7

Ch. 3 - Part 3

Chapter Quiz.

- measures of Central tendency.
- measures of variation.
- measures of position.

STAT.110

جمـال السعـدي

رياضيات - إحصاء

CH.3 Part 3

جمـــال السعــدي

Chapter quiz

Determine whether each statement is true of false.

If the statement is false, explain why.

1) When the mean is computed for individual data.

All values in the data set are used. ☑

─────────────────────────────

2) The mean can not be found for grouped data.

When there is an open class. ☑

─────────────────────────────

3) A Single, extremely large value can affect.

The median more than the mean. العكــس صح ☒

─────────────────────────────

4) One – half of all the data values will fall above.

The (mode) and one- half will fall below the (mode) ✗ ☒
median ✓ ‌‌‌‌‌‌‌‌‌‌‌‌‌‌‌‌‌‌‌‌‌‌‌‌‌‌‌‌ median ✓

─────────────────────────────

5) In a data set, the mode will always be unique ☒

ربما لا يوجد متوال متوال وحيد
May be * there is no mode * unique mode
ثنائي عديد المتوال
* Bimodal * Multimodal

─────────────────────────────

6) The range and (midrange) are both measures of variation ☒

variation ليس ضمن مقاييس الـ

⑦ One disadvantage of the median is that it is not unique. ☒

عيوب ── ليس وحيد

Because: the median is unique. وحيد

────────────────────────────

⑧ The mode and midrange are both measures of variation ☒

✗   ✗

────────────────────────────

⑨ If a person's score on an exam corresponds to 75th percentile,

then that person obtained 75 correct answer

out of 100 questions ☒

May be 3 correct answers from 4 questions.

> **Note**
>
> • **Mean – Median – Mode :**
> **are measures of central tendency.**
>
> • **Range – Variance – Standard deviation**
> **are measures of variation.**
>
> • **Percentiles – quartiles – deciles:**
> **are measures of position.**

• The coefficient of variation CVar :

الدلالة

Show relation between mean and standard deviation.

Select the best answer:

(10) what is the value of the mode when all values in the data set are different ? اذا كانت كل البيانات مختلفه ← لايوجد منوال

(a) 0   (b) 1   (c) there is no mode

يصنف

(11) When data are categorized as, for example, places of residence (Rural , Suburban , Urban) , the most appropriate measure of central tendency is the

ريفي ضاحية متسوب للمدينة الملائم ـ المناسب
تزعة مركزية

(a) Mean   (b) Median   (c) Mode   (d) Midrange

يماثل ـ يقابل

(12) P_{50} Corresponds to $P_{50} = Q_2 = D_5$

(a) Q_2 (b) D_5 (c) IQR   (d) a and b are correct

---

(13) Which is not part of the five – number summary ?

(a) Q_1 and Q_3 (b) the mean   (c) the median

(d) the smallest and largest data values.

---

(14) A statistic that tells the number of standard devotions a data value is <u>above or below</u> the <u>mean</u> is called

(a) A quartile   (b) A Percentile   (c) A Z-score ← تعريف

(d) A coefficient of variation

Complete these statements with the best answer

(16) A measure obtained from sample data is called a (statistic)

الحروف μ , σ^2 , σ

(17) Generally, Greek letters are used to represent (parameters)
الحروف $\overline{X}$, S^2 , S
,and Roman letters are used to represent (statistic).

(18) The positive square root of the variance is called :

(The standard deviation).

* الجذر التربيعى الموجب للتباييم يسمى الد أنحراف المعيارى .

(19) The symbol for the population standard deviation is (σ).

(20) When the sum of the lowest data value and the highest data

value is divided by 2 , the measure is called (Midrange) $= \dfrac{L+H}{2}$

(21) If the mode is to the left of the medium and the mean is to the

right of the medium then the distribution is (Positive) skewed.

(22) An extremely high or extremely low data value is called

(an Outlier).

A
L
S
A
A
D
I

Home work

(23) For the values:

12 , 15 , 13 , 14 , 15 , 16 , 17 , 16 , 17 , 18

Find each of these:

1. mean   2. median   3. Mode   4. Midrange

5. Range   6. variance   7. standard devation

Exercises:

For these situations, state which measure of central tendency – mean, median , or mode – should be used.

a) The <u>most</u> typical case is desired.   (Mode)

b) The distribution is <u>open- ended</u>.   (Median)

c) There is an <u>extreme value</u> in the data set.   (Median)

d) The data are <u>categorical</u>.   (Mode)

e) Further statistical Calculation will be needed.   (Mean)

f) The values are to be divided into <u>two</u> approximately equal <u>groups</u>, one group containing the larger values and one containing the smaller values.   (Median)

Home work

(24) the distribution of the number of errors that 10 students made on a typing test is shown

Errors	Frequency
0-2	1
3-5	3
6-8	4
9-11	1
12-14	1

Find each of these:

1. Mean.

2. Model class.

3. Variance.

4. Standard deviation.

Note

- The <u>median</u> is used for an <u>open – ended</u> distribution.

- The <u>mode</u> is the only measure of central tendency that can be used when the data are <u>nominal</u> or <u>categorical</u>.

- When the distribution is <u>extremely skewed</u> the <u>median rather than</u>, the <u>mean</u> as measure of central tendency.

- In <u>box plot</u>, if the <u>median is near the center</u> of the box, the distribution is approximately <u>symmetric</u>.

- When <u>all the data</u> transformed into <u>z – scores</u> the resulting distribution will have

Mean = 0 $\quad$ and standard deviation = 1

Exercises:

1.) What is a z score or standard score?

- A Z-score tells how many standard deviations the data value is above or below the mean. (تعريف)

2.) Define: Percentile rank, percentiles and Deciles.

- Percentile rank: $\dfrac{(\text{number of values below } x) + 0.5}{\text{Total number of values}}$ 100%

- Percentiles: divide the data set into 100 equal groups.

- Deciles : divide the data set into 10 equal groups

3.) What is the difference between a percentage and a percentile?

مقياس نسبي

- A percentile: is a relative measurement of position.

- A percentage: is an absolute measure of the part to the total.

4.) Define quartile.

Position in fourths that a data value holds in the distribution Q.

5. what is the relationship between quartiles and percentiles ?

$$Q_1 = P_{25}, \quad Q_2 = P_{50} \qquad , Q_3 = P_{75}$$

6. What is a decile ?
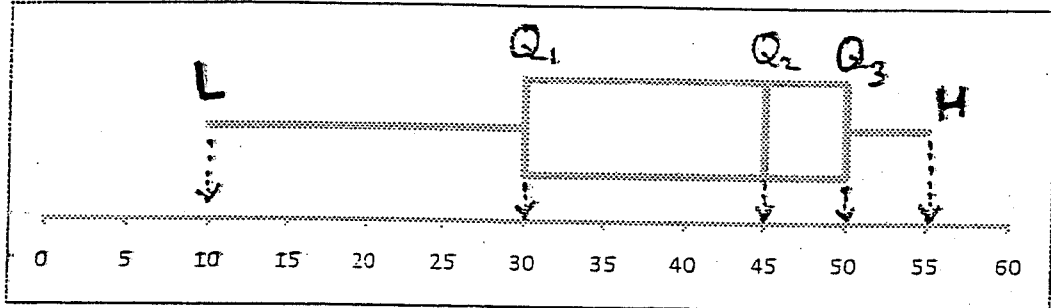
Position in tenths that a data value holds in the distribution D.

7. How are deciles related to percentiles ?

$$D_1 = P_{10} \quad , D_2 = P_{20} , D_3 = P_{30}, \ldots\ldots, D_{10} = P_{100}$$

8. To which <u>Percentile</u>, <u>Quartile</u> and <u>decile</u> does the median correspond ?

$$\text{Median} = P_{50} = Q_2 = D_5$$

Use the following boxplot to answer the following four questions:

The midrange value of the raw data for the above boxplot is ...
A) 42.5   B) 47.5   C) 32.5   D) 27.5

$$Midrange = \frac{L + H}{2} = \frac{10 + 55}{2} = 32.5$$

The mode value of the raw data for the above boxplot ...
A) is 30.   B) is 50.   C) can't be determined.   D) is 45.

mode الـ تعيين نمكن لا

من boxplot

The IQR value of the raw data for the above boxplot is ...
A) 40   B) 30   C) 20   D) 15

$$IQR = Q_3 - Q_1 = 50 - 30 = 20$$

What is the relationship among the mean, median and mode for the above boxplot?
A) The mean is the smallest value.   C) They are all equal.
B) Can't be determined.   D) The mean is the largest value.

* Left skew منحرف لليسار
mode أكبرها و mean هو المقاييس أصغر

Use the following to answer questions

The following table shows the distribution of the blood type for 32 students:

Classes	A	B	O	AB
Frequency	5	10	8	9

● The mean value
A) is B   B) is 8   C) is 2   D) cannot be calculated

The mean value can not be calculated

because : The data set are nominal .

* لا يمكن حساب قيمة الوسط الحسابي لأن البيانات أسمية

● The mode value
A) is AB   B) is 10   C) is B   D) cannot be calculated

المنوال
The mode : is $\dfrac{B}{\uparrow}$

الفئة المناظره لأعلى تكرار .

● Which measures of central tendency will always have unique values?
A) Mode and weighted mean.
B) Mode and median.
C) Mean and median.
D) Mean and mode.

* مقاييس النزعه المركزيه التي لها قيمه وحيده دائماً هي :
Mean and Median

* اما اختياره Mode لا يصلح لأن الـ Mode , ربما يكون

```
                    ┌→ no mode
                    ├→ unique
may be  ───────────→├→ bimodal
                    └→ multimodal
```

Use the following to answer questions

The following table shows the distribution of the blood type for 24 students:

Classes	A	B	O	AB
Frequency	8	3	6	7

The midrange value ...
A) is 5　　　B) is B　　　C) cannot be calculated　　D) is 1

midrange : cannot be calculated لا يمكن حسابه
لأن البيانات وصفية (الفئات ليست اعداد)

The mode value ...
A) is 8　　　B) is 6　　　C) cannot be calculated　　D) is A

المنوال هو الفئه ذات اكبر تكرار
mode = A

In a pie graph, how many degrees would be needed to represent the blood type A?
A) 120°　　　B) 28°　　　C) 105°　　　D) 90°

The number of degrees

$$= \frac{f}{n} \times 360°$$

$$= \frac{8}{24} \times 360° = 120°$$

● Which is <u>not</u> part of a five-number summary?

  A) The mean B) Q_1 and Q_3 C) The median D) The smallest and the largest data values

Five – number summary
are
Smallest data , Q_1 , Q_2 , Q_3 , Largest data.

The mean: is not part of
       five – number summary.

● Which is not part of a five-number summary?
  A) Q_2 B) The midrange C) The smallest and the largest data values D) Q_1 and Q_3

The midrange: is not part of

       five – number summary

● Which measures are mostly <u>affected by outliers?</u>
  A) Mean and median.
  B) Mean and IQR.
  C) Mode and median
  D) Mean and standard deviation

Mean and standard deviation

are affected by outliers, الوسط الحسابي
والانحراف المعياري أكثر تأثراً بالقيم الشاذه .

● Which measures are mostly affected by outliers?
  A) Mode and median. ✗
  B) Midrange and range. $\frac{H+L}{2}$ ✓ $H-L$ ✓
  C) Mode and midrange. ✗
  D) Mean and mode. ✗

Midrange and range
are affected by outliers

● Which measures are mostly affected by outliers?
  A) Mean and mode ✗
  B) Mode and midrange ✗
  C) Mean and midrange
  D) Mean and median ✗

✲ ملحوظه : الأكثر تأثراً بالقيم الشاذه هي المقاييس التي
يتم عليها عمليات حسابيه .

بالقيم الشاذه   لا يتأثر

● Which measures of central tendency are <u>not affected</u> largely by <u>outliers?</u>
  A) Mean and mode.
  B) Weighted mean and mean.
  C) Mean and midrange.
  D) Mode and median.

✲ أى ما من مقاييس النزعه المركزيه
لا تتأثر بالقيم الشاذه ؟

↳ _Mode and median_

المنوال والوسيط لا يتأثرام
بالقيم الشاذه لأنه لايتم عليها اى عمليات حسابيه .

● Approximately what percentage of normally distributed data values will fall within 1 standard deviation above or below the mean.
A) 68%   B) 95%   C) 99.7%   D) 13.5%

68% يكون الناتج 1 standard dev. لو كان *

حفظ ←
95% ″ ″ 2 ″ ″ ″ ″ *
99.7% ″ ″ 3 ″ ″ ″ ″ *

● Approximately what percentage of normally distributed data values will fall within 3 standard deviation above or below the mean
A) 95%   B) 99.7%   C) 68%   D) 34.1%

2 S   3 S   1 S

حفظ

3 standard deviation ⇒ 99.7 %

● Approximately what percentage of the standard normal distribution data values will fall between -2 and 2.
A) 99.7%   B) 13.5%   C) 95%   D) 68%

حل السؤال * Data values fall between -2 and 2 is 95 %

حالات أخرى ⌈ Data values fall between -3 and 3 is 99.7%

└ Data values fall between -1 and 1 is 68 %

* الحالات الثلاث السابقة حفظ -

● Which one of the following is referred to as a statistic?
  A) The sample mode   B) The sample data   C) The population mean   D) The population data

● Which one of the following is referred to as a parameter?
  A) Population mean   B) Population   C) Sample mode   D) Sample

$$\boxed{P}arameter \longrightarrow \boxed{P}opulation\ mean$$

● Which one of the following is referred to as a parameter?
  A) The population data.   C) The sample data.
  B) The sample variance.   D) The population variance.

★ Parameter (المعلمة)

population لم عباره عم مقياس لنصف المجتمع
وينتمي إلى

μ : mean

σ^2 : variance

σ : standard deviation

★ Statistic (الاحصاء)

sample لم عباره عم مقياس لنصف العينة
وينتمي إلى

$\bar{X}$: mean

S^2 : variance

S : standard deviation

معلومات
اجتماعية.

● This Pie graph represents snacks people eat at a sport game. What is the mode for this data

Super Bowl Snacks
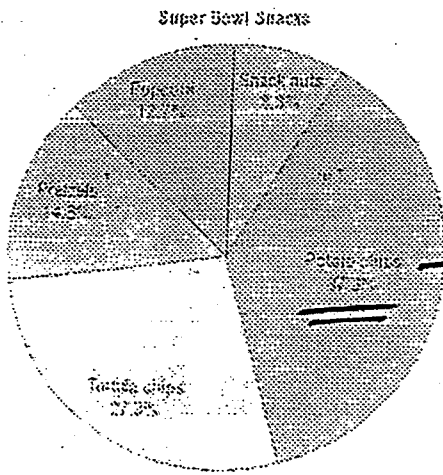
المنوال هو القطاع الذي له أكبر نسبة

∴ The mode is

potato chips

A) Potato chips   B) Tortilla chips   C) Pretzels   D) Popcorn

● What is the mean, median, mode of the following numbers? 1, 3, 6, 8, 12

A)   mean= 6   median= 6   mode= no mode
B)   mean= 5   median= 6   mode=1
C)   mean= 4   median= 6.5   mode= 3
D)   mean= 20   median= 3   mode=12

$$* \; mean \; = \; \frac{\Sigma x}{n} \; = \; \frac{1+3+6+8+12}{5} \; = \; 6$$

$$* \; median \; = \; 6$$

* القيمة التي تتوسط القيم بعد ترتيبها تصاعدياً

$$* \; mode \; : \; no \; mode \; \longleftarrow$$ لأن القيم مختلفة

● For a data set, half of the observations are always greater than the ...
A) mean.   B) weighted mean.   C) mode.   D) median.

* لمجموعة البيانات :
نصف عدد المشاهدات دائماً
أكبر من الوسط ← median
والنصف الأخر من المشاهدات اقل من الوسط

● If the standard deviation of a data set is 8.00, and a value $X=12.00$ has a z-score of 0.50, then the mean value

A) is 8   B) is -8   C) is 3.5   D) cannot be determined from the data given

$$\sigma = 8$$
$$X = 12$$
$$Z = 0.50$$

$$Z = \frac{X - \mu}{\sigma}$$

$$0.50 = \frac{12 - \mu}{8}$$

$$(0.50)(8) = 12 - \mu$$
$$4 = 12 - \mu \Rightarrow \mu = 8$$

● Find the average grade points for a student who has the following results:

Course	STAT	CHIM	BIO	ENG	ARAB
Grade	88	70	75	80	90
Credit	3	3	4	3	2

A) 73.33   B) 76.80   C) 80.40   D) 79.60

Average grade point

$$= \frac{88 \times 3 + 70 \times 3 + 75 \times 4 + 80 \times 3 + 90 \times 2}{3 + 3 + 4 + 3 + 2} = 79.6$$

Use the following to answer questions

أوزام

The weights (in grams) of the contents of several small bottles are 4, 2, 5, 4, 5, 2 and 4. Use this information to answer the following three questions:

● What is the value of the mean?

    A) 5.50   B) 4.57   C) 3.71   D) 4.00

أوزام لمحتويات عده زجاجات صغيره هي

4 , 2 , 5 , 4 , 5 , 2 , 4

$$mean: \quad \bar{X} = \frac{\Sigma x}{n} = \frac{26}{7} = 3.71$$

● What is the value of the standard deviation?

    A) 1.57   B) 0.80   C) 1.25   D) 0.89

$$S = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}} = \sqrt{\frac{106 - \frac{(26)^2}{7}}{7-1}} = 1.25$$

● What is the value of the coefficient of variation?

    A) 42.32%   B) 20.00%   C) 33.69%   D) 22.25%

Coefficient of variation

$$C\,Var. = \frac{S}{\bar{X}} \cdot 100\%$$

$$= \frac{1.25}{3.71} \cdot 100\%$$

$$= 33.69\%$$

ALSAADI

- A statistic that tells the number of standard deviations a data value is above or below the mean is called a ...
  A) coefficient of variation. B) z score. C) percentile. D) quartile.

Z score ⟶ {تعريف يحفظ}

: tells the number of standard deviations above or below the mean.

- If the value X=6 has a z-score of -0.50 and standard deviation 6 in a data set, then the mean value ...
  A) is 9.   B) is 12.   C) cannot be determined.   D) is -9.

$X = 6$ ، $Z\text{-score} = -0.50$ ، $\sigma = 6$

The mean : μ ?

∵ $Z = \dfrac{X - \mu}{\sigma}$ (قانون) ⟹ $-0.50 = \dfrac{6 - \mu}{6}$

⟹ $(-0.50) \cdot (6) = 6 - \mu$

$-3 = 6 - \mu$ ⟹ $\mu = 6 + 3$ ⟹ $\boxed{\mu = 9}$

- If the value X=6 has a z-score of -0.50 in a data set, then the mean
  A) is 5.50   B) is 6.00   C) is 6.50   D) cannot be determined from the given data

$Z\text{-score} = \dfrac{X - \mu}{\sigma}$ {لا يمكن ايجاد ال mean μ لأن المعادلة تحتوي على مجهولين μ, σ}

$-0.50 = \dfrac{6 - \mu}{\sigma}$

∴ mean can not be determined from the given data.

• If the mean of a set of data equals 18.00 and a value X=22.50 has a z-score of 0.65, then the variance must be:

A) 47.93   B) 73.94   C) 6.92   D) 3.24

$$Z - score = \frac{X - \mu}{\sigma}$$

$$0.65 = \frac{22.5 - 18}{\sigma} \implies \sigma = \frac{22.5 - 18}{0.65}$$

$$\implies Variance : \sigma^2 = \left(\frac{22.5 - 18}{0.65}\right)^2 = 47.928$$

• If a student scored X points on a test where the mean score was 82.4, the variance was 9, and the student's Z- score was 2.7 then X must be ...

A) 74.3   B) 10.4   C) 47.6   D) 90.5

$$Z - score = \frac{X - \mu}{\sigma}$$

$$2.7 = \frac{X - 82.4}{3}$$

$X = ?$
$\mu = 82.4$
$\sigma^2 = 9 \implies \sigma = 3$
$Z - score = 2.7$

$$\implies X - 82.4 = (3)(2.7)$$

$$\implies X = (3)(2.7) + 82.4 \implies X = 90.5$$

• What is the most appropriate measure of central tendency for the following data set?

A, C, B, B, C, A, B

A) Median.   B) Midrange.   C) Mean.   D) Mode.

٭ البيانات ليست رقميه (وصفيه)

∴ المقاييس المناسب لها هو ← Mode

انتشار

● The ... can be used to determine the spread of a data set.
A) (mode) and the standard deviation   C)   standard deviation and the (mean) ✗
B) (midrange) and the variance ✗   D)   variance and the range

تباین ✱

```
spread ≡ variation

Range — variance — standard deviatio
```

● The measures that can be used to determine the spread of a data set are the ...
A)   variance and the (mean) ✗   C)   variance and the (mode) ✗
B)   (midrange) and the standard deviation. ✗   D)   range and the standard deviation.

ليست ضمن

● If a data value is not within the range [Q_1 - 1.5(IQR), Q_3 + 1.5(IQR)], then this value is called ...
A) an outlier   B) the median   C) the third quartile   D) the first quartile

✱ القيم التي ليست ضمن الفترة
$$[Q_1 - 1.5(IQR), Q_3 + 1.5(IQR)]$$
تعرف بالقيم الشاذة

outlier

● If a data value is smaller than Q_1 - 1.5(IQR), this value is considered to be ...
A) the range.   B) the minimum.   C) a z-score.   D) an outlier.

✱ $Q_1 - 1.5(IQR)$ القيم التي هي أصغر من
✱ $Q_3 + 1.5(IQR)$ أو أكبر من
outlier   كلها تعرف بالقيم الشاذة

● A ... is referred to every measurement calculated for a study that is conducted on all students of KAU.
A) statistic.   B) population.   C) sample.   D) parameter.

المجتمع كاملاً
✱ المقياس الناتج من دراسة المجتمع كاملاً يسمى   parameter

● A ... is referred to every measurement calculated for a study that is conducted on a group of students from KAU.
A) sample.   B) population.   C) statistic.   D) parameter.

عينة من المجتمع
✱ المقياس الناتج من دراسة عينة من المجتمع يسمى   statistic

تعريفات مهمة

* Measures of average are also called <u>measures of central tendency</u> and include: the <u>mean</u>, <u>median</u>, <u>mode</u>, <u>midrange</u>, and <u>weighted mean</u>.

* <u>Measures of variation</u> such as the <u>range</u>, <u>variance and standard deviation</u> are used to describe <u>the spread</u> of data.

* The values that are <u>smaller than</u> $Q_1 - 1.5$ (IQR)

  or <u>larger than</u> $Q_3 + 1.5$ (IQR)

  are called <u>outliers</u>.

* The <u>weighed</u> <u>mean</u> is used when the values in a data set are <u>not equally represented</u>.

* A <u>statistic</u> is a characteristic or measure calculated using the data values of a <u>sample</u>.

* A <u>parameter</u> is a characteristic or measure calculated using all the data values of a specific <u>population</u>.

* <u>Variances and standard deviations</u> can be: used to determine the <u>spread</u> of the data. If the variance or standard deviation is large, the data are more dispersed.

Summarize data using measures of central tendency, such as the mean, median, mode, and midrange.

Describe data using the measures of variation, such as the range, variance, and standard deviation.

Identify the position of a data value in a data set using various measures of position, such as standard scores and quartiles.

Measures of average are also called measures of central tendency and include the mean, median, mode, midrange, and weighted mean.

When all the values in a data set occur with the same frequency is said to have no mode.

The midrange (MR) is a rough estimate of the middle and defined as the sum of the lowest and highest values in a data set divided by 2.

The mean cannot be computed for an open – ended frequency distribution.

The <u>mean is affected</u> by <u>extremely high or low</u> values and may not be the appropriate average.

The <u>median</u> is used to find the average of an <u>open-ended</u> distribution.

The <u>median</u> is affected <u>less than</u> the <u>mean</u> by <u>extremely high</u> or <u>extremely low</u> values.

The midrange is easy to compute.

The <u>midrange</u> gives the <u>midpoint</u>.

The midrange is affected by extremely high or low values in a data set.

<u>Large</u> coefficient of variation means <u>large variability</u>.

A <u>standard score</u> or <u>z score</u> is used when <u>direct comparison</u> of raw scores is <u>impossible</u>.

Boxplots are graphical representations of a five-number summary of a data set. Data set. The five specific values that make up five-number summary are minimum, Q_1, Q_2, Q_3 and maximum.

The range is the distance between highest value and lowest value.

The variance is the average of the squares of the distance between the mean and each value in a data set.

The standard deviation is the square root of the variance.