Conclusion – Univariate Insights

Age

* Age ranges from 18-64, with mean and median very close in value (mean = 39.207, median = 39.000). This indicates near zero skewness.
* There are no outliers.
* Q3 is 51 which means that 75% of customers are below age 51.

BMI

* BMI ranges from 15.96-53.13, with mean and median very close in value (mean = 30.663, median = 30.400). This indicates near zero skewness.
* There are outliers for this variable.
* Q3 is 35 which means that 75% of customers have a BMI below 35.

Number of Children

* Number of children ranges from 0-5 with mean and median very close in value (mean = 1.095, median = 1.000). This indicates near zero skewness.
* There are no outliers.
* Q3 is 2 which means that 75% of policy holders have less than 2 children.

Charges

* Charges range from $1,121.87-$63,770.43, a wide range. With the mean ($13,270.42) greater than the median ($9,382.03), the data is right-skewed.
* There are outliers for this variable all on the higher end (above $35,000).
* The mean is greater than the median, therefore the data is right-skewed.
* Q3 is $16,639.91. 75% of customer have been charged less than the value of Q3.

Sex

* More males are policy holders (676) compared to females (662). There is a 1% difference (49.5% vs. 50.5%).

Smoker

* Most policy holders do not smoke (1064). Non-smokers greatly outnumber smokers (79.5% of policy holders do not smoke).

Region

* Most policy holders live in the southeast region of the U.S. (364). The remainder of the policy holders are evenly split between the remaining regions.

Conclusion – Insights from Multivariate Analysis and Hypothesis Testing

* As age increases, charges tend to increase.
* Smokers tend to have higher charges. Based on hypothesis testing we can say that the medical claims of smokers are greater than those of non-smokers.
* Based on hypothesis testing we know that the proportion of smokers does not differ across the various regions.
* The BMI of smokers increases along with charges (significant jump at BMI = 30).
* Based on hypothesis testing we know that BMIs do not significantly differ based on gender.
* Based on hypothesis testing we know that BMI for women does not differ according to the number of children the policy holder has.