 = \frac{2+x+9}{3}$ $\quad 11+x=15$

$\therefore x = 15-11 = 4$

Median: الوسيط:

هو القيمة التي تتوسط البيانات بعد ترتيبها تصاعديا أو تنازليا

EXAMPLE 1:
Find the median to these values: 2, 5, 9, 7, 3
SOL.: first step we should arrange the value as sending or descending order.
Arranging data 2, 3, 5, 7, 9
Then the median equal to 5 (the value in the middle of values)

EXAMPLE 2:
Find the median to 2, 5, 9, 7, 3 , 12
Arranging data: 2, 3, 5, 7, 9, 12
Then the median equal to:

أستاذ/ قاسم مضوي – جوال 0502180703

MODE: المنوال

Is the value which having a higher frequency.

تعريفه: القيمة الأكثر تكرارا أو شيوعا.

EXAMPLE 1:
Find the mode: 6, 7, 2, 6, 5
SOL.: Mode = 6 (we say data having unimodal)

EXAMPLE 2:
Find the mode 6, 7, 2, 6, 5, 7
mode = 6, 7 (bimodal)

REMARKS:

ملاحظة مهمة جدا جدا:

إذا كان للبيانات منوال واحد البيانات تسمى وحيدة المنوال
أإذا كان للبيانات منوالين (٢ منوال) تسمى ثنائية المنوال.
إذا كان للبيانات أكثر من منوالين تسمى متعددة المنوال

one mode - - - unimodal
two mode ---- bimodal
three mode --- multimodal

أستاذ/ قاسم مضوي
0502180703

2- MEASURES OF VARIATIONS:
ثانيا: مقاييس الاختلاف:

المدى الربيعي midrange
التباين variance (تباين العينة يرمز له S^2)
الانحراف المعياري standard deviation (انحراف العينة يرمز له S)
معامل الاختلاف coefficient of variation (يرمز له C.V)
المقاييس أعلاه تسمى مقاييس الاختلاف.

Midrange: المدى الربيع

Midrange = highest value + lowest value / 2

EXAMPLE 1:
Find the midrange for 2, 9, 4

Midrange = $\frac{\text{highest value + lowest value}}{2} = \frac{2+9}{2} = 5.5$

VARIANCE: التباين

$$S^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} \quad \textbf{OR} \quad S^2 = \frac{\sum (X - \bar{X})^2}{n-1}$$

Where X is mean (average)

EXAMPLE 1:
Find the variance to 2, 4, 9
SOL:

أستاذ/ قاسم مضوي
0502180703

$$S^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$$

X	X2	
2	4	
4	16	
9	81	
Sum 15	**Sum 101**	

$$S^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{101 - \frac{(15)^2}{3}}{3-1} = 13$$

أستاذ/ قاسم مضوي – جوال 0502180703

STANDARD DEVIATION: الانحراف المعياري

تعريفه: هو الجذر التربيعي للتباين.

Definition: Is the square root of variance.

$$S=\sqrt{\frac{\sum x^2-\frac{\left(\sum x\right)^2}{n}}{n-1}}$$

EXAMPLE 1:

Find the standard deviation for 2, 4, 9

$$S=\sqrt{\frac{\sum x^2-\frac{\left(\sum x\right)^2}{n}}{n-1}}=\sqrt{\frac{101-\frac{(15)^2}{3}}{3-1}}=\sqrt{13}=3.61$$

COEFFICIENT OF VARIATION: معامل الاختلاف

Used to **compare** between two phenomena (i.e. age and salary).

مقياس يستخدم للمقارنة بين ظاهرتين مثل (المقارنة بين أعمار عمال ورواتبهم) وذلك بقسمة الانحراف المعياري للعمر على متوسط العمر ، ثم تقسيم الانحراف المعياري للرواتب على متوسط الرواتب ومن ثم المقارنة بينهما والقيمة الأكبر تكون هي الأكثر اختلافا.

$$C.V=\frac{S}{\overline{X}}\times 100\%$$

EXAMPLE 1:

Find C.V for 2, 4, 9.

We find X = 5, and S = 3.61 , then

$$C.V=\frac{S}{\overline{X}}\times 100\%=\frac{3.61}{5}\times 100=72.2\%$$

EXAMPLE 2:

The average age of the employees at a certain company is 30 years with standard deviation of 5 years ; the average salary of the employees is $40.000 with standard deviation of $ 5.000. which one has more variation age or income ?

age $C.V=\frac{S}{\overline{X}}\times 100\%=\frac{5}{30}\times 100=16.67\%$

salary $C.V=\frac{S}{\overline{X}}\times 100\%=\frac{5000}{40000}\times 100=12.5\%$

Then age has more variation.

أستاذ/ قاسم مضوي – جوال 0502180703

REMARKS: ملاحظة مهمة جدا

1. MEAN < MEDIAN < MODE

Then the distribution has a negative skewed or skewed to the left.

2. MODE < MEDIAN < MEAN

Then the distribution has a positive skewed or skewed to the right.

3. MODE = MEDIAN = MEAN

Then the distribution is symmetrical.

ملاحظة مهمة واحتمال ٩٩,٩% سؤال في الاختبار:

العلاقة بين المتوسط الحسابي والوسيط والمنوال.

١- إذا كان المتوسط الحسابي أقل قيمة من الوسيط والمنوال فإن التوزيع يتلوي جهة اليسار.

٢- إذا كان المتوسط الحسابي أكبر قيمة فمعنى هذا أن التوزيع يتلوي جهة اليمين.

٣- إذا تساوى المتوسط الحسابي والوسيط والمنوال فإن التوزيع يسمى متماثل.

3- MEASURES OF POSITIONS: ثالثا مقاييس المواضع

أستاذ/ قاسم مضوي
جوال 0502180703

1. STANDARD SCORE الدرجة المعيارية
2. QUARTILES الأرباعيات
3. OUTLIER القيمة المتطرفة

STANDARD SCORE:

$$Z = \frac{X - \bar{X}}{S}$$

EXAMPLE1:

If the mean of a set of data is 19 and 23.5 has z-sore of 0.75, then the standard deviation must be?

SOL.:

X = 23.5, and X = 19, Z = 0.75, S = ?

$$Z = \frac{X - \bar{X}}{S}$$

$$0.75 = \frac{23.5 - 19}{S}$$

$$0.75S = 4.5 \quad , \quad S = \frac{4.5}{0.75} = 6$$

EXAMPLE 2:

The mean of marks in test A is 80 and the standard deviation is 5,
if student get 70 in this test.
His relative position (z-score) in this exam is:

A) 6 B) 36 C)-3 D) -2

If the student in question **(EXAMPLE 2)** has a z-score = -1 in test B, one can say about his relative position that:
A- B better is better than A
B- information are not sufficient
C- Test (A) better than test (B)
D- Test (A) is similar to test (B)

2. QUARTILES الأرباعيات

ويقصد بها القيم التي تقسم البيانات إلى ثلاث أقسام وهي الربيع الأول Q1 والربيع الثاني Q2 وهو نفسه الوسيط ، والربيع الثالث Q3

Quartile divided the data into 4 groups.
Quartiles are denoted by Q1, Q2 , and Q3.
The second quartile (Q2) is median.

3. OUTLIER القيمة الشاذة

An outlier is an extremely high or an extremely how data value when compared with the rest of the data values.

(القيمة الشاذة هي أصغر قيمة أو أقل قيمة وواضحة جدا بالنسبة للبيانات المعطاة ويمكن استخراجها بسهولة من دون قانون أو قاعدة ، ولكن يوجد قانون للتأكد من أنها القيمة الشاذة أم لا).

Inter Quartile Range = IQR = Q3 – Q1

المدى الربيعي IQR

To determine whether a data value can considered as an outlier:
1. computer Q1 and Q3.
2. Find the IQR = Q3 – Q1
3. Compute 1.5 * IQR
4. Compute Q1 - 1.5 * IQR & Q3 + 1.5 * IQR

To determine whether a data value can be considered as an outlier.
5. Compare the data value with Q1 - 1.5 * IQR & Q3 + 1.5 * IQR

If x < Q1 - 1.5 * IQR or x > Q3 + 1.5 * IQR then x is considered as outlier.

EXAMPLE 1: Given the data set 5, 6, 12, 13, 15, 18, 22, 50 **Find the outlier ?**
المطلوب إيجاد القيمة الشاذة ؟

First of all we should arrange data , our data is already arranged.
Arranging of data : 5, 6, 12, 13, 15, 18, 22, 50
Q2 = (13+15)/2 = 14
Q1 = (6+12)/2 = 9
Q3= (18+22)/2 = 20
IQR = Q3- Q1 = 20- 9 = 11
1.5 * IQR = 1.5 * 11 = 16.5
Q1 – 1.5 * IQR = 9 – 16.5 = -7.5
Q3 + 1.5 * IQR = 20 + 16.5 = 36.5

Our range is -7.5 ——————————————— 36.5
Then 50 is out of this range, then 50 is outlier.

أستاذ/ قاسم مضوي – جوال 0502180703

4- EXPLORATORY DATA ANALYSIS:

Stem and leaf plot:الساق والأوراق

EXAMPLE 1:

If you have the following data 20, 32, 13, 14, 43, 02, 57, 23, 36, 32, 25, 31, 33, 32, 44, 32, 52, 44, 51, 45.

مثال ١: قم بإنشاء الساق والأوراق للبيانات السابقة

Construct a stem and leaf for the data.

SOL. الحل

First step we must arranging data as following: لازم نرتب البيانات ترتيبا تصاعديا كما يلي

02, 13, 14, 20, 23, 25, 31, 32, 32, 32, 32, 33, 36, 43, 44, 44, 45, 51, 52, 57

Stem	Leaf						
0	2						
1	3	4					
2	0	3	5				
3	1	2	2	2	2	3	6
4	3	4	4	5			
5	1	2	7				

BOX PLOT GRAPH:رسمة البوكس بولت

To draw Box plot graph we find Q1, Q2, and Q3 minimum and maximum value as following: لرسم الصندوق يجب إيجاد الربيع الأول والثاني والثالث

EXAMPLE 1:

Given the data set 5, 6, 12, 13, 15, 18, 22, 50 **construct Box plot ?**

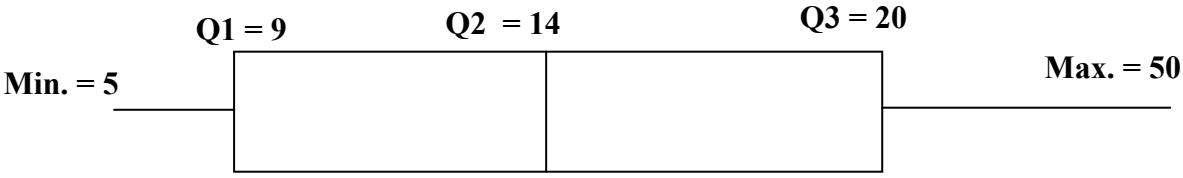
First of all we should arrange data , our data is already arranged.

Arranging of data : 5, 6, 12, 13, 15, 18, 22, 50

Q2 = (13+15)/2 = 14

Q1 = (6+12)/2 = 9

Q3= (18+22)/2 = 20



أستاذ/ قاسم مضوي – جوال 0502180703

REMARKS:ملاحظات مهمة جدا

1. **If the median is near the center of the box, the distribution is approximately symmetric.**
   إذا كان الوسيط Q2 في المنتصف فإن البيانات تعتبر متماثلة
2. **if the median fall to the left of the center of the box, the distribution is positively skewed.**
   إذا كان الوسيط Q2 في اليسار ، فإن التوزيع يعتبر ملتو جهة اليمين.
3.
4. **if the median fall to the right of the center of the box, the distribution is negatively skewed.**
   إذا كان الوسيط Q2 في اليمين ، فإن التوزيع ملتو جهة اليسار.

5. **If the lines are about the same length, the distribution is approximately symmetric.**
   إذا كان الخطان في طرفي الصندوق متساويان فإن التوزيع متماثل.

6. **If the right line is larger than the left line, the distribution is positively skewed.**
   إذا كان الخط يمين الصندوق أكبر من الخط يسار الصندوق فإن التوزيع متلو جهة اليمين.

7. **If the left line is larger than the right line , the distribution is negatively skewed.**
   إذا كان الخط يسار الصندوق أكبر من الخط يمين الصندوق فإن التوزيع ملتو جهة اليسار.

نهاية الفصل الثالث

والله ولي التوفيق

1431-1432H

أستاذ/ قاسم مضوي – جوال 0502180703