

## Stat 218 - Day 36 Prediction Intervals

Today we continue our study of regression methods by learning how to produce confidence intervals for two quantities:

- the mean value of the response variable at a particular value of the explanatory variable
- a single observation of the response variable at a particular value of the explanatory variable

We will let Minitab do the calculations for us, but we will make some general observations based on them.

### **Example: House prices**

Reconsider the data on price (in thousands of dollars) and size (in square feet) of a sample of 25 houses sold in the Bakersfield, California area for one week in April 2003 (RealEstate.mtw).

a) Reproduce the scatterplot with the regression line superimposed. Also report the equation of that line.

b) Use the regression line to predict the price of a 1800 square foot house.

c) What value would you use to estimate the mean price among all 1800 square foot houses in Bakersfield? Explain.

d) If you were to use an interval to predict the price of a 1800 square foot house and also use an interval to estimate the mean price among all 1800 square foot houses, which would you expect to be wider? Explain. [*Remember*: which varies more- averages or individual values?]

e) Use Minitab to produce both of these interval estimates (Stat> Regression> Regression, select “Options” and enter 1800 for the prediction interval). Record both intervals.

f) Which interval is wider? Is this consistent with your prediction in d)?

g) How do the midpoints of the intervals compare? Explain why this makes sense.

h) Repeat e) for a 1000 square foot house.

i) How do the widths of these intervals compare? How do their midpoints compare? Explain why this all makes sense.

j) Repeat, but change the confidence level to 90%. What effect does this have on the intervals? Explain why this makes sense.

- A prediction interval for an individual value is always wider than the corresponding confidence interval for a mean value.
- Both types of intervals are narrowest at the sample mean of the explanatory variable ( $x$ ) values.

k) Use Minitab to produce confidence and prediction bands for all square foot values within the data (Stat> Regression> Fitted line plot, select “Options,” and check off to display the confidence and prediction intervals.) Are these bands consistent with your earlier findings? Explain.