

Chapters 8 and 9

Linear Regression and Regression Wisdom

At the end of this chapter, you should be able to

- ◆ Calculate a regression line given summary statistics.
- ◆ Interpret the slope of the regression line in context.
- ◆ Know when the interpretation of the intercept of the regression line is reasonable given the context.
- ◆ Find predictions and residuals for points.
- ◆ Interpret a residual plot.
- ◆ Interpret the R^2 value for a regression in context.
- ◆ Describe and apply the limitations of regression.

Correlation and Regression

- ◆ Correlation = linear relationship between two variables.
- ◆ Summarize relationship with line.
- ◆ Called Regression line.
 - Explanatory variable (x)
 - Response variable (y)

Regression line

- ◆ Explains how response variable (y) changes in relation to explanatory variable (x).
- ◆ Use line to predict value of y for given value of x .

Regression line

- ◆ Need mathematical formula (model).
 - Different lines by sight.
- ◆ Predict y from x .
 - The _____ values are called _____
 - ◆ _____
 - The _____ values are called _____
 - ◆ _____

Regression Line

- ◆ Predicted values = Model.
 - Summary of relationship between x and y .
- ◆ Observed values = Model + Error
 - Model = summary of relationship between x and y .
 - Error = amount leftover that the model doesn't explain.

Regression line

- ◆ Look at vertical distance
- ◆ **Error** in regression line.
- ◆ Place line to make these errors as small as possible.

Least squares regression

- ◆ Most commonly used regression line.
- ◆ Puts line where sum of the squared errors as small as possible.
 - Minimizes _____
- ◆ Based on statistics

Regression line equation

where

Regression line equation - slope

- ◆ $b_1 =$ _____.
- Interpretation:

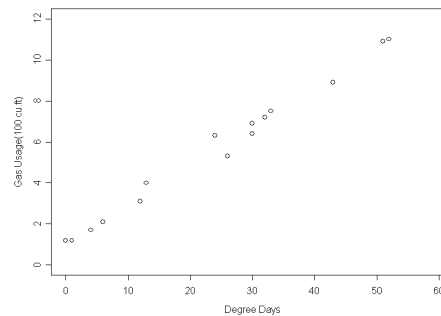
- Very important for interpreting data.

Regression line equation – intercept

- ◆ $b_0 =$ _____
- Interpretation:

- Usually not important for interpreting data.
 - Values of x are usually not close to 0.

Degree Days vs. Gas Usage (per month)



Calculating the regression line.

◆ Degree Days vs. Gas Usage

$$\bar{x} = 22.31, \bar{y} = 5.31, s_x = 17.74, s_y = 3.37, r = 0.9953$$

Calculating the regression line.

◆ Don't forget to write the equation.

Properties of regression line

◆ Regression line always goes through point

◆ r is connected to the value of b_1 .

Interpretations

◆ Interpretation: Slope

◆ Interpretation: Intercept

Predicted Values and Residuals

◆ Predicted Value and Residual for each point

◆ Predicted Values = Model

- Calculate from the Regression Line

◆ Residual = Error

- Calculate from Data and Predicted Value

Predicted Values

◆ From Regression Line = Model

- Ex. Predicted gas consumption when degree days = 43?

- Ex. Predicted gas consumption when degree days = 24?

Residuals

- ◆ From Predicted Value and Observed Value
 - Ex. Residual when degree days = 43?
 - Ex. Residual when degree days = 24?

Observed Values

- ◆ Predicted Value + Residual
- ◆ Mean
- ◆ Standard Deviation

Predicted Values

- ◆ Regression Line (Model)
- ◆ Mean
- ◆ Standard Deviation

Residuals

- ◆ Observed Value – Predicted Value
- ◆ Mean
- ◆ Standard Deviation

Variability

- ◆ In Observed Values
- ◆ In Predicted Values
- ◆ In Residuals

Variability in y explained by Regression Line

- ◆ R^2

Variability in y explained by Regression Line

- ◆ Interpretation of R^2

Properties of R^2

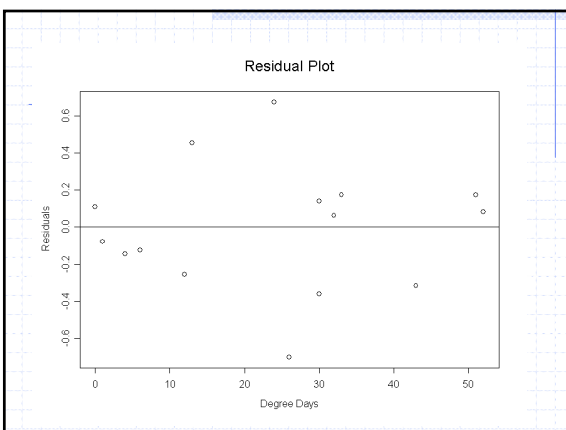
- ◆ Any value from _____
- ◆ Higher values of R^2
- ◆ Lower values of R^2

Residuals

- ◆ Error = relationship between x and y not modeled by regression line.
- ◆ Plot x and Error = Residual Plot

Residual Plot

- ◆ Special Scatterplot
 - Explanatory variable (x) on horizontal axis.
 - Residuals (e) on vertical axis.
 - Horizontal line at residual = 0.
- ◆ Good Residual Plot
 - _____
 - _____



Bad Residual Plots

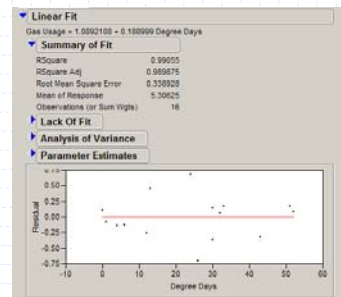
- ◆ Contain a pattern or outliers.
- ◆ Indicates model with regression line is not explaining everything about relationship between x and y .

Example of Other Residual Plots

Example of Other Residual Plots

Example of Other Residual Plots

Regression in JMP

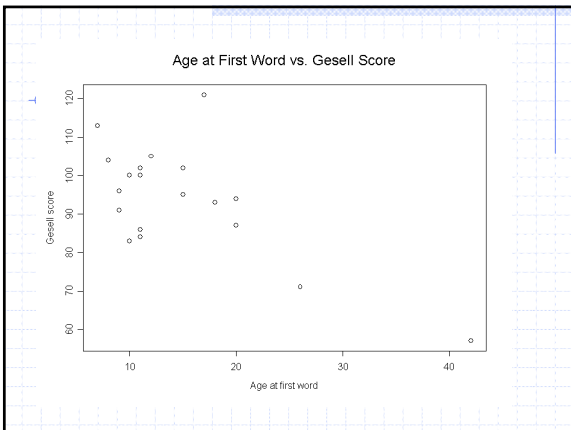


Outliers

- ◆ Observations outside overall pattern.
- ◆ Some examples

Outliers

- ◆ Some outliers are _____.
- Removing the outlier would markedly change regression line.
- Outliers in _____ are often _____.
- Usually have _____.

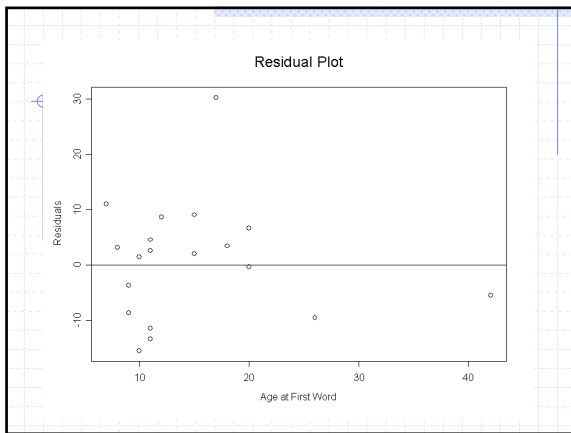


Age at first word vs. Gesell score

- ◆ Scatterplot:
- ◆ Regression:

Interpretations

- ◆ Slope:
- ◆ Intercept:
- ◆ R^2 :



Age at first word vs. Gesell score

- ◆ Residual Plot

Age at first word vs. Gesell score

- ◆ Remove (42,57) from data.

What should you do?

- ◆ Make sure data points recorded correctly.
- ◆ Collect more data.
- ◆ Remove outlier.

Cautions about regression

- ◆ Linear relationship only.
- ◆ Extrapolation

- _____
- _____

Cautions about regression

- ◆ _____
 - Makes relationship appear stronger.
 - Removes variation.
- ◆ _____
 - Important effect on variables, but not included in study.
 - Example:

Cautions about regression

- ◆ _____
 - Strong association between explanatory and response variables does not mean that explanatory variable causes response variable.
 - Ex: High positive correlation between number of TV sets per person and average life expectancy.

Proving Causation

- ◆ Experiment
 - Change values of x and control for lurking variables.
 - Not all problems can be solved by experiments
 - ◆ Smoking causes lung cancer.
 - ◆ Living near power lines causes leukemia.

Proving Causation

- ◆ Proving smoking causes lung cancer.
 - _____
 - _____
 - _____
 - _____
 - _____