STAT 250  
Dr. Kari Lock Morgan

Simple Linear Regression

SECTION 2.6
• Least squares line
• Interpreting coefficients
• Prediction
• Cautions

Want More Stats???
• If you have enjoyed learning how to analyze data, and want to learn more:
  • STAT 460 (Intermediate Applied Statistics)
  • STAT 461 (Analysis of Variance)
  • STAT 462 (Applied Regression Analysis)
  • All applied, only prerequisite is STAT 200 or 250

• If you like math and want to learn more of the mathematical theory behind what we've learned:
  • take STAT 414 (Probability)
  • and then STAT 415 (Mathematical Statistics)
  • Prerequisite: MATH 230 or MATH 231

Question of the Day
Can you estimate the temperature by listening to cricket's chirp?

MODELING
• We will fit a model to predict temperature based on cricket chirp rate

Crickets and Temperature

Linear Model
A linear model predicts a response variable, y, using a linear function of explanatory variables

Simple linear regression predicts on response variable, y, as a linear function of one explanatory variable, x
Regression Line

Goal: Find a straight line that best fits the data in a scatterplot.

Equation of the Line

The estimated regression line is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

where $x$ is the explanatory variable, and $\hat{y}$ is the predicted response variable.

- **Slope:** Increase in predicted $y$ for every unit increase in $x$
- **Intercept:** Predicted $y$ value when $x = 0$

Regression in Minitab

Stat -> Regression -> Fitted Line Plot

Regression Model

$\hat{\text{Temp}} = 37.68 + 0.23\text{Chirps}$

Which is a correct interpretation?

a) The average temperature is $37.68^\circ$

b) For every extra 0.23 chirps per minute, the predicted temperature increases by 1 degree

c) Predicted temperature increases by 0.23 degrees for each extra chirp per minute

d) For every extra 0.23 chirps per minute, the predicted temperature increases by $37.68^\circ$

Units

- It is helpful to think about units when interpreting a regression equation

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$\hat{\text{Temp}} = 37.68 + 0.23\text{Chirps}$

Explanatory and Response

- Unlike correlation, for linear regression it does matter which is the explanatory variable and which is the response

$$\text{Temp} = 37.68 + 0.23\text{Chirps}$$

$$\text{Chirps} = -157.8 + 4.25\text{Temp}$$
Regression Line

Which plot goes with the line \( \hat{y} = x + 10 \)?

Prediction

- The regression equation can be used to predict \( y \) for a given value of \( x \)
  \[ \text{Temp} = 37.68 + 0.23 \text{Chirps} \]
- If you listen and hear crickets chirping about 140 times per minute, your best guess at the outside temperature is 37.68 + 0.23 × 140 = 69.88°

Prediction

\[ 37.68 + 0.23 \times 140 = 69.88° \]

If the crickets are chirping about 180 times per minute, your best guess at the temperature is
(a) 60°
(b) 70°
(c) 80°

Prediction

\[ Temp = 37.68 + 0.23 \text{Chirps} \]
The intercept tells us that the predicted temperature when the crickets are not chirping at all is 37.68°. Do you think this is a good prediction?
(a) Yes
(b) No

Regression Line

- How do we find the best fitting line???
Predicted and Actual Values

- The actual response value, \( y \), is the response value observed for a particular data point.
- The predicted response value, \( \hat{y} \), is the response value that would be predicted for a given \( x \) value, based on a model.
- In linear regression, the predicted values fall on the regression line directly above each \( x \) value.
- The best fitting line is that which makes the predicted values closest to the actual values.

Residual

The residual for each data point is

\[
\text{actual} - \text{predicted} = y - \hat{y}
\]

- The residual is also the vertical distance from each point to the line.

Least Squares Regression

Least squares regression chooses the regression line that minimizes the sum of squared residuals:

\[
\text{minimize } \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

*Note: you will never have to find the equation by hand, this is just letting you know what technology is doing...
Alkalinity and Mercury

Does the relationship look linear?
(a) Yes
(b) No

Life Expectancy and Birth Rate

Which of the following interpretations is correct?
(a) A decrease of 0.89 in the birth rate corresponds to a 1 year increase in predicted life expectancy
(b) Increasing life expectancy by 1 year will cause birth rate to decrease by 0.89
(c) Both
(d) Neither

Regression Caution 1

• Do not use the regression equation or line to predict outside the range of x values available in your data (do not extrapolate!)

• If none of the x values are anywhere near 0, then the intercept is meaningless!

Regression Caution 2

• Computers will calculate a regression line for any two quantitative variables, even if they are not associated or if the association is not linear

• ALWAYS PLOT YOUR DATA!

• The regression line/equation should only be used if the association is approximately linear

Regression Caution 3

• Outliers (especially outliers in both variables) can be very influential on the regression line

• ALWAYS PLOT YOUR DATA!


Regression Caution 4

• Higher values of x may lead to higher (or lower) predicted values of y, but this does NOT mean that changing x will cause y to increase or decrease

• Causation can only be determined if the values of the explanatory variable were determined randomly (which is rarely the case for a continuous explanatory variable)
To Do

• Read Section 2.6
• Do HW 2.6 (due Friday, 12/4)