Andrew Gordon

Professor Mike Nelson

ECO 602

19 September 2021

Week 3 Reading Questions

Q1:

Scatterplot, Cleveland dotplot, QQplot, coplot

Q2:

Histogram, box plot

Q3:

A conditional variable is usually mentioned in the context of a coplot, which is a conditional scatterplot that shows the relationship between two variables x and y. A coplot holds several scatterplots to show how the x-y relationship changes when a third variable in the situation, z, is manipulated. This new, manipulated variable is the conditional variable. An example of this would be the effect of temperature (z) on the relationship between bird and fish populations (x & y).

Q4:

Measures of spread or dispersion:

* Variance
* Standard deviation
* Coefficient of variance
* Median absolute deviation
* Range
* Interquartile range

Q5:

1. Standard deviation represents how far apart most values are from the mean of the dataset. A low standard deviation means that the majority of the values are found closer to the mean.
2. Interquartile Range represents the range of data values between the 25th and 7th quantiles. This range starts and ends at the middle of the lower and upper halves of the dataset respectively.

Q6:

One dataset that I have collected counts the number white tailed deer observed from the roadside at different times of the day.

An important reason to perform data exploration on this dataset would be to determine if deer are more active during a particular hour (day or night). This can be represented using a histogram, which shows frequencies (number of deer observed) according to a specific variable (time of day).

A second reason for data exploration would be to see if the relationship between time of day and deer activity is impacted by a conditional variable such as temperature or weather. A coplot can be used to identify if changes in a third variable (temperature in Celsius) has an effect on the frequency of deer found at particular times of day.