Section 9-3 Regression

- **Regression Equation**
  Given a collection of paired data, the regression equation

$$\hat{y} = a + b_1 x$$

algebraically describes the relationship between the two variables

- **Regression Line**
  (line of best fit or least-squares line)

Apr 27-8:37 AM

The Regression Equation

x is the independent variable (predictor variable)

$\hat{y}$ is the dependent variable (response variable)

$\hat{y} = a + b_1 x$ $\quad a = y$ - intercept

$y = mx + b$ $\quad b_1 =$ slope

- **Round to three significant digits**

Apr 27-8:41 AM

The regression line can be used to make prediction.

When making predictions, determine if the data set has a correlation by checking the r-value and critical value.

1. **If there is not a significant linear correlation, the best predicted y-value is $\bar{y}$.**

$\bar{y}$ is the mean for the y coordinates. Calculate a 1 var stat for L_2 (y column) to find it.

2. **If there is a significant linear correlation, the best predicted y-value is found by substituting the x-value into the regression equation.**

Apr 27-8:44 AM

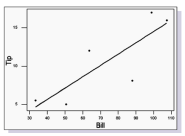
Guidelines for Using The Regression Equation

1. **If there is no significant linear correlation, don't use the regression equation to make predictions.**
2. **When using the regression equation for predictions, stay within the scope of the available sample data.**
3. **A regression equation based on old data is not necessarily valid now.**
4. **Don't make predictions about a population that is different from the population from which the sample data was drawn.**

Apr 27-8:46 AM

Ex1) Write a linear equation to represent the correlation between the bill and the tip. Then predict what you should leave for tip if you bill was $45.75.

L1 bill	L2 tip
33.46	5
50.68	5
87.92	8.08
98.84	17
63.6	12
107.34	16



```
LinRegTTest
 y=a+bx
 β≠0 and ρ≠0
↑p=.0417567494
 df=4
 a=-.3472791722
↓b=.1486141477
```

y = -.347 +.149x

r value = .838
critical value = .811
Yes, it is linear.

y = -.247 + (.149)(45.75)
tip = 6.77

May 3-8:20 AM

Ex2) **What is the best predicted size of a household that discard 0.50 lb of plastic?**

Data from the Garbage Project

x Plastic (lb)	0.27	1.41	2.19	2.83	2.19	1.81	0.85	3.05
y Household	2	3	3	6	4	2	1	5

Using a calculator:

$a = 0.549$

$b_1 = 1.48$

$y = .549 + 1.48x$

```
LinRegTTest
 y=a+bx
 β≠0 and ρ≠0
↑p=.0086721828
 df=6
 a=.5492703618
↓b=1.479851857
```

r value = .842
critical value = ±.701
Yes, it is linear.

y = 0.549 + 1.48 (0.50)
y = 1.3

A household that discards 0.50 lb of plastic has approximately one person.

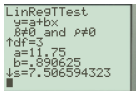
Apr 27-8:46 AM

Ex3) Find the regression line and determine the y value if x = 4 for the following data set.

x	1	2	3	5	5
y	19	12	4	20	18

regression equation:

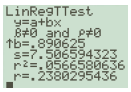
y = 11.75 + .891x



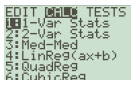
r value = .238

critical value = .878

There is NOT a linear correlation.



Use the mean value for y to be the best prediction.



14.6

May 3-7:46 PM

- **Marginal Change**
  the amount a variable changes when the other variable changes by exactly one unit
- **Outlier**
  a point lying far away from the other data points
- **Influential Points**
  points which strongly affect the graph of the regression line

Apr 27-8:47 AM