Questions: Introduction to data analysis

Michelle Arnetta

Summary

A selection of questions for the study guide on introduction to data analysis.

Before attempting these questions, it is highly recommended that you read Guide: Introduction to data analysis.

Q1

Cantor's Confectionery records how many customers had entered its store for the past 6 days. This can be summarized by the following list: $\{17, 91, 23, 18, 27, 22\}$.

1.1 Find the mean of this data set.

1.2 Find the median of this data set.

1.3 Why would it be important to find the median of this data set, instead of only finding the mean?

1.4 Why would it not be important to find the mode of this data set?

1.5 Can you think of any potential reasons for the existence of the outlier?

Q2

For each data set below, find the mean, median, mode (if possible), range, and interquartile range.

2.1 $\{14, 71, 82, 39, 39, 71, 71, 71, 48\}$

2.2 $\{919, 3293, 912, 7775\}$

2.3 $\{3, 72, 800, 8, 763, 8, 3, 9028, 763, 39, 3\}$

Q3

Examine the following scenarios, and identify which type of data visualization would be most suitable for each scenario. Your options include: bar chart, line graph, histogram, pie chart, and scatter plot.

3.1 A business wants to visualize how the budget allocations breakdown of each department compares to the totality of the budget.

3.2 A florist wants to visualize the distribution of flower stem lengths.

3.3 A teacher wants to visualize their students' exam scores against the number of hours they studied.

3.4 A bakery wants to visualize the trends in its chocolate bread sales revenue over time.

Q4

For each scenario below, determine which of the following options would be most appropriate to use: correlation test, regression analysis, confidence interval, PDF, PMF, and CDF.

4.1 A researcher wants to test whether there is a relationship between cholestrol levels and heart disease risk.

4.2 A researcher wants to model the relationship between cholestrol levels and heart disease risk in terms of a linear function.

4.3 You flip a coin three times. Each flip has a 0.5 chance of the coin landing on 'tails'. You know that this can be modelled as a Binomial distribution, and you want to find the probability that the coin lands on 'tails' all three times.

4.4 A biologist collects a random sample of 100 birds and calculates their mean wingspan. They want to find a range of values in which the mean wingspan of all birds lies, with 90% confidence.

After attempting the questions above, please click this link to find the answers.

Version history and licensing

v1.0: initial version created 9/25 by Michelle Arnetta as part of a University of St Andrews VIP project.