

AFM Unit 9 Regression Day 1 notes

A mathematical model is an equation that best describes a particular set of paired data.

These mathematical models are referred to as regression models and are used to estimate one variable based upon another variable.

Linear Regression: STAT CALC #4 LinReg (ax+b)



Correlation Coefficient:

the quantity r, measures the strength and direction of a linear relationship between 2 variables.

-1 ≤ r ≤ 1 If x and y have a strong positive linear correlation, r is close to +1

If a perfect positive fit, then r = 1. (the slope of the line will be 1)

Positive values: as values for x increases, values for y increase.

Negative correlation: r will be close to -1. An r value of -1 indicates a perfect negative fit. (slope = -1)

Negative values: as values for x increase, values for y decrease.

No correlation: If there is no linear correlation or a very weak linear correlation, r is close to 0.

**To turn the Diagnostics on:**

1. Press 2nd CATALOG (above the numeral zero) to display the Catalog in alpha mode (note the A in the upper right hand corner).

Interpolation: Predicting values with an equation of best fit within the range of data

Extrapolation: Predicting values with an equation of best fit outside the range of data.

Example 1. Is there a relationship between Math SAT scores and the number of hours spent studying for the test?

A study was conducted involving 20 students as they prepared for and took the Math section of the SAT Exam.

y = 25.33x + 353.16 a.) Determine a linear regression model equation to represent the data.

yes b.) Graph in the calculator and decide whether the equation is a "good fit".

733 c.) If a student studied for 15 hours, what is the expected Math SAT score?

interpol. d.) Is part c an example of interpolation or extrapolation? given x find y Trace

14.5 e.) If a student obtained a Math SAT score of 720, how many hours did the student most likely spend studying?

2886 f.) If a student spent 100 hours studying, what would be the expected Math SAT score? given y find 2nd Trace inters.

extrap g.) Is part "f" an example of interpolation or extrapolation?

Hours Spent Studying	Math SAT Score
4	390
9	580
10	650
14	730
4	410
7	530
12	600
22	790
1	350
3	400
8	590
11	640
5	450
6	520
10	690
11	690
16	770
13	700
13	730
10	640

1. Write the equation of the line in standard form  $Ax + By = C$  which passes through (1, 1) and is parallel to the line  $x - 3y = 6$ .

$$y = mx + b$$

$$y = \frac{1}{3}x + b$$

$$1 = \frac{1}{3}(1) + b \quad b = \frac{2}{3}$$

$$y = \frac{1}{3}x + \frac{2}{3}$$

$$3y = x + 2$$

$$-2 = x - 3y$$

$$x - 3y = 6$$

$$\frac{-3y}{-3} = \frac{-x + 6}{-3}$$

$$y = \frac{1}{3}x - 2$$

2. Find the fifth term of the expansion of  $(x - 3y)^{10}$

$${}_{10}C_4 x^6 (-3y)^4$$

$$17,010 x^6 y^4$$

- ① A must be pos.  
 ② can't have fract.  
 ③ x, y on same side

3. Boston has a 32% chance of beating Cleveland this year. What is the probability that Boston will beat Cleveland in the best of five series?

$${}_5C_3 (.32)^3 (.68)^2$$

$$15.2\%$$

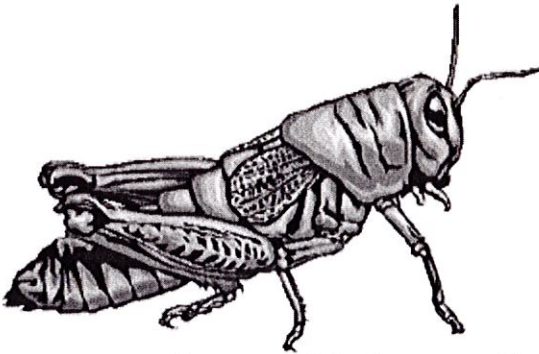
4. A physics exam consists of 12 multiple choice questions and 6 open ended problems in which all of the work must be shown. If the students must answer 8 of the multiple choice and 3 of the open ended problems, in how many ways can the questions and problems be chosen?

$${}_{12}C_8 \cdot {}_6C_3 = 9,900$$



Example 2.

y x



Pierce (1948) *mechanically* measured the frequency (the number of wing vibrations per second) of chirps (or pulses of sound) made by a striped ground cricket, at various ground temperatures. Since crickets are ectotherms (cold-blooded), the rate of their physiological processes and their overall metabolism are influenced by temperature. Consequently, there is reason to believe that temperature would have a profound effect on aspects of their behavior, such as chirp frequency. In general, it was found that crickets did not sing at temperatures colder than 60° F. or warmer than 100° F.

Chirps/Second	Temperature (° F)
20.0	88.6
16.0	71.6
19.8	93.3
18.4	84.3
17.1	80.6
15.5	75.2
14.7	69.7
15.7	71.6
15.4	69.4
16.3	83.3
15.0	79.6
17.2	82.6
16.0	80.6
17.0	83.5
14.4	76.3

$y = .203x + .46$  a) Determine a linear regression model equation to represent this data.

.83 average. b) Graph and determine whether the new equation is a "good fit".

19.7 c) If the ground temperature reached 95° F, then at what rate would you expect the crickets to be chirping? *give x*

extrap d) Is part c an example of interpolation or extrapolation? *Trace*

86.4° e) With a listening device, you discovered that on a particular morning the crickets were chirping at a rate of 18 chirps per second. What was the approximate ground temperature that morning?

interp. f) Is part "e" an example of interpolation or extrapolation? *y = 18 2nd trace intersect*

6.95 (slows down) g) If the ground temperature should drop to freezing (32° F), what happens to the cricket's chirping rate?

Round all answers to nearest ten thousandths.

1. The table shows statistics for nine NHL career scoring leaders.

Games Played, x	1487	1767	1269	1348	1282	1432	1756	1425	1514
Goals Scored, y	894	801	741	731	717	708	694	692	692

$y = .053x + 663.42$

a. Find an equation of the least-squares line.

$.14$

b. Find the correlation coefficient.

$742$

c. Mike Modano played 1499 games. Predict the number of goals he might have scored.

2. Brian wanted to determine the relationship that might exist between speed and miles per gallon of an automobile. Let x be the average speed of a car on the highway measured in miles per hour and let Y represent the miles per gallon of the automobile. The following data are collected.

X	50	55	55	60	60	62	65	65
Y	28	26	25	22	20	20	17	15

$y = -.83x + 70.39$

a. Use a graphing calculator to find the line of best fit.

For every 1mph you lose .83 miles per gallon

b. Interpret what the slope means.

c. Predict the miles per gallon of a car traveling 61 miles per hour.

$56.13$

d. Predict the speed of a car that gets 24 miles per gallon.

3. A doctor wished to determine whether a relationship exists between the height of a female and weight. She obtained the heights and weights of 10 females aged 18-24. Let height be the independent variable, X, measured in inches, and weight be the dependent variable, Y, measured in pounds.

X	60	61	62	62	64	65	65	67	68	68
Y	105	110	115	120	120	125	130	135	135	145

$y = 4.13x - 140.95$

a. Use a graphing calculator to find the line of best fit.

go up 4 lbs for every inch

b. Interpret what the slope means.

$131.43$

c. Predict the weight of a female aged 18 to 24 whose height is 66 inches.

$63.72$

d. Predict the height of a female whose weight is 122 pounds.

For each pair of variables, tell whether you think the correlation is positive, negative, or approximately zero. Briefly give your reasons.

- Average daily temperature in Waterville during January and average daily heating cost in Waterville during January. *negative*
- A person's income and the value of his or her house. *positive*
- The amount of annual taxes a person pays and his or her height. *zero*
- The height of a person and weight of a person *positive*
- The height of a person and the average height of his or her parents *weak (maybe pos)*
- The value of an automobile and its age *negative*



$$y = ab^x$$

Fit an exponential curve to the data.

X	1	3	4	5
y	3.00	6.75	10.13	15.19

STAT CALC #4

$$r = .956$$

STAT CALC #10

$$r = .9995$$

EXPRG

The average price of an adult ticket to the movies in the United States has increased,

a. Use the movie price data to plot y versus year x. Represent the years 1948, 1958, ... as x = 0, 10,

b. linear or exponential? *better fit.*

$y = 0.37(1.06)^x$  c. write the equation.

~~8.81~~ d. price of a movie ticket in 2000? *Trace 52*

~~16.22~~ e. price of a movie ticket in 2010? *Trace 62*

X = year	Y = price
1948	\$0.36
1958	\$0.68
1967	\$1.22
1978	\$2.34
1988	\$4.11

2. A study of nutrition in developing counties collected data from the Egyptian village of Nahya. Here are the mean weights for 170 infants in Nahya who were weighed each month during their first year of life.

Age (in months)	1	2	3	4	5	6	7	8	9	10	11	12
Mean Wt. (kg)	4.3	5.1	5.7	6.3	6.8	7.1	7.2	7.2	7.2	7.2	7.5	7.8

a. Find the correlation coefficient for each type of equation. (linear, exponential) round to 4 places

\_\_\_\_\_

\_\_\_\_\_ b. Which equation best fits the data?

3. Brian wanted to determine the relationship that might exist between speed and miles per gallon of an automobile. Let x be the average speed of a car on the highway measured in miles per hour and let Y represent the miles per gallon of the automobile. The following data are collected.

X	50	55	55	60	60	62	65	65
Y	28	26	25	22	20	20	17	15

a) Plot the data. Find the regression equation and correlation coefficient (r).

b) Interpret what the slope means.

c) Predict the miles per gallon of a car traveling 61 miles per hour.

d) Predict the speed of a car that gets 24 miles per gallon.

4. A doctor wished to determine whether a relationship exists between the height of a female and weight. She obtained the heights and weights of 10 females aged 18-24. Let height be the independent variable, measured in inches, and weight be the dependent variable, Y, measured in pounds.

X	60	61	62	62	64	65	65	67	68	68
y	105	110	115	120	120	125	130	135	135	145

a) Plot the data. Find the regression equation and correlation coefficient (r).

b) Interpret what the slope means.

c) Predict the weight of a female aged 18 to 24 whose height is 66 inches.

d) Predict the height of a female whose weight is 122 pounds.

5. Suppose you are studying frogs that live in a nearby wetland area. The data below was collected by a local conservation organization. They indicate the number of frogs estimated to be living in the wetland area over a 5-year period.

a. Write the exponential equation to model this data.

b. What will the population be in 8 years?

c. What will be the population in 2020?

Year	Estimated Population
2006	120
2007	101
2008	86
2009	72
2010	60