AP Statistics
Mr Murphy
4.2 Linear Regression: Fitting a Line to Bivariate Data

Assignment 4-2
Pg. 217 #5.17, 5.19, 5.23, 5.27, 5.29, 5.30

Goals:
1. Write the equation for a regression line
2. Interpret the slope and y intercept of a regression line in context
3. Detect when extrapolation occurs

- *y* is called the **dependent** or **response variable**, and *x* is referred to as the **independent**, **predictor**, or **explanatory variable**.

- The relationship $y = a + bx$ is the equation of a line. The value of *b*, called the **slope** of the line, is the amount by which *y* increases when *x* increases by 1 unit. The value of *a*, called the **intercept** (sometimes the ***y* – intercept**) of the line, is the height of the line above the value *x* = 0.

- The slope of the least-squares line is

  $b = \dfrac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$ and the *y* – intercept is $a = \bar{y} - b\bar{x}$

  We write the equation of the **Least Squares Regression Line (LSRL)** as

  $$\hat{y} = a + bx$$

  where the ∧ above indicates that $\hat{y}$ (read as *y* – hat) is a **prediction** of *y* resulting from the substitution of a particular *x* value into the equation.

Ex1 You are given the following set of observations for variables x and y.

x	-3	-1	1	3
y	8	4	5	-1

Write the LSRL for this data.

Put data in L_1 and L_2. Stat -> Calc -> 8. LinReg(a + bx) -> L_1, L_2, Y_1

Ex2 Studies have shown that people who suffer sudden cardiac arrest (SCA) have a better chance of survival if a defibrillator shock is administered very soon after cardiac arrest. How is survival rate related to the time between when cardiac arrest occurs and when the defibrillator shock is delivered?

Data are given below where *y* = survival rate (percent) and *x* = mean call-to-shock time (minutes) for a cardiac rehabilitation center (where cardiac arrests occurred while victims were hospitalized and so the call-to-shock time tended to be short) and for four communities of different sizes:

Mean call-to-shock time, x	2	6	7	9	12
Survival rate, y	90	45	30	5	2

Find the least square regression line. What is the slope? What is the y intercept?

Interpreting Slope and *y*-intercept of LSRL

- **Slope - ALWAYS use the words "predicted average" for free response questions, as in:** ***for every one unit increase in x the "predicted average" increase/decrease in y is ____ (units).***
- ***y*-intercept - ALWAYS use the word "predicted" for free response, as in:** ***when x = 0 (units) the predicted y value is ____(units).***

Interpret the slope in the context of the problem.

Interpret the y intercept in the context of the problem.

Use this line to predict the SCA survival rate for a community with a mean call-to-shock time of 5 minutes.

Should this line be used to predict the SCA survival rate for a community with a mean call-to-shock time of 20 minutes? Why/why not?

This is an example of **extrapolation.**

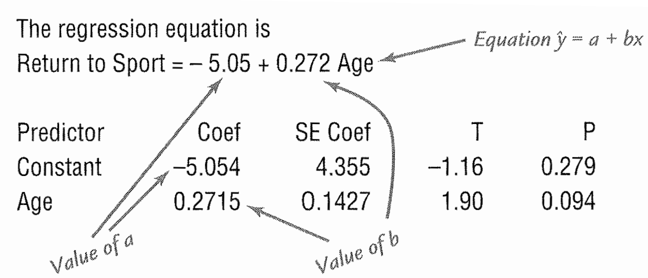
Ex3 How quickly can athletes return to their sport following injuries requiring surgery? We are given the following data on x = age and y = days after arthroscopic shoulder surgery before being able to return to their sport, for 10 weight lifters:

x	33	31	32	28	33	26	34	32	28	27
y	6	4	4	1	3	3	4	2	3	2

Calculate the LSRL. Interpret the slope and y intercept in context.

Take a look at the following summary. This is called a MINITAB output. MINITAB is a very popular and widely used statistical program.

Figure 5.11 Partial MINITAB output for Example 5.7.


The regression equation is
Return to Sport = – 5.05 + 0.272 Age

Predictor	Coef	SE Coef	T	P
Constant	–5.054	4.355	–1.16	0.279
Age	0.2715	0.1427	1.90	0.094

Ex4 A random sample of moving times (in minutes) and weights (in pounds) were recorded for 20 moving jobs requiring three-man crews, and the results of the regression analysis are shown below. The equation for the LSRL is

Predictor	Coef	StDev	T	P
Constant	21.84	25.54	0.86	0.404
Weight	0.036538	0.002977	12.27	0.000

S = 30.32 R – Sq = 89.3% R – Sq(adj) = 88.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	138434	138434	150.60	0.000
Residual Error	18	16546	919		
Total	19	154980			

(a) $\widehat{\text{Weight}} = 21.84 + 0.037(\text{Time})$

(b) $\widehat{\text{Time}} = 21.84 + 0.037(\text{Weight})$

(c) $\widehat{\text{Weight}} = 25.54 + 0.003(\text{Time})$

(d) $\widehat{\text{Time}} = 25.54 + 0.003(\text{Weight})$

(e) $\widehat{\text{Time}} = 0.037 + 21.84(\text{Weight})$

Checkpoint:
Multiple Choice

1. If you're attempting to predict a value of the response variable using a value of x that is outside the range of observed x values in your data set, you're conducting a process of:

(a) predicting the slope of the regression line.
(b) interpolation.
(c) computing residuals.
(d) extrapolation.
(e) slope interpretation.

Free Response
(1999 Q1) Lydia and Bob were searching the internet to find information on air travel in the United States. They found data on the number of commercial aircraft in the United States during the years 1990 - 1998. The dates were recorded as years since 1990. Thus, the year 1990 was recorded as year 0. They fit a least squares regression line to the data. The computer output for their regression are given below.

Predictor	Coef	Stdev	t-ratio	p
Constant	2939.93	20.55	143.09	0.000
Years	233.517	4.316	54.11	0.000
s = 33.43				

(a) What is the value of the slope of the LSRL? Interpret the slope in the context of this situation.
(b) What is the value of the intercept of the LSRL? Interpret the intercept in the context of this situation.
(c) What is the predicted number of commercial aircraft flying in 1992?