

Stat 150 – Day 30
Statistical Significance with Regression Models

Recall that we studied the concept of *statistical significance* in the context of comparing two groups. We assessed statistical significance by calculating a *p-value*.

a) Fill in the blanks to interpret what a p-value is: The p-value is the probability of

_____ ,

assuming that _____ .

b) Does a small p-value provide evidence that the groups differ significantly, or that they do not differ significantly? Explain.

c) In the context of simple linear regression, a small p-value provides evidence of what?

d) In the context of multiple linear regression, a small p-value for a particular predictor variable provides evidence of what?

e) In the context of simple or multiple linear regression, what can you conclude from a not-so-small p-value?

Example 1: House Prices

Reconsider the data on house prices in Arroyo Grande in 2006.

a) Again fit a regression model for predicting house price from size (Stat> Regression> Regression). Report the p-value, and summarize what you can conclude from it.

b) Repeat for each of the other two predictor variables (number of bathrooms, number of bedrooms). Report the p-value from each simple linear regression analysis, and summarize what you can conclude from it.

Number of bathrooms:

Number of bedrooms:

c) Now fit a multiple regression model using all three predictor variables. First report the overall p-value, and summarize what you can conclude from it.

d) Now use this multiple regression model to report the three individual p-values. Are any of the predictor variables significant at the .05 level?

e) Repeat for a two-predictor model using size and number of bathrooms as predictors. Report all p-values, and summarize what you can conclude from them.

Example 2: Speaking and Intelligence

Reconsider the data on predicting Gesell aptitude score based on the age at which a child first speaks.

a) Fit a simple linear regression model. Report the p-value. Is the relationship between these two variables statistically significant?

b) Remove the child who took the longest to speak. Report the appropriate p-value. Now is the relationship between these two variables statistically significant?

c) Now remove the child who took the second longest to speak. Report the appropriate p-value. Now is the relationship between these two variables statistically significant?

Example 3: Wine Prices

Reconsider the wine data.

a) Fit the simple linear regression model for predicting $\ln(\text{price})$ from age. Report the p-value. Is the relationship between these variables statistically significant?

b) Now fit a multiple regression model for predicting $\ln(\text{price})$ from age, summer temperature, harvest rain, and winter rain. Report the overall p-value, and explain what it reveals.

c) Now report the four individual p-values from your multiple regression model. Which variables are statistically significant predictors at the .05 level?