Compiled by: Mr. Loren L. Spencer

Recipe for Success: Definition Template for Regression

Slope β_1 = b = $\frac{\Delta y}{\Delta x}$ (using y = a+ bx notation)

_____ is the expected amount of change in _____
Value of the slope *Restate the definition of Y*

for a 1 unit increase in _____ .
Restate the definition of x

y- Intercept β_0 = a = constant (using y = a+ bx notation)

We would expect to have_____ _____ if the
Value of y-intercept *Restate the definition of Y*

amount of _____ = zero.
Restate the definition of x

Correlation Coefficient (r) = $\sqrt{R^2}$

- $|r| \geq .75$ There is a strong _____linear relationship between
(+ or –)--*use the sign of the slope*

_____ and _____.
Restate the definition of Y *Restate the definition of x*

- $.40 < |r| < .75$ There is a moderately strong _____linear
(+ or –)--*use the sign of the slope*

relationship between _____and_____.
Restate the definition of Y *Restate the definition of x*

- $|r| < .40$ There is a weak _____linear relationship between
(+ or –)--*use the sign of the slope*

_____ and _____.
Restate the definition of Y *Restate the definition of x*

Coefficient of Determination (R^2)

_____% of the variation in the _____ can be explained by
Restate the definition of Y

changes in the_____
Restate the definition of x

S The **standard deviation of the residuals** is _____ and measures the variance in

S = $\sqrt{\frac{\sum x^2}{(n-2)}}$ _____ for a given amount of _____.
Restate the definition of Y *Restate the definition of x*

Standard Error of the Slope:

The standard error of the slope is_____. Because the slope is estimated from the sample, other samples are likely to have differing slopes. The standard error of the slope quantifies the amount of variation in sample slopes that could be expected from different samples.

Compiled by: Mr. Loren L. Spencer

An Example Computer Print-Out

Before Challenger went of at 31°F, each of the 23 earlier launches experienced from zero to three O-ring failures. There was some speculation that the number of O-ring failures was related to the temperature at lift-off. A computer printout, performed too late, is shown below.

Source	df	SS	MS	F
Regression	1	4.30166	4.30166	9.66
Residual	21	9.35052	0.445263	
Variable	Coef	s.e. Coeff	t	P
Constant	4.79365	1.409	3.4	0.0027
Temperature	-0.0626587	0.02016	-3.11	0.0052
s = .06673		R-sq = 31.5%	R-sq(adj) = 28.2%	

Explanatory Variable (x): Temperature

Response Variable (y): The number of o-ring failures

Least Squares equation: $\widehat{failures}$= 4.79365 + (-0.0626587)(temperature)

Slope: $\frac{-0.062587(failures)}{1\ Temperature}$
We would expect a 0.062587 decrease in o-ring failures for every 1 degree increase in temperature.

y-intercept: 4.79365
We would expect to have 4.79365 o-ring failures if the temperature was zero degrees

Correlation Coefficient: r = - $\sqrt{.315}$ = -.5612 *(r is negative because the slope is negative)*
There is a moderately strong **negative** linear relationship between the amount of o-ring failures and temperature.

Coefficient of Determination: R-sq = 31.5% or R^2 = 31.5%
31.5% of the variation in the number of O-ring failures can be explained by changes in temperature.

Standard Deviation of the Residuals: .06673
The standard deviation of the residuals is .06673 and measures the amount of variation in o-ring failures that we can expect for a given temperature.

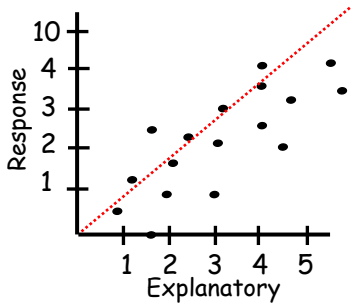
Note: Residuals = the number of actual o-ring failures – the predicted number of failures.
Residuals are the vertical distance an observed value is from the predicted.
Remember: A residual plot needs to be random and with no pattern for a given equation to be appropriate

Standard Error of the Slope: .02016
The standard error of the slope is 02016. Because the slope is estimated from the sample, other samples are likely to have differing slopes. The standard error of the slope quantifies the amount of variation in sample slopes that could be expected from different samples.

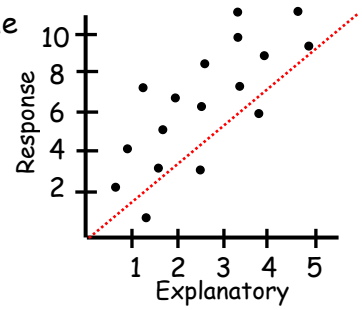
Analyzing a Scatterplot:

If you need to analyze a scatterplot draw a line at 45 degrees through the origin.

If the majority of the points are below the line, the explanatory variable tends to be **larger than** the Response Variable.



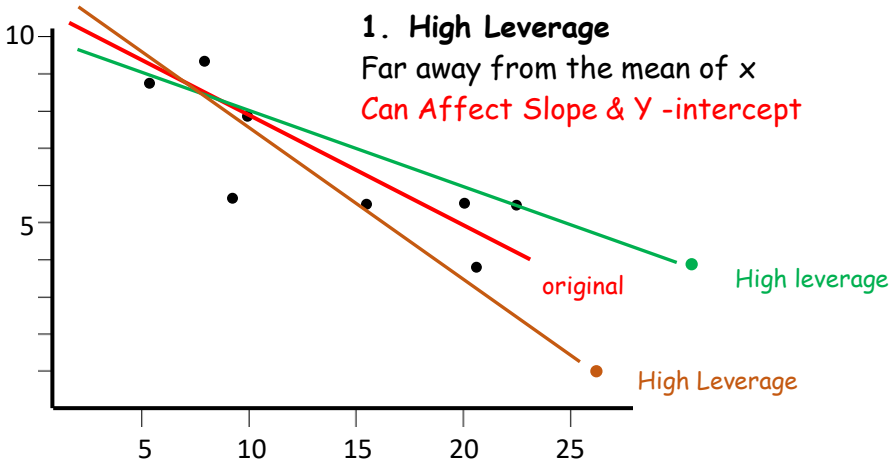
If the majority of the points are above the line, the explanatory variable tends to be **less than** the Response Variable.



Know what a correlation of 1 and -1 and 0 look like

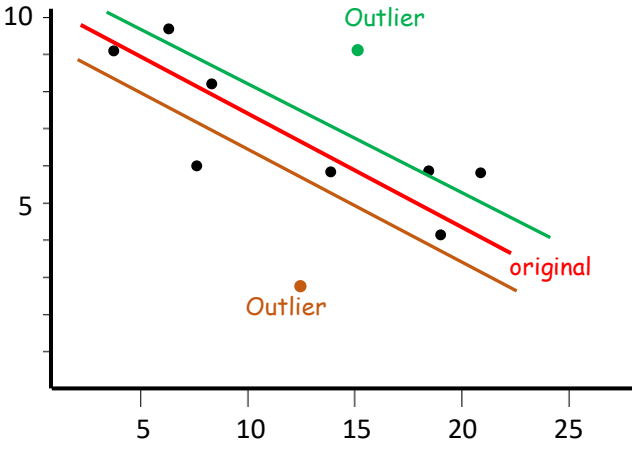
A slope of zero is not useful for prediction and horizontal lines have slopes of zero.

Influential Points: (3 Types)

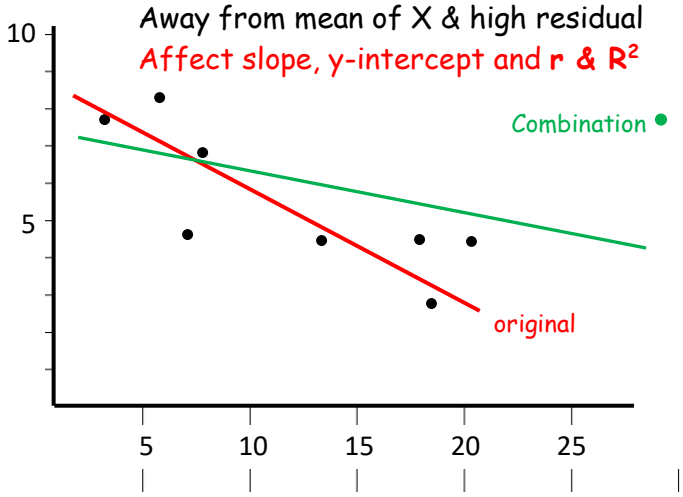


1. High Leverage
Far away from the mean of x
Can Affect Slope & Y -intercept

2. Outliers
Close to the mean of x
but have a high residual and
substantially reduce the **r & R^2**



3. Combination
Away from mean of X & high residual
Affect slope, y-intercept and **r & R^2**



Residuals

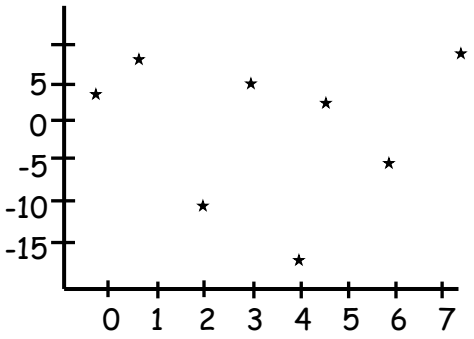
(residual = original – expected)

Residual = actual – predicted

Residual = y - $\hat{y}$ (remember $\hat{y}$ is predicted from the equation and lies on the regression line)

Be able to calculate the residuals using graphs and equation

The number of students taking AP Statistics at a high school during the years 2000-2007 is fitted with a least square regression line. The graph of the residuals & some computer output is as follows.

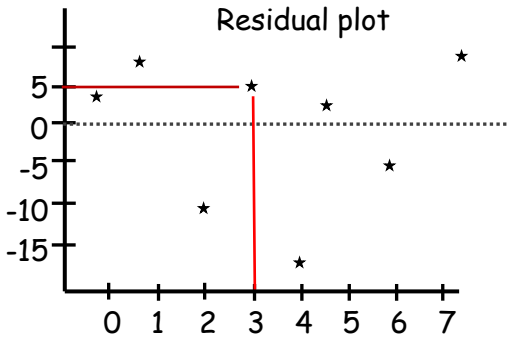


Dependent variable is: Students

Variable	Coeff	s.e	t	p
Constant	11	6.299	1.75	0.1313
Years	13.9826	1.506	9.25	0.0001

S = 9.758 R-sq =93.4% R-sq(adj) = 92.4%

How many students took AP Statistics in the year 2003?



The residual for 2003 is 5

$\widehat{\text{students}}$ = 11 + 13.986(years)

$\hat{y}$ = 11 + 13.986(x)
→ $\hat{y}$ = 11 + 13.986(3)

→ $\hat{y}$ = 52.958

Expected/Predicted ≈ 53

Residuals = Observed - Expected

→ 5 = Observed – 53

→ 58 = **Observed/Actual**

Remember: The residual must be random, centered about zero and without pattern for the regression line to be appropriate.

Compiled by: Mr. Loren L. Spencer

Recipe for Success: The Regression, Scatterplot & Residual Graphs

1. Turn on STAT Diagnostics

- Press **MODE**
- ↓ **STATDIAGNOSTICS:**
- → Highlight **ON**
- Press **ENTER**
- Press **2nd Mode/Quit**

2. Input the Data

- Enter "**x**" values into L_1
- Enter "**y**" values into L_2

3. Calculate the Regression Statistics

- Regression Equation **y = a + bx**
- Slope: $B_1 = b$
- Y-intercept: $B_0 = a$
- Correlation Coefficient: **r**
- Coefficient of Determination: r^2

- Press **STAT** → Highlight ***CALC***
- ↓ **8:LinReg (a + bx)**
- ↓ XList: Press **2nd** L_1 Enter
- ↓ YList: Press **2nd** L_2 Enter
- ↓ Store RegEQ: Press **2nd ALPHA TRACE** ENTER
- Press **2nd Mode/Quit**

4. Graphing:
Scatter Plot vs. Regression Equation

- Press **2nd STAT PLOT**
- Highlight **1: Plot 1** Press ENTER
- Highlight **On** Press ENTER
- ↓ Highlight First Graph Press ENTER
- ↓ XList: Press **2nd** L_1 Enter
- ↓ YList: Press **2nd** L_2 Enter
- Press **ZOOM 9**

5. Calculating Predicted Values
Caution: Do not make predictions outside the range of x-values.

- Press **2nd TABLESET**
- Input x-value
- Press **2nd TABLE**
- **OR** Input an x value into the equation and solve for y

6. Residuals:
The **vertical distance** from a given data point to the line of best fit

- A **positive** residual means the **actual is greater** than the predicted-**above the regression line**
- A **negative** residual means the **actual is less** than the predicted-**below the regression line**

7. Calculating Residuals
(actual – predicted)

- Press **STAT** → Highlight **EDIT** & Press **ENTER**
- ↑ Highlight L_3
- Press **2nd STAT/LIST**
- ↓ Highlight **7 RESID** Press ENTER
- Press ENTER again
- Press **ZOOM 9**

8. Graphing Residuals
(actual – predicted)

- Press **2nd STAT PLOT**
- Highlight **1: Plot 1** Press ENTER
- Highlight **On** Press ENTER
- ↓ Highlight First Graph Press ENTER
- ↓ XList: Press **2nd** L_1 Enter
- ↓ YList: Press **2nd** L_3 Enter
- Press **ZOOM 9**