

Warm Up



What percent of females thought that they were almost certain to be married in the next ten years?

	Female	Male
Almost no chance	119	103
Some chance, but probably not	150	171
A 50-50 chance	447	512
A good chance	735	710
Almost certain	1174	756

Objective: SWBAT create residual plots to assess how well a regression fits the data

Agenda:

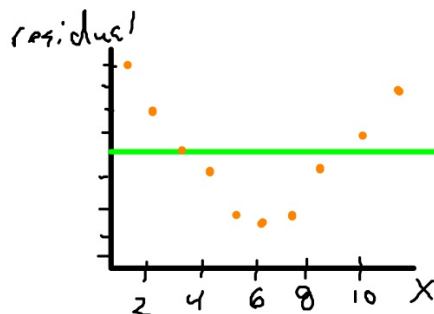
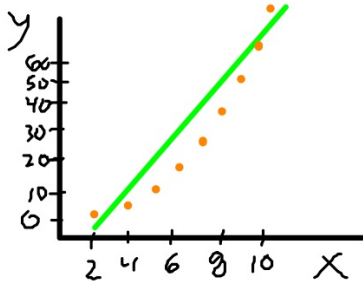
- Warm Up
- Notes
- Practice
- Reflection

HW: pg. 194 #61-70

Notes: Residual Plots

A residual plot graphs the residuals ($y - \hat{y}$) against the explanatory variable.

If a regression line fits the data well and illustrates the trend, the residual plot will have no pattern and the residuals will be small.



Is a line
a good fit
for this data?

Notes: Standard Deviation of the Residuals

The standard deviation of the residuals will tell us how far off on average our predictions are (because the mean is 0 for a least-squares regression line)

$$S = \sqrt{\frac{\sum \text{residuals}^2}{n-2}} = \sqrt{\frac{\sum (y_i - \hat{y})^2}{n-2}}$$

Why $n-2$?

(how many points do we need to make a line?)

Interpreting Computer Output

Minitab

Predictor	Coef	SE Coef	T	P
Constant	-3.751	1.453	-2.58	0.027
Points per game	0.43721	0.05205	8.40	0.000

$s = 1.23531$ $R\text{-Sq} = 87.6\%$ $R\text{-Sq}(\text{adj}) = 86.3\%$

explanatory variable

Standard dev. of the residuals

y-intercept

slope

r^2

JMP

Summary of Fit

RSquare	0.875852	r^2
RSquare Adj	0.863437	
Root Mean Square Error	1.23531	standard dev. of the residuals
Mean of Response	8.083333	
Observations (or Sum Wgts)	12	

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	-3.750604	1.453342	-2.58	0.0274*
Points per game	0.4372144	0.052053	8.40	<.0001*

explanatory variable

y-intercept

slope

Finer Points about Correlation & Regression

1. The difference between explanatory and response variables matters for regressions. If you switch them, you will get a different line.
2. Correlation & regression lines are only meaningful for linear relationships (look up Anscombe's Quartet).
3. Correlation & regression lines are sensitive to outliers.

example: pg. 185 ("Gesell Scores")

Association vs. Causation

Remember, a strong correlation between variables does not imply a cause.

(Having a lot of cars doesn't make you live longer; something else is going on.)

Practice

Chapter 3.2 Quizzes

Reflection

How does a residual plot help us assess whether a regression fits the data?