***ASSIGNMENT-4(vchalla3)***

A graph with a square and line

Description automatically generated

A boxplot is a way of visualizing the distribution of a set of data. The horizontal line in the middle of the box represents the median, which is the value that splits the data in half, with half the values above the median and half the values below it. The edges of the box represent the quartiles, which are the values that split the data into quarters. The whiskers extend from the quartiles to the most extreme values in the data set, excluding any outliers.

The boxplot for age shows that the median age is 40 years old. The quartiles are 30 and 50 years old, meaning that the middle 50% of the people in the data set are between the ages of 30 and 50. The whiskers extend to 20 and 60 years old, meaning that the most extreme values in the data set are 20 and 60 years old.

The boxplot for BMI shows that the median breast size is 30. The quartiles are 25 and 35, meaning that the middle 50% of the people in the data set have BMI between 25 and 35. The whiskers extend to 15 and 45, meaning that the most extreme values in the data set are 15 and 45.

The box plot for Children shows that the median duration is 40. The quartiles are 30 and 50, meaning that the middle 50% of the people in the data set have Children between 30 and 50. The whiskers extend to 20 and 60, meaning that the most extreme values in the data set are 20 and 60.

It is difficult to say much more about the data without knowing what the variables represent. However, the boxplots do give us some information about the distribution of the data, such as the median, quartiles, and range.

A group of blue bars

Description automatically generated

The image shows four histograms, one for each of the variable’s age, BMI, children, and expenses. A histogram is a graph that shows the distribution of a set of data. It is a type of bar chart that shows the number of observations in each bin (range of values).

The histogram for age shows that many people in the data set are between the ages of 20 and 50. There are also a significant number of people under the age of 20 and over the age of 50.

The histogram for BMI shows that many people in the data set have a BMI between 20 and 30. There are also a significant number of people with a BMI below 20 and above 30.

The histogram for children shows that many people in the data set have 0 or 1 child. There are also a significant number of people with 2 or 3 children.

The histogram for expenses shows that many people in the data set have expenses between $0 and $10,000. There are also a significant number of people with expenses between $10,000 and $20,000, and between $20,000 and $30,000. There are also a few people with expenses over $30,000.

Overall, the histograms show that the data set is relatively diverse, with a wide range of values for each of the four variables.

Here are some specific observations about the histograms:

* The age histogram is right skewed, meaning that there are more people in the younger age groups than in the older age groups.
* The BMI histogram is also right skewed, meaning that there are more people with a lower BMI than with a higher BMI.
* The children histogram is left-skewed, meaning that there are more people with 0 or 1 child than with 2 or more children.
* The expenses histogram is also left-skewed, meaning that there are more people with lower expenses than with higher expenses.

A blue bar graph with white text

Description automatically generated

The image you sent contains three bar plots:

* Bar Plot of Gender: This plot shows the number of people in the dataset who are male and female. There are slightly more female (500) than male (400) people in the dataset.
* Bar Plot of Smoker: This plot shows the number of people in the dataset who are smokers and non-smokers. There are more non-smokers (800) than smokers (300) in the dataset.
* Bar Plot of Region: This plot shows the number of people in the dataset who live in the Southeast, Southwest, Northwest, and Northeast regions. The Northeast region has the most people (350), followed by the Southeast (250), Southwest (150), and Northwest (100) regions.

A group of graphs with different sizes

Description automatically generated with medium confidence

The age distribution is shown in the first two charts. The first chart shows the overall age distribution, while the second chart shows the age distribution for children. The overall age distribution is skewed to the younger side, with a median age of around 30 years. The children's age distribution is also skewed to the younger side, with a median age of around 5 years.

**BMI distribution**

The BMI distribution is shown in the third chart. The BMI distribution is approximately normal, with a mean BMI of around 25. This suggests that the population is slightly overweight on average.

**Children distribution**

The children’s distribution is shown in the fourth chart. The children distribution is skewed to the left, with a median of 1 child per household. This suggests that most households have 1 or fewer children.

**Expenses distribution**

The expenses distribution is shown in the fifth chart. The expenses distribution is skewed to the right, with a median of around $20,000. This suggests that most households have expenses of $20,000 or less, but there are a few households with much higher expenses.

**Overall**

The overall distribution of age, BMI, children, and expenses in the population is shown in the image. The age distribution is skewed to the younger side, the BMI distribution is approximately normal, the children distribution is skewed to the left, and the expenses distribution is skewed to the right.

R-squared : 0.7608041536118314

MSE: 0.009020779497640052