Faculty of Computers &Artificial Intelligence
2nd Term (June 2021) Final Exam
Medical Informatics Program
Course Code: MBS312 **Level:** 3rd level
Subject: Biostatistics

Benha University
Date: 12 / 06 /2021
Time: 3 Hours
Total Marks: 50 Marks
Examiner : Dr. Mohamed Abdelgawad

Answer all the following questions [4 questions in 4 pages]

Question No. 1 [15 Marks]

A. To study the relationship between the student intelligence X and achievement in the biostatistics exam Y, the data were as follows:

$\Sigma X = 60$, $\Sigma Y = 70$, $\Sigma X^2 = 406$, $\Sigma Y^2 = 536$, $\Sigma XY = 374$, $N=10$.

Find the Correlation Coefficient between the two variables and determine its type. Also, find the equation of the regression line for the data to predict Y when X=20 ?.

B. Write briefly the characteristics of a good estimator and prove that the sample variance S^2 is an unbiased estimator of σ^2 ?

Question No. 2 [15 Marks]

A. Consumer reports tested n=15 brands of vanilla yogurt and found the following numbers of calories per serving: 160, 200, 220, 230, 120, 180, 140, 130, 170, 180, 80, 120, 100, 170, 190. Find the sample mean and standard deviation. We assume that the sample was taken from approximately normally distributed population. Calculate 95% confidence interval for the mean, $t_{(0.025,14)} = 2.145$?

B) A student received an **A** in Biostatistics (3 credits), a **C** in Mathematics (3 credits), a **B** in Machine Learning (4 credits), and a **D** in Databases (2credits). Assuming **A**=4 grade points, **B**=3grade points, **C**=2grade points, **D** = 1 grade point, and **F** = 0 grade points. Find the student's grade point Average (GPA).

Question No. 3 [10 Marks]

A. In two factories A and B located in the same industrial area, the Average weekly wages (in rupees) and the Variance are as follows:

Factory	Average	Variance
A	34.5	25
B	28.5	20.25

Which factory A or B has more variability (CVar) in individual wages?

B. An electronics company manufactures resistors that have a mean resistance of 100 ohms and a standard deviation of 10 ohms. The distribution of resistance is normal. Find the probability that a random sample of n= 25 resistors will have an average resistance less than 95 ohms.

Question No. 4 [10 Marks]

A. Consider the following data:18, 15, 12, 6, 8, 2, 3, 5, 20, 10

Find the percentile rank of 12.

B. Complete the blanks

Class	Frequency	Class boundaries	Class Midpoints	Cumulative Frequency	Percentage
100 – 104	2	……….	……….	……….	……….
105 – 109	8	……….	……….	……….	……….
110 – 114	18	……….	……….	……….	……….
115 – 119	13	……….	……….	……….	……….
120 – 124	7	……….	……….	……….	……….
125 – 129	1	……….	……….	……….	……….
130 – 134	1	……….	……….	……….	……….
Total	……….			……….	……….

Faculty of Computers &Artificial Intelligence
2st Term (June 2021) Final Exam
Medical Informatics Program
Course Code: MBS312 **Level:** 3rd level
Subject: Biostatistics

Benha University
Date: 12 / 06 /2021
Time: 3 Hours
Total Marks: 50 Marks
Examiner : Dr. Mohamed Abdelgawad

Model Answer

Solution Question No. 1 [15 Marks]

A. The linear correlation coefficient is given by

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

$$r = \frac{10(374) - 60(70)}{\sqrt{[10(406) - (60)^2][10(536) - (70)^2]}} = -1$$

The correlation coefficient is perfect negative. To get the equation of the regression line

$$a = \frac{\sum y \ \sum x^2 - (\sum x)(\sum xy)}{n \ \sum x^2 - (\sum x)^2} = 13$$

$$b = \frac{n \ \sum xy - (\sum x)(\sum y)}{n \ \sum x^2 - (\sum x)^2} = -1$$

$\boldsymbol{y'}=\boldsymbol{a}+\boldsymbol{b}X=13-X=13-20= -7$

B. 1- Characteristics of a good estimator: (unbiased)

The point estimator $\hat{\Theta}$ is an **unbiased estimator** for the parameter θ if

$$E(\hat{\Theta}) = \theta$$

If the estimator is not unbiased, then the difference

$$E(\hat{\Theta}) - \theta$$

is called the **bias** of the estimator $\hat{\Theta}$.

2- Characteristics of a good estimator: (with the least variance)

Suppose that $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimator of $\hat{\theta}$. Since $\hat{\theta}_1$ has smaller variance than $\hat{\theta}_2$. Then $\hat{\theta}_1$is minimum variance than $\hat{\theta}_2$ is the best.

To prove that the sample variance S^2 is an unbiased estimator of $\boldsymbol{\sigma}^2$

$$E(S^2) = E\left[\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}\right] = \frac{1}{n-1} E \sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$= \frac{1}{n-1} E \sum_{i=1}^{n}(X_i^2 + \overline{X}^2 - 2\overline{X}X_i) = \frac{1}{n-1} E\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i^2) - nE(\overline{X}^2)\right]$$

since $E(X_i^2) = \mu^2 + \sigma^2$ and $E(\overline{X}^2) = \mu^2 + \sigma^2/n$, we have

$$E(S^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n}(\mu^2 + \sigma^2) - n(\mu^2 + \sigma^2/n)\right]$$

$$= \frac{1}{n-1}(n\mu^2 + n\sigma^2 - n\mu^2 - \sigma^2)$$

$$= \sigma^2$$

Solution Question No. 2 [15 Marks]

A. The sample statistics were 159.3 for the sample mean and 43.5 for the standard deviation.

1- α=0.95→ α=0.05→ α/2=0.025,
Standard deviation= S = 43.5 ,n=15
95%confidence interval for μ is given by:

$t_{0.025,14} = 2.145$

$t_{0.025,14}(s/\sqrt{n}) = 2.145\ (43.5 / \sqrt{14}) = 11.6$

$159.3 \pm 2.145\ (43.5 / \sqrt{14}) \rightarrow$

$(159.3 - 24.94\ ,\ 159.3 + 24.94) \rightarrow (134.6,\ 184.24)$

B.

Course	Credits (w)	Grade (X)
Biostatistics	3	A (4 points)
Mathematics	3	C (2 points)
Machine Learning	4	B (3 points)
Databases	2	D (1 points)

$$\bar{x} = \frac{\sum \boldsymbol{w_i x_i}}{\sum \boldsymbol{w_i}} = \frac{3(4) + 3(2) + 4(3) + 2(1)}{3 + 3 + 4 + 2} = 2.7$$

The grade point average is 2.7

Solution Question No. 3 **[10 Marks]**

A. C.V.(A) = $\boldsymbol{\frac{\sigma}{\bar{x}}.100 = \frac{5}{34.5}.100 = 14.49}$

C.V.(B) = $\boldsymbol{\frac{\sigma}{\bar{x}}.100 = \frac{4.5}{28.5}.100 = 15.79}$

Factory B has greater variability in individual wages, since C.V. of factory B is greater than C.V of factory A.

B.

Note that the sampling distribution of $\overline{X}$ is normal, with mean $\mu_{\overline{X}} = 100$ ohms and a standard deviation of

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

Therefore, the desired probability corresponds to the shaded area in Fig. 7-7. Standardizing the point $\overline{X} = 95$ in Fig. 7-7, we find that

$$z = \frac{95 - 100}{2} = -2.5$$

and therefore,

$$P(\overline{X} < 95) = P(Z < -2.5) = 0.0062$$

Solution Question No. 4 **[10 Marks]**

A. Step 1. Ascendingly

2, 3, 5, 6, 8, 10, 12, 15, 18, 20

Step 2. Count how many numbers that are less than 12.

Step 3. Since we have 6 numbers < 12, then:

$$P = \frac{6 + .5}{10} . 100 = 65\text{th percentile}$$

The percentile rank of 12 is the 65th percentile.

B. Complete the blanks

Class	Frequency	Class boundaries	Class Midpoints	Cumulative Frequency	Percentage
100 – 104	2	99.5 – 104.5	102	2	4%
105 – 109	8	104.5 – 109.5	107	10	16%
110 – 114	18	109.5 – 114.5	112	28	36%
115 – 119	13	114.5 – 119.5	117	41	26%
120 – 124	7	119.5 – 124.5	122	48	14%
125 – 129	1	124.5 – 129.5	127	49	2%
130 – 134	1	129.5 – 134.5	132	50	2%
Total	50				100%

GOOD LUCK,

نموذج اجابة امتحان الاحصاء العضوية المستوى الثالث معلوماتية طبية برامج خاصة

كلية الحاسبات والذكاءالاصطناعى

د/ محمد عبدالجواد احمد عبدالجواد

مدرس – كلية العلوم – قسم الرياضيات

تاريخ الإمتحان ٢٠٢١/٦/١٢ الزمن ٣ ساعات