ALQUSOUR ACADEMY
اكاديمية القصور

Q.A.J.U.S.T

BIOSTATISTICS

#لجينة_كلية_الصيدلة

لطلبة الصيدلة والعلوم الطبية

Subject:

First Exam – Part Two

Prod. Date: **20/10/2015** Pages: **12** Price: **40**

ساعـــات الـــدوام الرسمـــي .....

السبت - الخميس: 11:00 ظهراً - 12:00 ليلاً

الجمعـــــــة: 2:00 ظهراً - 12:00 ليلاً

تحذير: محاضراتنا (الملخصات) متوفرة لدى أكاديمية القصور .

اربد / بجانب اربيلا مول / سلام سنتر / الطابق الرابع

للإستفسار والتسجيل: 0795 33 99 34 / 0785 70 60 08

Like facebook.com/alqusouracademy.page

ملاحظة: تعتمد المادة على السلايدات

Descriptive Statistics

❖ Mathematical presentation:
It includes Measures of Central Tendency and measures of variability.

First: Measures of Central Tendency

Definitions:
(a) Statistic:
Is a descriptive measure computed from the data of a sample.
(b) Parameter:
Is a descriptive measure computed from the data of a population.

Measures of Central Tendency:
- The following are called measures of central tendency *(measures of location):*
Mean, Median, Mode.

(1)The mean (arithmetic mean):

- Is the most commonly used measure of central tendency.
- It is the sum of the numbers divided by the sample size.

The sample mean of a sample of size n is:

$$\bar{x} = \frac{\sum_{i=1}^{n} xi}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

The mean of population with size N is: $\mu = \frac{\sum_{i=1}^{N} xi}{N}$

Example:

$let\ x_1 = 2, x_2 = 3, x_3 = 7, x_4 = 6, x_5 = 12$

$$Then\ \bar{x} = \frac{2+3+7+6+12}{5} = 6$$

- Mean is affected by extreme values (outliers).
- It has the same unit of original data.

(2) The Median:

It is a number such that half the numbers less than it, and half the numbers greater than it. When finding the median it is important to write the sample in the ordered form.

It is not affected by extreme values. (less sensitive)

Example:

Find the median of the following data sets:

(a) 2, 4, 3, 5, 8, 6, 8.

(b) 4, 6, 7, 8, 1, 2, 8, 9.

Solution:

(a) Remember to re-arrange first from the smallest value to the largest (ex. rewrite the sample in the ordered form):

2 3 4 5 6 8 8
↑

We can use the following formula: $\frac{n+1}{2}$ to find median **position**, then: $\frac{7+1}{2} = 4$

Thus, median position is 4, so the median is 5.

(b) First re-arrange, then we have 1 2 4 6 7 8 8 9

Here, the number of values is even so you can take the average of the middle two values: $\frac{6+7}{2} = \mathbf{6.5}$

---

(3) The Mode:

A mode is a value that occurs with maximum frequency.

The data may have exactly one mode (unimodal), two modes (bimodal), or more than two modes (multimodal).

Example:

Find the mode of the following data sets:

(a) 2, 3, 2, 4, 8, 2. (b) 1, 2, 3, 3, 2, 1, 1, 2. (c) 1, 3, 4, 7, 6, 10

Solution:

(a) mode = 2 (b) mode = 1 and 2. (c) No mode.

Example:

The following table contains a list of weights in Kg of 10 goats randomly selected from Talal farm in the city if Irbid.

15	16	22	15	13
12	25	16	14	35

Find:

(a) Find the mode and median of the weights.

Solution:

First we arrange the data:

12 13 14 15 15

16 16 22 25 35

Then:

mode = 15 and 16, median = 15.5

(b) The mean of the weights.

$$\bar{x} = \frac{12 + 13 + \cdots + 25 + 35}{10} = 18.3$$

Important:

If data includes outliers, then it is better to use median not mean.

Example:

Given the following data set, which is better - the mean or median?

2 4 5 8 9 12 100

Here, Mean =20 and Median = 8

Which is better to use in this case? Mean or Median?

Answer:

The median will give a better indication of center because the value 100 is an outlier and the median not affected by this value.

Second: Measures of Dispersion (Variability)

- It is the variation from the mean. (how far the values are from the mean)
- Variation gives information on the spread or variability of data.

Variability measures include:

(1) Range: largest value minus smallest value

$$R = xl - xs$$

: xl is largest value, xs: is smallest value.

- Range is very sensitive to the smallest and largest values (poor measure).

(2) The variance:

If we subtract the mean from each of its value, square the result and then adding the squared difference and divide that by the sample size minus one…. The result is equal to variance.

Sample variance for a sample of size n:

$$s^2 = \frac{\sum_{i=1}^{n}(xi-\bar{x})^2}{n-1}.$$

Note: Population variance is σ^2 $(sigma^2)$.

Note: (n – 1) is called degree of freedom.

(It means if we have the mean of 5 values for example, and we know 4 of them then we can find the fifth missing value).

Remark:

The sum of the **deviations** of the values from their mean is equal to zero :

$\sum(xi - \bar{x}) = 0$, for this reason we square it.

Example:

$let\ x_1 = 2, x_2 = 3, x_3 = 7, x_4 = 6, x_5 = 12$

be a sample. Then:

$$\bar{x} = \frac{2+3+7+6+12}{5} = 6$$

Now, $(2-6)+(3-6)+(7-6)+(6-6)+(12-6) = 0$

as we noted in the remark above.

Important:
We don't write a unit for the variance. (because it is squared)

(3) Standard deviation:
Standard deviation (s) for a sample:

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

Standard deviation (σ) for a population:

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

- Standard deviation is commonly used.
- It shows variation around the mean.
- Standard deviation has the same units as the original data.

Note: We can use calculator to calculate the mean and standard deviation. (see last page)

(4) Coefficient of variation: (it is strong measure)
Coefficient of variation (C.V) for a sample:

$$C.V = \frac{s}{\bar{x}}(100)$$

Properties of the Coefficient of variation:

1. Useful for comparing the variability of two or more variables.
2. It is independent of the unit of measurement.
3. Measures the relative variation.
4. always in percentage (%).

Example:
The following table contains a list of weights in Kg of 10 goats randomly selected from Talal farm in the city of Irbid.

15 16 22 15 13
12 25 16 14 35

Find:

(a) The range of the weights.

First we arrange the data:

12 13 14 15 15
16 16 22 25 35

$Range = 35 - 12 = 23$

(b) The mean of the weights.

$$\bar{x} = \frac{12 + 13 + \cdots + 25 + 35}{10} = 18.3$$

(c) The variance of the weights.

$$s^2 = \frac{(12 - 18.3)^2 + \cdots + (35 - 18.3)^2}{10 - 1} = 50.7$$

(d) The Coefficient of variation.

$$C.V = \frac{s}{\bar{x}}(100) = \frac{7.1}{18.3}(100) = 38.8$$

Percentiles:

Provides information about how the data are spread (variable) over the range of data.

The p^{th} percentile of a data set is a value such that at least p percent of the items take on this value or less and at least $(100 - p)$ percent of the items take on this value or more.

(5) Quartiles:

1^{st} Quartile (Q1) = 25^{th} Percentile,

2^{nd} Quartile (Q2) = 50^{th} Percentile = **median**,

3^{rd} Quartile (Q3) = 75^{th} Percentile.

- The first quartile, Q_1, is the value for which 25% of the observations are smaller than it and 75% are larger than it.
- Q_2 is the same as the median (50% are smaller, 50% are larger).
- Only 25% of the observations are greater than the third quartile.

Formulas to calculate Quartiles (it gives the position of the answer):

$$Q1 = \frac{n+1}{4}$$

$$Q1 = \frac{2(n+1)}{4}$$

$$Q1 = \frac{3(n+1)}{4}$$

Note: Calculation method is not mentioned in the slides.

---

(6) Interquartile range : IQR = Q3 – Q1

(7) Percent IQR = $\frac{IQR}{Range}$ **x 100**

Questions

1) The sum of the deviations of the values from their mean is equal to:

a. one.
b. the standard deviation.
c. zero
d. the mean.

2) The mean, median, and mode are the most common measures of dispersion.

a. True.
b. False

3) The only measure of central tendency that can be found for nominal data is the:

a. mean
b. median
c. mode
d. midrange

4) The measure of variation that is most affected by a few large or small numbers is:

a. mean
b. median
c. mode
d. range

The number of rooms for 15 homes recently sold were: 8 8 8 5 9 8 7 6 6 7 7 7 7 9 9 .

Answer the following 3 questions:

5) What is the mean?

a. 8.0
b. 7.0
c. 6.0
d. 9.0
e. 7.4

6) What is the median?

a. 8.0
b. 7.0
c. 6.0
d. 9.0

7) What is the mode?

a. 8.0
b. 7.0
c. 6.0
d. 9.0

8) The _____ is often the preferred measure of central tendency if the data have extreme values.

a. Mean
b. Median
c. Mode
d. Range

A researcher has collected the following sample data. 3 5 12 3 2
The mean of the sample is 5. (Answer Q9 to Q12)

9) The variance is:

a. 80
b. 4.062
c. 13.2
d. 16.5

10) The standard deviation is:

a. 8.944
b. 4.062
c. 13.2
d. 16.5

11) The coefficient of variation is:

a. 72.66%
b. 81.24%
c. 26.4%
d. 33.0%

Answers Key

Q	Answer
1	c. zero
2	b. False
3	c. mode
4	d. range
5	e. 7.4
6	b. 7.0
7	b. 7.0
8	b. Median
9	d. 16.5
10	b. 4.062
11	b. 81.24%
