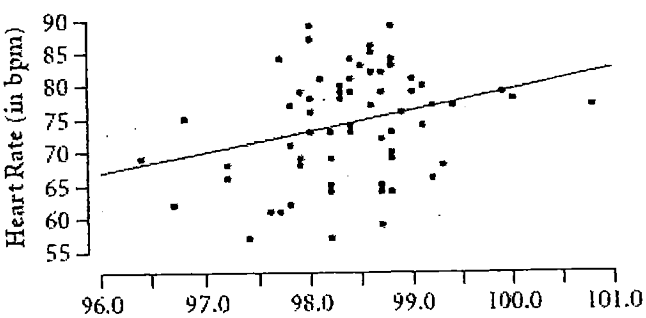
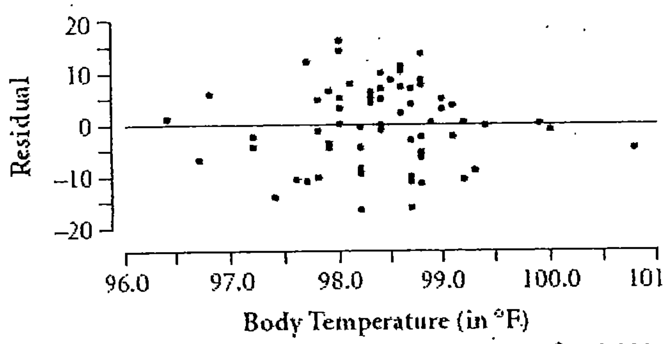
AP Statistics
Chapter 15 Review

Name: Key
Hour: ______

1. As part of a 1992 study attempting to verify that the "true" mean body temperature is 98.6°F, the heart rates, in beats per minute (bpm), and body temperatures of 26 women were taken at several different times during two consecutive days, resulting in a total of 65 readings. Assume that these women can be considered a random sample of the population of adult women. The scatterplot (with residual plot) and corresponding regression analysis of heart rate versus body temperature are shown here.





Heart Rate = 3.13 *Body Temp* − 233.6; $r^2 = 0.082$

R squared = 0.082
Dependent variable: Heart Rate
$s = 7.83$

Source	Sum of Squares	DF	Mean Square	F Ratio	prob
Regression	346.15	1	346.15	5.65	0.023
Residual	3858.31	63	61.24		

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	-233.6	129.53	-1.80	0.077
Body Temp	3.13	1.32	2.377	0.021

Slope $\quad b_1$

invT(.99, 63) = 2.387

Which of the following best represents a 98% confidence interval for the slope of the regression line and its interpretation?

A. 3.13 ± 2.390(1.32); the heart rate increase per degree increase in body temperature is 3.13 bpm, give or take about 3 bpm.

B. 3.13 ± 2.390(61.24); the heart rate increase per degree increase in body temperature is 3.13 bpm, give or take about 122 bpm.

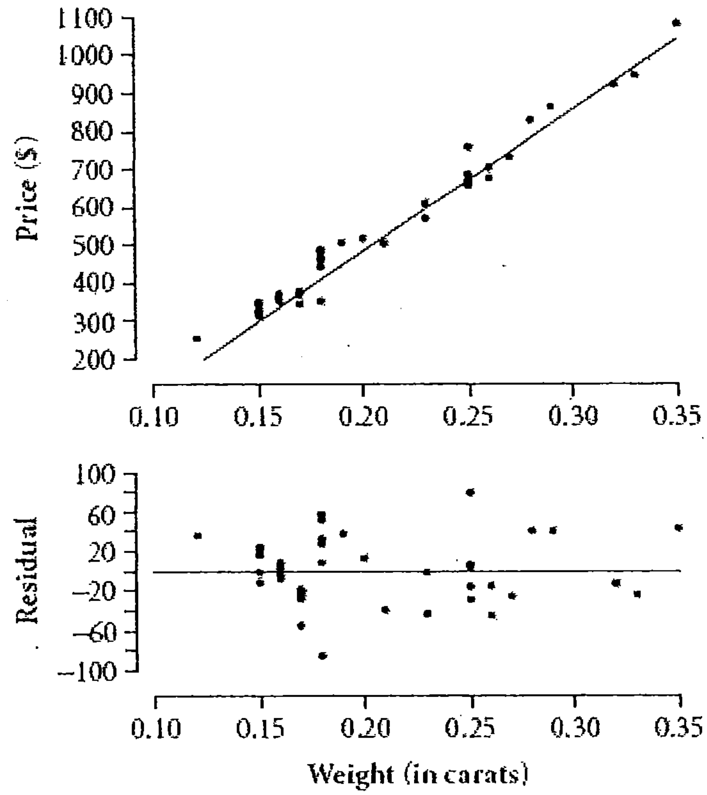
C. 3.13 ± 2.390(7.83); the heart rate increase per degree increase in body temperature is 3.13 bpm, give or take about 16 bpm.

D. 3.13 ± 2.326(7.83); the heart rate increase per degree increase in body temperature is 3.13 bpm, give or take about 15 bpm.

E. 3.13 ± 2.326(61.24); the heart rate increase per degree increase in body temperature is 3.13 bpm, give or take about 120 bpm.

Use the following data to answer #2-6:

The price of a diamond ring is based partly on the market value of its gold content, the worker's time needed to craft the ring, and the cost of the diamond. The weight and price of 48 diamond rings weighing from 0.12 to 0.35 carats (one carat = 0.2 gram) and priced between \$223 and \$1086 were taken from an advertisement placed in a Singapore [illegible] this range of weight that were available in Singapore on that date. The scatterplot (with residual plot) and corresponding regression analysis of the cost versus the number of carats are shown.



R squared = 0.978
Dependent variable: Price
s = 31.84

Source	Sum of Squares	DF	Mean Square	F Ratio	prob
Regression	2098596.0	1	2098596	2069.991	< .0001
Residual	46635.7	46	1014		

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	-259.6	17.3	14.99	< .0001
Weight (carats)	3721.0	81.8		

Source: *Journal of Statistics Education* 4(3) (1996), www.amstat.org/publications/jse/v4n3/datasets.chu.html.

2. Find the equation of the LSR line. Be sure to define any variables used.

$\hat{y} = -259.6 + 3721x$

$\hat{y}$ = predicted price in dollars
x = weight in carats

3. Interpret the slope of the LSR line.

3721: For each additional carat the price of the ring increases by an average of \$3721.

4. Find s_{b_1} on the chart. Interpret this value in context.

81.8: The standard error of the slope of the LSR line relating diamond weights in carats and price in dollars for all diamond rings in the sample.

5. Find the missing test statistic and *P*-value from the chart.

$t = \frac{3721 - 0}{81.8} = 45.489$ $P \approx 0$ $df = 46$

6. Based on the answer to #5, what is your conclusion?

Since the p-value $= 0 < \alpha = .05$, we reject H_0. There is significant evidence of a linear relationship between weight in carats and price in dollars for all diamond rings.

7. Suppose a test of significance of the null hypothesis that the true slope of a regression line is 0 produces a *P*-value less than 0.05. Which of the following is an accurate conclusion?

A. The 95% confidence interval estimate of the slope will not include 0.
B. The 95% confidence interval estimate of the slope will include 0.
C. The 95% confidence interval estimate of the slope will not include 0.05.
D. The 95% confidence interval estimate of the slope will include 0.05.
E. There is insufficient evidence to reject the null hypothesis that the slope of the true regression line is 0.

8. In a study of the performance of a computer printer, the size (in kilobytes) and the printing time (in seconds) for each of 22 small text files were recorded. A regression line was a satisfactory description of the relationship between size and printing time. The results of the analysis are shown below.

Dependent variable: Printing Time

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	53.3315	1	53.3315	140
Residual	7.62381	20	0.38115	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	11.6559	0.3153	37	≤ 0.0001
Size	3.47812 slope	0.294 s_{b_1}	11.8	≤ 0.0001

R squared = 87.5% R squared (adjusted) = 86.9%
s = 0.6174 with 22 − 2 = 20 degrees of freedom

p-value = 2.086
invT(.975, 20)

Which of the following should be used to compute a 95 percent confidence interval for the slope of the regression line?

A. $3.478 \pm 2.086 \times 0.294$
B. $3.478 \pm 1.96 \times 0.617$
C. $3.478 \pm 1.725 \times 0.294$
D. $11.656 \pm 2.086 \times 0.315$
E. $11.656 \pm 1.725 \times 0.315$

9. Do better home run hitters in professional baseball have higher batting averages? The following table shows a random sample of 15 American League baseball players' statistics from the 2003 season.

Home Runs	Batting Average
12	243
2	271
15	317
6	245
16	265
19	275
13	299
14	245
33	317
14	284
2	204
8	259
18	253
11	295
7	261

Source: *The World Almanac and Book of Facts 2004* (New York: World Almanac, 2003), pp. 894–95.

(a) Find a good-fitting model that can be used to predict the batting average given the number of home runs

$\hat{y} = 239.562 + 2.314x$

$\hat{y}$ = predicted batting average

x = # of homeruns

(b) Do these data provide convincing evidence that the batting average is related to the number of home runs for American League baseball players?

① $H_0: \beta_1 = 0 \quad H_a: \beta_1 \neq 0$

② t-Test for the slope of the true regression equation relating the # of home runs and batting average for all American League baseball season in 2003

③ The scatterplot of the sample appears to be linear. SRS is stated and we assume each player is measured independently. The boxplot of residuals is ~N. The residual plot shows no apparent pattern or fanning

④ $t = \dfrac{2.314 - 0}{.8862} = 2.611 \quad df = 15 - 2 = 13 \quad$ p-value = .022

⑤ Since the p-value $.022 < \alpha = .05$, we reject the H_0. There is evidence that the batting average is related to the # of homeruns for American League baseball players.

10. Windmills generate electricity by transferring energy from wind to a turbine. A study was conducted to examine the relationship between wind velocity in miles per hour (mph) and electricity production in amperes for one particular windmill. For the windmill, measurements were taken on twenty-five randomly selected days, and the computer output for the regression analysis for predicting electricity production based on wind velocity is given below. The regression model assumptions were checked and determined to be reasonable over the interval of wind speeds represented in the data, which were from 10 miles per hour to 40 miles per hour.

Predictor	Coef	SE Coef	T	P
Constant	0.137	0.126	1.09	0.289
Wind velocity	0.240	0.019	12.63	0.000

S = 0.237 R-Sq = 0.873 R-Sq (adj) = 0.868

(a) Use the computer output above to determine the equation of the least squares regression line. Identify all variables used in the equation.

$\hat{y} = .137 + .240x$

$\hat{y}$ = predicted electricity production (amperes)
x = wind velocity (mph)

(b) How much more electricity would the windmill be expected to produce on a day when the wind velocity is 25 mph than on a day when the wind velocity is 15 mph? Show how you arrived at your answer.

$\hat{y}_{25} = .137 + .240(25) = 6.137$

$\hat{y}_{15} = .137 + .240(15) = 3.737$

$6.137 - 3.737$

$\boxed{\text{2.4 more amps}}$

(c) What proportion of the variation in electricity production is explained by its linear relationship with wind velocity?

.873 = 87.3% of the variation in electricity is explained...

(d) Is there statistically convincing evidence that electricity production by the windmill is related to wind velocity? Explain.

Since the p-value $0 < \alpha = .05$, we reject H_0. There is evidence that electricity production by the windmill is related to wind velocity.

11. A study was designed to explore subjects' ability to judge the distance between two objects placed in a dimly lit room. The researcher suspected that the subjects would generally overestimate the distance between the objects in the room and that this overestimation would increase the farther apart the objects were.

The two objects were placed at random locations in the room before a subject estimated the distance (in feet) between those two objects. After each subject estimated the distance, the locations of the objects were rerandomized before the next subject viewed the room.

After data were collected for 40 subjects, two linear models were fit in an attempt to describe the relationship between the subjects' perceived distances (y) and the actual distance, in feet, between the two objects.

$$\text{Model 1: } \hat{y} = 0.238 + 1.080 \times (\text{actual distance})$$

The standard errors of the estimated coefficients for Model 1 are 0.260 and 0.118, respectively.

$$\text{Model 2: } \hat{y} = 1.102 \times (\text{actual distance})$$

The standard error of the estimated coefficient for Model 2 is 0.393.

(a) Provide an interpretation in context for the estimated slope in Model 1.

The predicted perceived distance increases by 1.080 feet on average for each additional foot of actual distance between the objects.

(b) Explain why the researcher might prefer Model 2 to Model 1 in this context.

The y-intercept of 0 in model 2 matches the predicted distance of 0.

(c) Using Model 2, test the researcher's hypothesis that in dim light participants overestimate the distance, with the overestimate increasing as the actual distance increases. (Assume appropriate conditions for inference are met.)

$H_0: B_1 = 1 \quad H_a: B_1 > 1$

t-Test for slope of the true regression equation (B_1) relating the actual and perceived distance between objects for all students.

we are told to assume all conditions are met.

$t = \frac{1.102 - 1}{.393} = .260 \quad df = 40-2 = 38 \quad$ p-value $= .398$

Since the p-value $= .398 > \alpha = .05$, we fail to reject H_0. There is not evidence that a participant will overestimate distance in dim light.

The researchers also wanted to explore whether the performance on this task differed between subjects who wear contact lenses and subjects who do not wear contact lenses. A new variable was created to indicate whether or not a subject wears contact lenses. The data for this variable were coded numerically (1 = contact wearer, 0 = noncontact wearer), and this new variable, named "contact," was included in the following model.

$$\text{Model 3: } \hat{y} = 1.05 \times (\text{actual distance}) + 0.12(\text{contact}) \times (\text{actual distance})$$

The standard errors of the estimated coefficients for Model 3 are 0.357 and 0.032, respectively.

(d) Using Model 3, sketch the estimated regression model for contact wearers and the estimated regression model for noncontact wearers on the grid below.

with contacts:

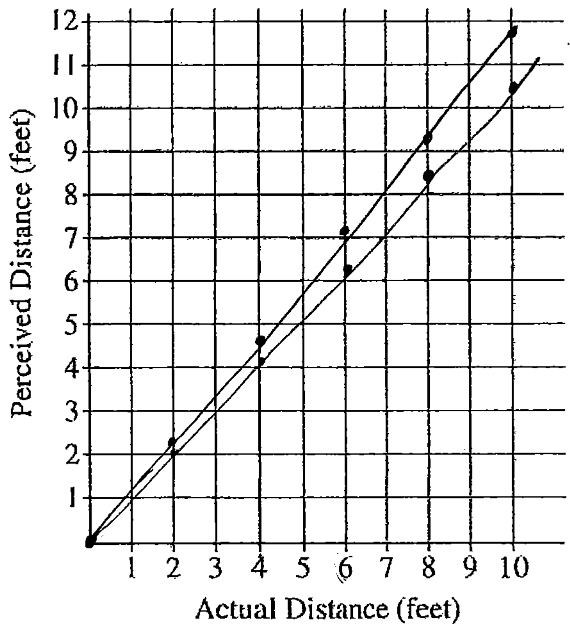
$\hat{y} = 1.05x + .12(1)x$

$\hat{y} = 1.17x$

without contacts:

$\hat{y} = 1.05x + .12(0)x$

$\hat{y} = 1.05x$



X	contacts	no contacts
2	2.34	2.1
4	4.68	4.2
6	7.02	6.3
8	9.36	8.4
10	11.7	10.5

$\hat{y}$ = Perceived distance
x = actual distance

(e) In the context of this study, provide an interpretation of the estimated coefficients for Model 3.

Individuals with or without contacts have an average increased perceived distance of 1.05 ft for each foot of actual distance. Those with contacts increased an additional .12 feet for each ft of actual distance.

(or) Individual wearing contacts have an average increase in perceived distance of 1.17 ft for each foot of actual distance. While those without contacts average an increase of 1.05 feet per 1 foot of actual