AP Stats

Name ANSWER KEY

Review: Regression

1. Some pizza facts.

Pizza	Fat (g)	Calories
Pizza Hut's Hand Tossed	9	305
Domino's Deep Dish	16	382
Pizza Hut's Pan	14	338
Domino's Hand Tossed	9	327
Little Caesar's Pan!Pan!	10	309
Little Caesar's Pizza! Pizza!	11	313
Pizza Hut's Stuffed Crust	13	349

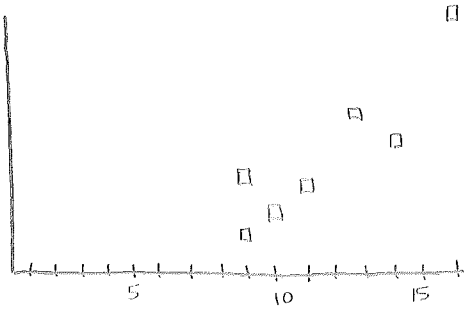
A. Identify the explanatory and response variables.

FAT CALORIES

B. Construct a scatterplot. Describe the relationship.

FORM: STRAIGHT

STRENGTH: VERY STRONG



C. Compute the correlation coefficient.

$r = .8846$

D. Based on the value of the correlation, is it reasonable to conclude that more fat in the pizza causes the pizza to have more calories? Explain.

No, this is an observation, does NOT prove cause/effect

E. Find the equation of the least-squares line.

$\hat{y} = 8.97x + 226.74$ $\quad \widehat{CAL} = 8.97(fat) + 226.74$

F. When the fat content changes by 1 gram, by how much does the calorie content tend to change.

8.97 calories (the slope)

G. Explain the y-intercept of the least-squares line.

A pizza with zero grams of fat contains 226.74 calories

H. What would you predict the number of calories are in a pizza with a fat content of 11 grams?

$\hat{y} = 8.97(11) + 226.74$ $\quad \hat{y} = 325.45$

I. Can I use this to predict the number of calories in a pizza that has a fat content of 20 grams?

Not a good idea, that would be an example of extrapolation. 20 grams is out of the given domain (9 to 16).

J. What calorie amount would you predict for a pizza with a fat content of 11 grams? What is the residual corresponding to the observation (11, 313)?

$\hat{y}_{11} = 325.45$ $\quad$ Residual $= y - \hat{y}$ $\quad R = 313 - 325.45 =$ **Residual: -12.45**

K. Calculate the SSE. What does this represent?

$SSE = \sum (y_i - \hat{y})^2 = 971.68$

$r^2 = \frac{(SST - SSE)}{SST}$

L. Calculate the SST. What does this represent?

$SST = \sum (y_i - \bar{y})^2 = 4468.85$

M. **Calculate** and interpret the value of r^2.

$r^2 = .7826$ 78.26% of the variation in CALORIES can be explained by the linear relationship with FAT.

N. Find and interpret the value of s_e.

$s_e = s = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{971.68}{5}} = 13.94$

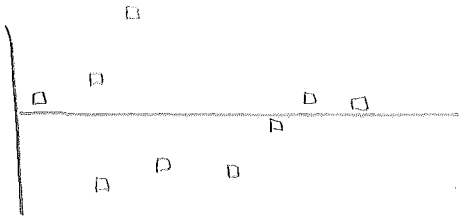
The average amount that the observed CALORIES differ from the predicted calories is approx 13.94 calories.

$\hat{y} = -.78x + 106.47$ $r = -.526$ $r^2 = .277$

2. The table below describes the relationship between latitude and average July temperature in the twelve largest U.S. cities.

City	Latitude (x)	July Temp (y)
New York	40	77
Los Angeles	34	74
Chicago	42	75
Houston	29	84
Philadelphia	40	77
Phoenix	33	94
San Diego	32	71
San Antonio	29	85
Dallas	32	86
San Jose	37	70
Detroit	42	74
Indianapolis	39	75

a. Draw the residual plot.



b. Is a linear plot appropriate? Explain.

Not bad. It's appropriate. It is just not a strong correlation.

c. Phoenix has a very large positive residual. How would the slope of the regression line change if it were removed from the data set? $\hat{y} = -.62x + 99.45$ $r = -.56$ $r^2 = .314$

Not as steep, so slope is a smaller negative #

3. Alana's favorite exercise machine is a stair climber. On the "random" setting, it changes speeds at regular intervals, so the total number of simulated "floors" she climbs varies from session to session. She also exercises for different lengths of time each session. She decides to explore the relationship between the number of minutes she works out on the stair climber and the number of floors it tells her that she's climbed. She records minutes of climbing time and number of floors climbed for six exercise sessions. The computer output from a linear regression analysis of the data is shown below.

Predictor	Coef	SE Coef	T	P
Constant	-3.822	5.458	-0.70	0.522
Minutes	5.2150	0.2779	18.76	0.000

S = 2.34720 R-Sq = 98.9% R-Sq(adj) = 98.6%

a. What is the equation of the LSRL?

$\hat{y} = -3.822 + 5.215x$

$\widehat{\#\text{FLOORS}} = -3.822 + 5.215(\#\text{ OF MIN WORK out})$

b. Find and interpret the correlation coefficient.

$r = \sqrt{.989} = .994$ Very strong, positive, linear correlation.

c. Find and interpret the coefficient of determination.

98.9% OF the VARIATION IN # OF FLOORS climbed CAN BE explained by the VARIATIONS IN # OF MIN WORKED out.

d. Find and interpret the standard deviation of the error.

$S_e = S = 2.34720$

STANDARD DEV. OF RESIDUAL

The average amount that the observed FLOORS differ from the predicted # OF FLOORS is Approx 2.34720 FLOORS.

5. Do aircraft with a higher typical speed also have a greater average flight length? Find the equation that best describes this relationship and use it to predict the Flight length for a speed of 445 mph.

Speed (mph)	Flight Length (mi)
518	2882
539	5063
529	3321
498	1363
513	2451
498	1493
504	1963
516	2379
467	1126
524	3253
523	2995
495	2331
464	1167
486	2135

Speed ($\hat{y}$) (mph)	Flight Length (mi)
441	782
440	742
458	1101
414	702
432	798
416	602
374	345
388	442
412	570
387	487
389	468
384	500
380	413

$\hat{y} = 15.65x - 5573.34$

Flight length $= 15.65(\text{speed}) - 5573.34$

$\hat{y} = 15.65(445) - 5573.3$

$\hat{y} = 1390.91$ mi

4. The number of hours a student studied for the exam (x) was correlated with the students' score on the exam(y) to obtain the least squares regression line: $\textit{predicted exam} = 5.19(\textit{hours}) + 67.19$.

a. Find and interpret the slope.

Slope: 5.19 As the number of hours studying increases by an hour, the predicted exam score increases by 5.19 units.

b. Find and interpret the y-intercept.

A student who does not study for the exam (zero hours), will receive a score of 67.19%.

c. Find the residual amount for a student who studied 3 hours and made an 83 on the exam.

$\hat{y} = 5.19(3) + 67.19$

$\hat{y} = 82.76$

$\text{Res} = y - \hat{y}$

$83 - 82.76 = .24$ underpred

5. Find the correlation coefficient and the least squares regression line for each set of data shown below. Then find the residual plot. Discuss what you found.

a.

x	10	8	13	9	11	14	6	4	12	7	5
y	8.04	6.95	7.58	8.81	8.83	9.96	7.24	4.26	10.84	4.82	5.68

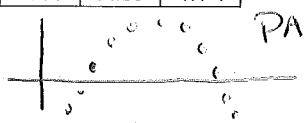
Line is Good!

$r = .82$ $\hat{y} = .51x + 2.96$ Residual Plot  Scattered

b.

x	10	8	13	9	11	14	6	4	12	7	5
y	9.14	8.14	8.74	8.77	9.26	8.1	6.13	3.1	9.13	7.26	4.74

Line is NOT Best Fit

$r = .82$ $\hat{y} = .5x + 3$ Residual Plot  PATTERN! NO BUENO!!

c.

x	8	8	8	8	8	8	8	8	8	8	19
y	6.58	5.76	7.71	8.84	7.04	5.25	5.56	7.91	6.89	8.47	12.5

LINE IS NOT Best Fit

Slope is undefined! CAN NOT DO THIS!