BHR 210/ BHM 310: Statistics for Researchers

QUESTION ONE

PART A

As part of the human resource department of Karatina University, you are tasked to summarize the

maximum educational level attained by each of the 900 employees at the institution. From the records, you

find that 200 have no college degree (None), 50- have a diploma (DIP), 200 have a bachelors degree (BA),

300 have a masters degree (MA), and 150 have a PhD.

a) Make a frequency table with two columns; Education Level and Number. (2 marks)

b) Make a relative frequency table. (2 Marks)

c) Is Education Level a quantitative or qualitative variable? Explain.

(1 mark)

d) Is Number a quantitative or qualitative variable? Explain. (1 marks)

e) Draw a bar-graph to represent the above information. (2 marks).

NB: The x-xis should represent education level with number on the y-axis.

f) What is the probability that an employee of Karatina University picked at random would be a holder of a diploma qualification?

PART B

The table below shows the corresponding height(in cm) and weight (in kilograms) of a sample of 10 traders at Karatina open air market.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Name | Wairimu | Oloo | Kiptoo | Ann | Etyang | Paul | Jane | Joan | Saul | Carol |  |
| Height | 165 | 172 | 167 | 179 | 150 | 140 | 180 | 200 | 176 | 220 |  |
| Weight | 67 | 73 | 69 | 79 | 64 | 50 | 78 | 84 | 80 | 90 |  |

a) Compute the arithmetic mean for the height and weight of the traders. (1 mark)

b) Compute the median height and weight of the traders. (1 mark)

c) When is the median an appropriate measure of center than the arithmetic mean? Explain. (1 mark)

d) Draw a scatter plot of the height (x - axis) and weight (y -axis). (4 marks)

e) Compute the standard deviation for (i) height and (ii) weight of the traders. (2 marks)

f) Compute the correlation coefficient between height and weight of the traders. (3 marks)

g) Write a sample R code that you would use to capture the height of the traders. (2 marks)

h) Write a sample R code that you would use to compute the mean of the height (1 mark)

QUESTION TWO

A tire manufacturer believes that the tread life of its snow tires can be described by a Normal model with a

mean of 32,000 miles and a standard deviation of 2500 miles.

a) Compute the z-score for a tire that lasts for 40,000 miles. (3 marks)

b) Approximately what fraction of these tires can be expected to last less than 30,000 miles? (6 marks)

c) Approximately what fraction of these tires can be expected to last between 30,000 and 35,000 miles? (6 marks)

QUESTION THREE

A recent study of Kenya Revenue Authority (KRA) audits showed that, for estates worth less than 15 million, about 1 out of 7 of all estate tax returns are audited, but that probability increases to 0.5 for estates worth over Ksh. 15 million. Suppose a tax accountant has three clients who have recently filed returns for estates worth more than Ksh. 15 million. What are the probabilities that:

a) All three will be audited? (3 marks)

b) None will be audited? (3 marks)

c) At least one will be audited? (5 marks)

d) What did you assume in calculating these probabilities? (4 marks)

QUESTION FOUR

Is there a relationship between total team salary and the performance of teams in the Kenya National Soccer

League (NFL)? For the 2019–2020 season, a linear model predicting Wins (out of 16 regular season games)

from the total team Salary (Ksh. millions) for the 20 teams in the league is:

Wins = -16.32 + 0.219Salary

a) What is the explanatory or independent variable? (1 mark)

b) What is the response or dependent variable? (1 mark)

c) What does the slope mean in this context? (2 marks)

d) What does the y-intercept mean in this context? Is it meaningful? Explain. (4 marks)

e) If one team spends Ksh. 10 million more than another on salary, how many more games on average would

you predict them to win?

(3 marks)

f) If a team spent ksh. 120 million on salaries and won 8 games, would they have done better or worse than predicted? (4 marks)

QUESTION FIVE

a) Distingusih between a sample and population.

(3 marks)

b) Compare and contrast the arithmetic mean and the geometric mean, giving situations where each is a

more appropriate measure of center.

(4 marks).

c) List and discuss four sampling techniques that researchers use to select a sample from the population (8

marks).