Module 9 prep guide

Mason Deja

7/3/2107

BIVARIATE EDA – QUANTITATIVE

1. Bivariate data is when two variables are measured on the same individuals.
2. The response variable is the variable that one is interested in explaining something. The synonym for the response variable is the dependent variable.
3. The explanatory variable is used to help explain or allow one to predict the response variable. The synonym for the response variable is the independent variable.
4. The x-coordinate is the explanatory variable and the y-coordinate is the response variable.
5. The symbol used for the sample correlation coefficient is r. The symbol used for the population correlation coefficient is r.
6. The four things described for a bivariate EDA with two quantitative variables are form of the relationship, presence (or absence) of outliers, association or direction of the relationship and strength of the relationship.
7. Form determines if the “cloud” of points forms a line or some sort of curve, this is also referred to being linear or nonlinear.
8. An outlier looks like a dot that is not near any other dots or simply looks out of place.
9. The words used to describe association are positive, none and negative. A positive association resembles an increasing function which goes from lower left to upper right and most of the individuals are above average or below average for both variables. A negative association resembles a decreasing function which goes from upper left to lower right. No association has no rhyme or reason to the points on the scatter plots.
10. The sign of the correlation coefficient tells us if the scatter plot is positive, negative or no association. It is also a measure of strength.
11. The magnitude of the absolute value shows the strength of the scatterplot. The closer to 1 indicates a stronger association.
12. Correlation coefficient values are always between -1 and 1.
13. You shouldn’t use the correlation coefficient if the relationship isn’t linear because the R is meaningless
14. In an observational study and experimental studies the correlation coefficient is used to determine the strength of the relationship. A strong relationship will have an absolute value of >0.8 for an observational study and >0.95 for controlled experiments. For moderate relationships its >0.6 for observational and >0.9 for controlled. Weak relationships will have an absolute value of >0.4 for observational and >0.8 for controlled experiments.
15. “Correlation is not causation” means just because there is a strong correlation doesn’t mean that the explanatory value caused the response variable.