Unit 4, Module 13: Assessing the Fit of a Line

**Learning Goal:** Use residuals, standard error and *r*2 to assess the fit of a linear model.

In this activity you will practice pulling together what you have learned in Topic 3.3. You will use residuals, se and *r*2 to analyze how well the regression line fits the data.

1. Researchers took body measurements for 136 children at ages 2, 9, and 18. The children were born in 1928-29 in Berkeley CA. Here we investigate the relationship between height at age 2 and height at age 18.

**Simple linear regression results:**

Dependent Variable: Height\_18

Independent Variable: Height\_2

Height\_18 = 45.797 + 1.444 Height\_2

Sample size: 136

R (correlation coefficient) = 0.5486734

R-sq = 0.3010425
Estimate of error standard deviation: 7.4216731

1. In this analysis, what is the explanatory variable? Why does this make sense?
2. Do taller 2-year-olds also tend to be taller 18-year-olds? Support your answer.
3. Is the association between heights at age 2 and age 18 strong or weak? Support your answer.
4. Use the context of the problem to describe what the standard error tells us in this situation.
5. Use the context of the problem to describe what the R-sq value tells us in this situation.
6. In the data set there is a boy who was 92 cm tall at age 2 and 189 cm tall at age 18. What is the residual for this boy? What does the residual tell us?
7. Here we have measurements from 252 men. Body Fat is the percentage of a man’s weight that comes from fat, as opposed to muscle. Weight is measured in pounds. Abdomen is a measurement (cm) around the body at the stomach.
8. Which measurement, weight or abdomen, is a better predictor of the percent of body fat? Support your answer using the scatterplot, se and *r*2.



**Simple linear regression results:**

Dependent Variable: BodyFat

Independent Variable: Weight

BodyFat = -12.052 + 0.1744 Weight

Sample size: 252

R (correlation coefficient) = 0.612414

R-sq = 0.37505091

Estimate of error standard deviation: 6.6290191

**Simple linear regression results:**
Dependent Variable: BodyFat Independent Variable: Abdomen
BodyFat = -39.023 + 0.6282 Abdomen
Sample size: 252
R (correlation coefficient) = 0.81483224
R-sq = 0.66395158
Estimate of error standard deviation: 4.8610253

1. Calculate two predictions for the percentage of body fat of a man who weighs 175 pounds and has an abdomen measurement of 100 cm. Show your work. Which prediction do you think is more accurate? Why?
2. What percentage of the total variation in body fat is explained by weight? By abdomen measurement? What other variables could explain the variation we see in men’s body fat percentages?
3. In general do we want a large or small R-sq value? Why?
4. In general do we want a large or small se value? Why?