Statistics Sample Revision Questions

John Karuitha

5/12/2021

QUESTION ONE

The Data

You are provided with the following dataset for the average amount of meat in kgs and the weight for 10 individuals.

REQUIRED:

1. A scatter plot of weight (y-axis) versus meat intake (x -axis).
2. A comment on the relationship between weight and meat intake.
3. The arithmetic mean, geometric mean and harmonic mean of weight and meat intake. When is each of these means appropriate?
4. The median, Q1, Q3, and Interquantile range for weight and meat intake.
5. The variance, standard deviation of weight and meat intake.
6. The correlation coefficient between weight and meat intake.
7. A regression equation with weight as the dependent variable (y) and meat intake as the independent variable x.
   - $y = \beta_0 + \beta_1 x$.
   - $weight = \beta_0 + \beta_1 Meat_i ntake$.
8. Standardize both the weight and meat intake variables (z-scores).
9. 3 separate lines of R code that capture individual, Weight and meat intake using the `c` function.
10. Use the data.frame function to combine the three variables above into a data table.
11. Use R to compute the metrics you have manually computed in questions 1-7.

QUESTION TWO

People who eat a lot of meat have a high risk of developing heart disease. A person who eats 500g of meat daily has a 70% (0.5) chance of getting heart disease. A doctor meets 3 patients in a day each of whom consumes 500g of meat daily. What is the probability that the three will develop heart disease? What assumption did you make?

QUESTION THREE

Suppose you take the average score in the statistics exam for a group of students. You find that the data is normally distributed with the average score at 60 marks with a standard deviation of 10 marks. What is the probability that a student picked at random will have more than 70 marks? What is the probability that a student picked at random fails, that is, scores less than 40 marks?

Table 1: Height vs Meat Intake for 10 Individuals

Individual	Meat_intake	Weight
1	5.0	57
2	7.5	60
3	2.0	55
4	0.0	58
5	15.0	100
6	10.0	80
7	3.0	48
8	20.0	120
9	4.0	50
10	12.0	48

* This is a fictitious dataset and should not be taken as medical proof

QUESTION FOUR

Be sure to comfortable with the theory and definitions over and above the computations and their interpretations.