Chapter 11 Review

MULTIPLE CHOICE.

1. Which of the following statements is (are) true?
   I. In the computer output for regression, s is the estimator of σ, the standard deviation of the residuals.
   II. The t-test statistic for the H_0: $\beta = 0$ has the same value as the t-test statistic for H_0: $\rho = 0$.
   III. The t-test for the slope of the regression line is always two sided (H_A: $\beta \neq 0$).

   A. I only
   B. II only
   C. III only
   D. I and II only
   E. I and III only

Use the following output in answering questions 2-4:
A study attempted to establish a linear relationship between IQ score and musical aptitude. The following table is a partial printout of the regression analysis and is based on a sample of 20 individuals.

The regression equation is
MusApt = –22.3 + 0.493 IQ

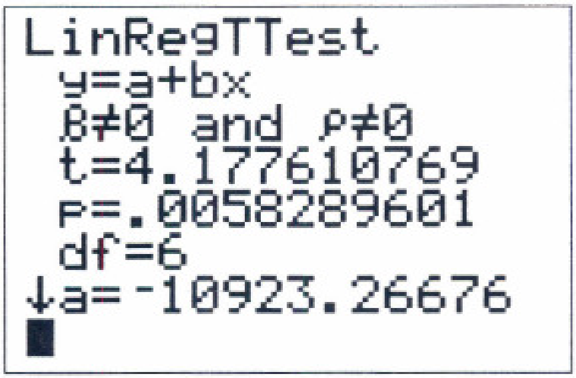
Predictor	Coef	St Dev	t ratio	P
Constant	–22.26	12.94	–1.72	.102
IQ	0.4925	0.1215		
s = 6.143	R-sq = 47.7%	R-sq(adj) = 44.8%		

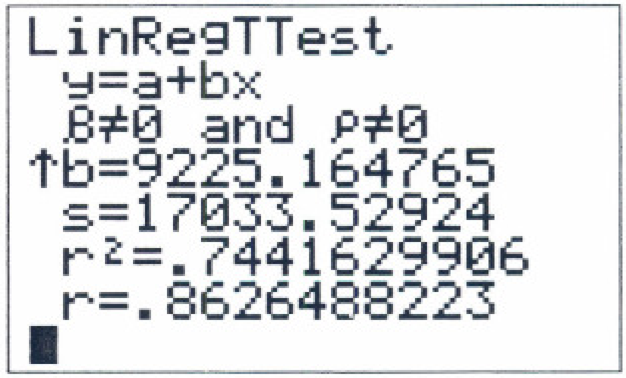
2. The value of the t-test statistic for H_0: $\beta = 0$ is
   A. 4.05
   B. -1.72
   C. 0.4925
   D. 6.143
   E. 0.0802

3. A 99% confidence interval for the slope of the regression line is
   A. 0.4925 ± 2.878(6.143)
   B. 0.4925 ± 2.861(0.1215)
   C. 0.4925 ± 2.861(6.143)
   D. 0.4925 ± 2.845(0.1215)
   E. 0.4925 ± 2.878(0.1215)

4. Which of the following best interprets the slope of the regression line?
   A. For each increase of one IQ point, the Musical Aptitutde score increases by 0.4925 points.
   B. As IQ score increases, so does the Musical Aptitude score.
   C. For each increase of one IQ point, the Musical Aptitude scores is prediced to increase by 0.4925 points.
   D. For each additional point of Musical Aptitude, IQ is predicted to increase by 0.4925 points.
   E. There is a strong predictive linear relationship between IQ score and Musical Aptitude.

5. The two screens shown below were taken from a TI-83/84 LinRegTTest. What is the standard error of the slope of the regression line (s_b)?





A. 17033.53 B. 6953.91 C. 2206.98
D. 9225.16 ± 17033.53 E. 3115.84

6. A group of 12 students take both the SAT Math and the SAT Verbal. The least-squares regression line for predicting Verbal Score from Math Score is determined to be $Verbal\ Score = 106.56 + 0.74(Math\ Score)$. Further, $s_b = 0.11$. Determine a 95% confidence interval for the slope of the regression line.
A. 0.74 ± 0.245 B. 0.74 ± 0.242
C. 0.74 ± 0.240 D. 0.74 ± 0.071
E. 0.74 ± 0.199

FREE RESPONSE.
Use the following table to answer questions 1-5.
The following table gives the ages in months of a sample of children and their mean height (in inches) at that age.

Age	18	19	20	21	22	23	24	25	26	27	28
Height	30.0	30.7	30.7	30.8	31.0	31.4	31.5	31.9	32.0	32.6	32.9

1. Find the correlation coefficient and the least-squares regression line for predicting height (in inches) from age (in months).

2. Draw a scatterplot of the data and the LSRL on the plot. Does the line appear to be a good model for the data?

3. Construct a residual plot for the data. Does the line still appear to be a good model for the data?

4. Use your LSRL to predict the height of a child of 35 months. How confident should you be in this prediction?

5. Interpret the slope of the regression line found in question #1 in the context of the problem.

6. In 2002, there were 23 states in which more than 50% of high school graduates took the SAT test. The following printout gives the regression analysis for predicting SAT Math from SAT Verbal from these 23 states.

The regression equation is

Predictor	Coef	St Dev	t ratio	P
Constant	185.77	71.45	2.60	0.017
Verbal	0.6419	0.1420	4.52	0.000
$s = 7.457$	R-sq = 49.3%		R-sq(adj) = 46.9%	

a. What is the equation of the least-squares regression line for predicting Math SAT score from Verbal SAT score?

b. Interpret the slope of the regression line and interpret in the context of the problem.

c. Identify the standard error of the slope of the regression line and interpret in the context of the problem.

d. Identify the standard error of the residuals and interpret it in the context of the problem.

e. Assuming that the conditions needed for doing inference for regression are present, what are the hypotheses being tested in this problem, what test statistic is used in the analysis, what is its value, and what conclusion would make you make concerning the hypothesis?

7. For the regression analysis of question #6:
   a. Construct and interpret a 95% confidence interval for the true slope of the regression line.

   b. Explain what is meant by “95% confidence interval” in the context of the problem.

8. It has been argued that the average score on the SAT test drops as more students take the test (nationally, about 46% of graduating students took the SAT). The following data are the Minitab output for predicting SAT Math score from the percentage taking the test (PCT) for each of the 50 states. Assuming that the conditions for doing inference for regression are met, test the hypothesis that the scores decline as the proportion of students taking the test rises. That is, test to determine if the slope of the regression line is negative. Test at the 0.01 level of significance.

The regression equation is SAT Math = 574 – 99.5 PCT

Predictor	Coef	St Dev	t ratio	P
Constant	574.179	4.123	139.25	0.000
PCT	–99.516	8.832	–11.27	0.000
$s = 17.45$	R-sq = 72.6%	R-sq(adj) = 72.0%		

9. Some bored researchers got the idea that they could predict a person's pulse rate from his or her height (earlier studies had shown a very weak linear relationship between pulse rate and weight). They collected data on 20 college-age women. The following table is part of the Minitab output of their findings.

The regression equation is

Pulse = [ ] Height

Predictor	Coef	St Dev	*t* ratio	*P*
Constant	52. 00	37.24	1.40	0.180
Height	0.2647	0.5687	[ ]	
$s = 10.25$	R-sq = 1.2%	R-sq(adj) = 0.0%		

a. Determine the t-ratio and the P-value for the test.

b. Construct a 99% confidence interval for the slope of the regression line used to predict pulse rate from height.

c. Do you think there is a predictive linear relationship between height and pulse rate? Explain.

d. Suppose the researcher was hoping to show that there was a positive linear relationship between pulse rate and height. Are the t-ratio and P-value the same as in part (a)? If not, what are they?

10. The following table gives the number of manatees killed by powerboats along the Florida coast in the years 1977 to 1990, along with the number of powerboat registrations (in thousands) during those years:

Year	Powerboat Registrations	Manatees Killed
1977	447	13
1978	460	21
1979	481	24
1980	498	16
1981	513	24
1982	512	20
1983	559	34
1984	559	34
1985	585	33
1986	614	33
1987	645	39
1988	675	43
1989	711	50
1990	719	47

a. Test the hypothesis that there is a positive linear relationship between the number of powerboat registrations and the number of manatees killed by powerboats. Assume the conditions needed to do interference for regression have been met.

b. Use a residual plot to assess the appropriateness of the model.

c. Construct and interpret a 90% confidence interval for the true slope of the regression line (that is, find a 90% confidence interval for the predicted number of additional manatees killed for each additional registered powerboat).

Chapter 11 KEY

MULTIPLE CHOICE.

1.

The correct answer is (d). II is true since it can be shown that $t = \frac{b}{s_b} = r\sqrt{\frac{n-2}{1-r^2}}$.

III is not true since, although we often use the alternative H_A: $\beta \neq 0$, we can certainly test a null with an alternative that states that there is a positive or a negative association between the variables.

2.

The correct answer is (a). $t = \frac{b}{s_b} = \frac{0.4925}{0.1215} = 4.05$.

3.

The correct answer is (e). For $n = 20$, df $= 20 - 2 = 18 \Rightarrow t^* = 2.878$ for $C = 0.99$.

4.

The correct answer is (c). Note that (a) is not correct since it doesn't have "predicted" or "on average" to qualify the increase. (b) Is a true statement but is not the best interpretation of the slope. (d) Has mixed up the response and explanatory variables. (e) Is also true ($t = 4.05 \Rightarrow P$-value $= 0.0008$) but is not an interpretation of the slope.

5.

The correct answer is (c). The TI-83/84 does not give the standard error of the slope directly. However,

$$t = \frac{b}{s_b} \Rightarrow s_b = \frac{b}{t} = \frac{9225.16}{4.18} = 2206.98.$$

6.

The correct answer is (a). A 95% confidence interval at $12 - 2 = 10$ degrees of freedom has a critical value of $t^* = 2.228$ (from Table B; if you have a TI-84 with the `invT` function, `invT(0.975,10)=2.228`). The required interval is $0.74 \pm (2.228)(0.11) = 0.74 \pm 0.245$.

FREE RESPONSE.

1.

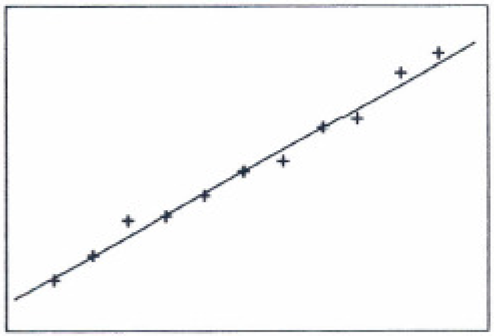
$r = 0.9817$, height $= 25.41 + 0.261$(age)

(Assuming that you have put the age data in `L1` and the height data in `L2`, remember that this can be done on the TI-83/84 as follows: `STAT CALC LinReg(a+bx) L1,L2,Y1`.)

2.

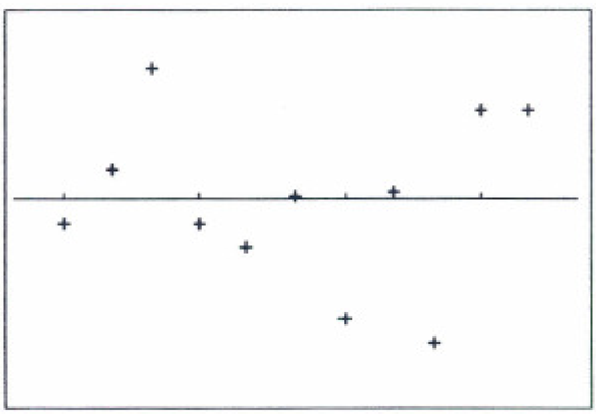
The line does appear to be a good model for the data.

(After the regression equation was calculated on the TI-83/84 and the LSRL stored in `Y1`, this was constructed in `STAT PLOT` by drawing a scatterplot with `Xlist:L1` and `Ylist:L2`.)



3.

The residual pattern seems quite random. A line still appears to be a good model for the data.



(This scatterplot was constructed on the TI-83/84 using `STAT PLOT` with `Xlist:L1` and `Ylist:RESID`. Remember that the list of residuals for the most recent regression is saved in a list named `RESID`.)

4.

Height = 25.41 + 0.261(35) = 34.545 (`Y1(35) = 34.54`). You probably shouldn't be too confident in this prediction. 35 is well outside of the data on which the LSRL was constructed and, even though the line appears to be a good fit for the data, there is no reason to believe that a linear pattern is going to continue indefinitely. (If it did, a 25-year-old would have a predicted height of 25.41 + 0.261(12 × 25) = 103.71″, or 8.64 feet!)

5.

The slope of the regression line is 0.261. This means that, for an increase in age of 1 month, height is predicted to increase by 0.261 inches. You could also say, that, for an increase in age of 1 month, height will increase on average by 0.261 inches.

6.

a. $Math = 185.77 + 0.6419(Verbal)$.

b. $b = 0.6419$. For each additional point scored on the SAT Verbal test, the score on the SAT Math test is predicted to increase by 0.6419 points (or: will increase *on average* by 0.6419 points). (Very important on the AP exam: be very sure to say "is predicted" or "on average" if you'd like maximum credit for the problem!)

c. The standard error of the slope is $s_b = 0.1420$. This is an estimate of the variability of the standard deviation of the estimated slope for predicting SAT Verbal from SAT Math.

d. The standard error of the residuals is $s = 7.457$. This value is a measure of variation in SAT Verbal for a fixed value of SAT Math.

e. • The hypotheses being tested are $H_o: \beta = 0$ (which is equivalent to $H_o: \rho = 0$) and $H_A\ \beta \neq 0$, where β is the slope of the regression line for predicting SAT Verbal from SAT Math.

• The test statistic used in the analysis is $t = \frac{b}{s_b} = \frac{0.6419}{0.1420} = 4.52$, df = 23 − 2 = 21.

7.

a. df = 23 − 2 = 21 $\Rightarrow t^* = 2.080$. The 95% confidence interval is: $0.6419 \pm 2.080(0.1420) = (0.35, 0.94)$. We are 95% confident that, for each 1 point increase in SAT Verbal, the true increase in SAT Math is between 0.35 points and 0.94 points.

b. The procedure used to generate the confidence interval would produce intervals that contain the true slope of the regression line, on average, 0.95 of the time.

8.

I. Let β = the true slope of the regression line for predicting SAT Math score from the percentage of graduating seniors taking the test.

$$H_0: \beta = 0.$$

$$H_A: \beta < 0.$$

II. We use a linear regression t test with $\alpha = 0.01$. The problem states that the conditions for doing inference for regression are met.

III. We see from the printout that

$$t = \frac{b}{s_b} = \frac{-99.516}{8.832} = -11.27$$

based on 50 − 2 = 48 degrees of freedom. The P-value is 0.000. (Note: The P-value in the printout is for a two-sided test. However, since the P-value for a one-sided test would only be half as large, it is still 0.000.)

IV. Because $P < 0.01$, we reject the null hypothesis. We have very strong evidence that there is a negative linear relationship between the proportion of students taking SAT math and the average score on the test.

9.

a. $t = \dfrac{b}{s_b} = \dfrac{.2647}{.5687} = 0.47$, df $= 20 - 2 = 18 \Rightarrow P$-value $= 0.644$.

b. df $= 18 \Rightarrow t^* = 2.878$; $0.2647 \pm 2.878(0.5687) = (-1.37, 1.90)$.

c. No. The P-value is very large, giving no grounds to reject the null hypothesis that the slope of the regression line is 0. Furthermore, the correlation coefficient is only $r = \sqrt{0.012} = 0.11$, which is very close to 0. Finally, the confidence interval constructed in Part (b) contains the value 0 as a likely value of the slope of the population regression line.

d. The t-ratio would still be 0.47. The P-value, however, would be half of the 0.644, or 0.322 because the computer output assumes a two-sided test. This is a lower P-value but is still much too large to infer any significant linear relationship between pulse rate and height.

10.

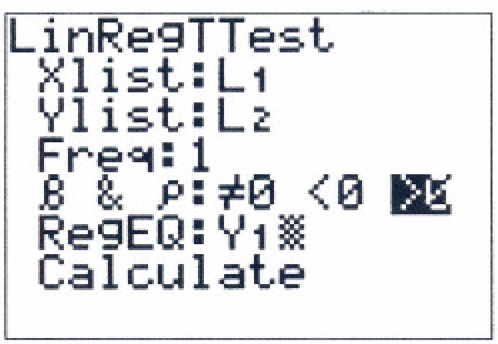
a. **I**. Let β = the true slope of the regression line for predicting the number of manatees killed by powerboats from the number of powerboat registrations.

$$H_0: \beta = 0.$$
$$H_A: \beta > 0.$$

II. We use a t-test for the slope of the regression line. The problem tells us that the conditions necessary to do inference for regression are present.

III. We will do this problem using the TI-83/84 as well as Minitab.

- On the TI-83/84, enter the number of powerboat registration in `L1` and the number of manatees killed in `L2`. Then go to `STAT TESTS LinRegTTest` and set it up as shown below:



After "Calculate," we have the following two screens.

```
LinRegTTest
 y=a+bx
 β>0 and ρ>0
 t=9.675470539
 p=2.5545306E-7
 df=12
↓a=-41.43043895
```

```
LinRegTTest
 y=a+bx
 β>0 and ρ>0
↑b=.1248616923
 s=4.276387771
 r²=.8863794853
 r=.9414772888
```

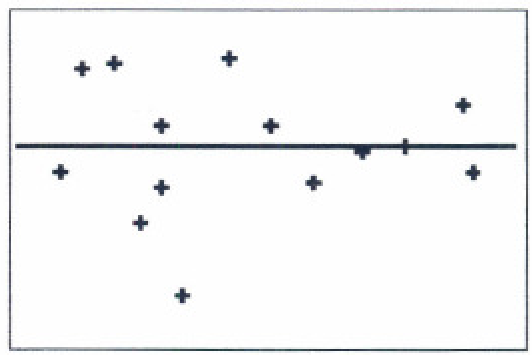
The Minitab output for this problem is:

The regression equation is Man = – 41.4 + 0.125 PB Reg

Predictor	Coef	St Dev	t ratio	P
Constant	–41.430	7.412	–5.59	0.000
PB Reg	0.12486	0.01290	9.68	0.000
s = 4.276	R-sq = 88.6%	R-sq(adj) = 87.7%		

IV. Because the P-value is very small, we reject the null. We have very strong evidence of a positive linear relationship between the number of powerboat registrations and the number of manatees killed by powerboats.

b. Using the residuals generated when we did the linear regression above, we have:



There appears to be no pattern in the residual plot that would cause us to doubt the appropriateness of the model. A line does seem to be a good model for the data.

c. (i) Using the TI-83/84 results, df = 12 $\Rightarrow t^*$=1.782. We need to determine s_b. We have $t = \frac{b}{s_b} \Rightarrow s_b = \frac{b}{t} = \frac{0.125}{9.68} = 0.013$. The confidence interval is 0.125 ± 1.782(0.013) = (0.10, 0.15).

(ii) Directly from the Minitab output: 0.125 ± 1.782(0.013) = (0.10, 0.15).