Question 14.

1. The linear model built is below:

A screenshot of a computer

Description automatically generated

From above, to check whether the interaction term is necessary for the model or not.

Hypothesis:

H0: β2 = 0

H1: β2 != 0

As, p-value for beta 2 is 0.01<0.05, at 5% level of significance, it is evident that the interaction term is useful for the model. Overall, the f-statistic of 120.1 on 3 and 617 DF with p-value 2.2e-16 < 0.05, at 5% level of significance the model seems useful in predicting the monthly consumption of antibiotics with 36% of the variation of monthly consumption being explained by the model.

1. Prediction equation with 95% confidence is below:

A close-up of a computer code

Description automatically generated

We are 95% confident, that the average monthly consumption of antibiotics will be in between 1.10 and 12.94 per 1000 inhabitants per day for a county with 10% of the population over 25 years old in a metropolitan area who completed higher education.

1. Actual independent variable:

1. Scatter plot:

A graph of blue dots

Description automatically generated

The scatter plot between population density and average monthly antibiotic consumption seems to be not correlated with no direction and no strength and 1 outlier.

1. Linear model:

A screenshot of a computer error

Description automatically generated

For actual independent variable, the model seems adequate with p-value 0.0004 < 0.05 and f-statistic value of 12.58 on 1 and 619 DF.

1. Residual plot:

A graph of a plot

Description automatically generated

The residual of the model seems to be within the +-2 standard deviation with few outliers above 2s.

2. Log of independent variable:

1. Scatter plot:

A graph of blue dots

Description automatically generated

The above scatter plot for log shows a strong correlation with a positive direction and good strength and 1 outlier around 40.

1. Linear model with log of independent:

A screenshot of a computer

Description automatically generated

The log model with f-statistic 138.9 on 1 and 619 DF and p-value 2.2e-16 < 0.05 seems to be adequate model.

1. Residual plot:

A graph of a number of dots

Description automatically generated

The residual seems to be within +- 2s with 1 outlier above 2s while the rest seems to be close to mean value.

3. Square root of independent variable:

1. Scatter plot of sqrt:

A diagram of a number of blue dots

Description automatically generated

The above scatter plot shows a weaker correlation with 1 outlier and positive direction, the strength is not strong.

1. Linear model: A screenshot of a computer

   Description automatically generated

The linear model with sqrt of independent variable shows that f-statistic of 43.35 on 1 and 619 DF with p-value 3.67e-11 < 0.05 seems to be adequate model for predicting average consumption.

1. Residual plot:

A graph of a number of dots

Description automatically generated

The above residual plot for sqrt shows that most of the residuals are within 2 standard deviations while there are few outliers present in the data.

Final answer: After comparing the above models with and without transformations, the best model is the model with log of independent variable with highest f-statistic value, low p-value, stronger correlation and least outlier in residual plot.