Mod 11 prep guide

Mason Deja

7/16/17

Linear regression

1. The two goals of linear regression analysis is used to explain variability in the response variable and predicting future values of the response variable.
2. The response variable or the dependent variable is to be predicted or explained and the explanatory or independent variable is used for predicting or explaining.
3. The response variable is on the y axis and the x is the explanatory variable.
4. The generic equation of a line is y=mx+b. the y is the response variable, x is the explanatory variable, m is the slope of the line and b is the y intercept.
5. The slope is the change in value of the response variable for a single unit change in the value of the explanatory variable. The intercept is the value of the response variable when the explanatory variable is equal to zero.
6. The statistical word that must be used when describing slope and intercept is average or on average. The reason you have to use average is because interpretations for the slope and intercept represent the average change.
7. You can predict the value of the response variable by adding the equation of the line and the value of the explanatory variable and then multiplying it by a value on the y axis.
8. An extrapolation is when you make a prediction outside of the range of observed values of x.
9. The residual is a measure of how far off the prediction is from what is actually observed.
10. The RSS stands for residual sum of squares which says the best fit line is the one line out of all possible line that has the minimum RSS.
11. The two assumptions that will be used mostly in this course is 1) a line describes the data and 2) homoscedasticity.
12. 4
13. The coefficient of determination is the proportion of the total variability in the response variable that is explained by knowing the value of the explanatory variable and the best fit model. The symbol used is r2.
14. Values of coefficient of determination that have values closer to 1 signify a stronger relationship.
15. The 12 questions are what is the response variable, what is the explanatory variable, comment on the linearity and homoscedastity, what is the equation of the best fit line, interpret the value of the slope, interpret the value of the intercept, make a prediction given the value of the explanatory value, compute a residual given the values of both the explanatory and response variables, identify an extrapolation in the context of a prediction problem, What is the proportion of variability in the response variable explained by knowing the value of the explanatory variable, What is the correlation coeﬃcient and How much does the response variable change if the explanatory variable changes by X units?