AP STAT Worksheet: LSRL, Correlation, & Predictions Name Key

Ch 3.2: Least Squares Regression Lines

1. Karl keeps his savings under his mattress. He began with $500 from his mother and add $100 each year. His total savings y after x years are given by the equation savings = 500 + 100(years)

a) After 20 years, how much will Karl have under his mattress? Karl will have $2,500 under his mattress after 20 years.

b) If Karl had added $200 instead of $100 each year to his initial $500, what is the equation that describes his savings after x years.

The equation savings = 500 + 200(years) best models this relationship.

2. The following computer output is the result of plotting hours studied for an exam against the score obtained on the exam.

The regression equation is

Predictor	Coef	St Dev	t ratio	P
Constant	59.026	2.863	20.62	.000
Hours	6.767	1.092	6.20	.000

s = 6.135 R-sq = 74.7% R-sq(adj) = 72.8%

a) Write an equation (in context) for exam score after hours studying.

pred exam score = 59.026 + 6.767 (hours studied) The equation best models this relationship.

b) Interpret the slope in context of the problem.

On average, for each addition hour studied, the predicted exam score will increase 6.767 points.

c) Interpret the y-intercept in context of the problem.

The predicted exam score will be 59.026 if a student does not study at all, hours = 0, for the exam.

3. The following calculations are from a data set plotting school grade point average (GPA), **y**, against IQ test score , **x**, for 78 seventh-grade students.

$\bar{x}$ = 108.9 s_x = 13.17 $\bar{y}$ = 7.447 s_y = 2.10 r = 0.6337

a) Find the equation of the least squares line for predicting GPA from IQ.

$b = r\left(\frac{s_y}{s_x}\right) = .6337\left(\frac{2.10}{13.17}\right) = .101$

$a = \bar{y} - b\bar{x} = 7.447 - (.101)(108.9) = -3.5519$

pred GPA = -3.5519 + .101(IQ test score)

b) What percent of the observed variation in these students' GPAs can be explained by the linear relationship between GPA and IQ? The coefficient of determination is .4016. This means approximately 40.2% of the variation in student GPA is accounted for by the LSRL of GPA on IQ test scores.

c) One student has an IQ of 103 but a very low GPA of 0.53. What is the predicted GPA for a student with IQ = 103? What is the residual for this particular student and what does the residual say about this student?

pred GPA = -3.5519 + .101(103)

pred GPA = 6.8511

Resid = obs - pred = 0.53 - 6.8511 = -6.3211

A residual value of -6.32 implies this student is an underachiever.

4. Keeping water supplies clean requires regular measurement of levels of pollutants. The measurements are indirect - a typical analysis involves forming a dye by a chemical reaction with the dissolved pollutant, then passing light through the solution and measuring its "absorbence." To calibrate such measurements, the laboratory measures known standard solutions and uses regression to relate absorbence to pollute concentration. This is usually done every day. Here is one series of data on the absorbence for different levels of nitrates. Nitrates are measured in milligrams per liter of water.

Nitrates	50	50	100	200	400	800	1200	1600	2000	2000
Absorbence	7.0	7.5	12.8	24.0	47.0	93.0	138.0	183.0	230.0	226.0

a) Chemical theory says that these data should lie on a straight line. If the correlation is not at least 0.997, something went wrong and the calibration procedure is repeated. Find the correlation and decide if the calibration must be done again.

The correlation coefficient was 0.999, therefore the calibration does not need to be repeated.

b) What is the equation of the LSRL for predicting absorbence from concentration? If the lab analyzed a specimen with 500 mg of nitrates per liter, what do you expect the absorbence to be?

The LSRL for this data is pred Absorbence = 1.6571 + 0.1133(concentration). A specimen with 500 mg of nitrates per liter is expected to have an absorbence level of 58.31.

c) What type of prediction is this? pred Absorbence = 1.6571 + 0.1133(500) = 58.3071

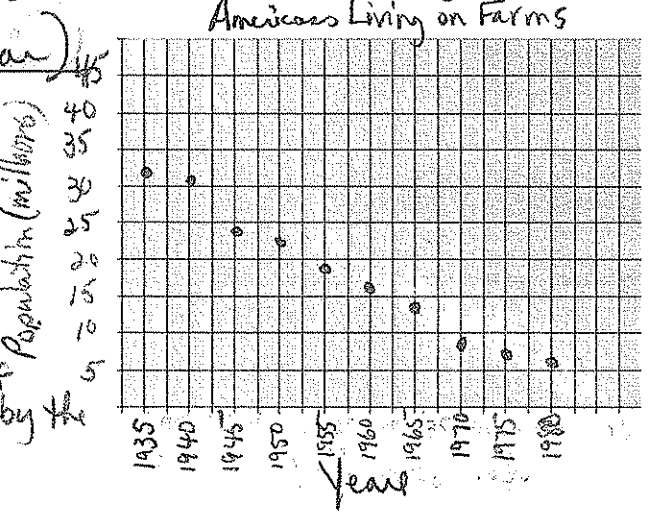
This type of prediction is called interpolation.

5. The number of people living on American farms has steadily declined during this century. Here are the data on the farm population (millions of persons) from 1935 to 1980.

Year	1935	1940	1945	1950	1955	1960	1965	1970	1975	1980
Population	32.1	30.5	24.4	23.0	19.1	15.6	12.4	9.7	8.9	7.2

a) Make a scatterplot. State and graph the least-squares regression line (in context) of farming on year.

LSRL: pred pop on farms = 1166.93 − 0.5868(year)



b) According to the regression line, how much did the farm population decline, on the average, each year during this period? Farm population declined, on average, 0.5868 million people per year during this period.

c) What is the *proportional reduction in error* (also called the coefficient of determination)? Interpret this value in context of the problem.

The coefficient of determination is 0.977. Approx 97% of the variation in pred farm population is accounted for by the LSRL of farm pop on year after 1930.

d) Predict the number of people living on farms in 1990.

This type of prediction is called extrapolation

Is this result reasonable? Explain.

The predicted number of people living on farms in 1990 is −.802 million. This type of prediction is called extrapolation. This result is not reasonable because it is a negative value for population

e) What conclusions can you make concerning making predictions obtained from values of x which stray far from your data?

Use caution when making predictions outside of the data set. Model was created using given data. Predictions become less accurate when made outside given domain.

d) pred pop = 1166.93 − 0.5868(1990) = −.802