

Text Exercise Set 31

NAME:

31-1 In Exercise 11-3, the relationship between the number of hours of training to perform a certain job and the number of minutes necessary to perform the job is being investigated, with regard to the prediction of time to perform the job from training hours. The Task Training Data, displayed as Table 11-1, was recorded for several subjects. For convenience, we have redisplayed the data here in Table 31-3.

- (a) Verify that the least squares line (found in Exercise 11-3) is $prf = 6.5 - 0.07(trn)$, and write a one sentence interpretation of the slope of the least squares line.

Table 31-3

Task Training Data

Each subject in a sample spends time training before performing a certain task. The number of hours of training and the number of minutes to perform the task are recorded for each subject.

Training Time (hrs.)	Performance Time (min.)
15	5.2
15	5.9
30	3.7
30	4.7
45	3.8
45	3.1

- (b) A 0.05 significance level is chosen for a hypothesis test to see if there is any evidence that the linear relationship between training time and performance time is negative, or in other words, evidence that the slope in the linear regression to predict performance time from training time is negative. Complete the four steps of the hypothesis test below. You should find that the t test statistic is $t_4 = -3.986$.

Step 1 H_0 :

H_1 :

$\alpha =$

Step 2

Step 3

Step 4

31-1 - continued

- (c) Considering the results of the hypothesis test, decide which of the Type I or Type II errors is possible, and describe this error.
- (d) Decide whether H_0 would have been rejected or would not have been rejected with each of the following significance levels: (i) $\alpha = 0.01$, (ii) $\alpha = 0.10$.

31-2 In Exercise 11-4, the relationship between temperature and the breaking strength of a particular alloy is being studied, with regard to the prediction of breaking strength from temperature. The Temperature and Strength Data, displayed as Table 11-2, was recorded for several pieces of the alloy. For convenience, we have redisplayed the data here in Table 31-4.

Temperature (°F)	Breaking Strength (lbs.)
9	396
50	351
60	320
97	281
123	263
140	227
152	155
178	116
225	61
262	36
277	16

- (a) Verify that the least squares line (found in Exercise 11-4) is $brk = 416.5 - 1.5(tmp)$, and write a one sentence interpretation of the slope of the least squares line.
- (b) A 0.05 significance level is chosen for a hypothesis test to see if there is any evidence that the linear relationship between temperature and breaking strength is significant, or in other words, evidence that the slope in the linear regression to predict breaking strength from temperature is different from zero. Complete the four steps of the hypothesis test below. You should find that the t test statistic is $t_9 = -18.286$.

31-2(b) – *continued*

Step 1 H_0 :
 H_1 :
 $\alpha =$

Step 2

Step 3

Step 4

- (c) Considering the results of the hypothesis test, decide which of the Type I or Type II errors is possible, and describe this error.

- (d) Decide whether H_0 would have been rejected or would not have been rejected with each of the following significance levels: (i) $\alpha = 0.01$, (ii) $\alpha = 0.10$.

31-3 The relationship between high school grade point average and the weekly TV viewing time for high school students in a particular state is being studied, with regard to the prediction of grade point average from weekly TV hours. The individuals selected for the STUDENT DATA, displayed as Data Set 23-2 at the end of Unit 23, are treated as comprising a simple random sample. The portion of the data to be used is displayed on the right.

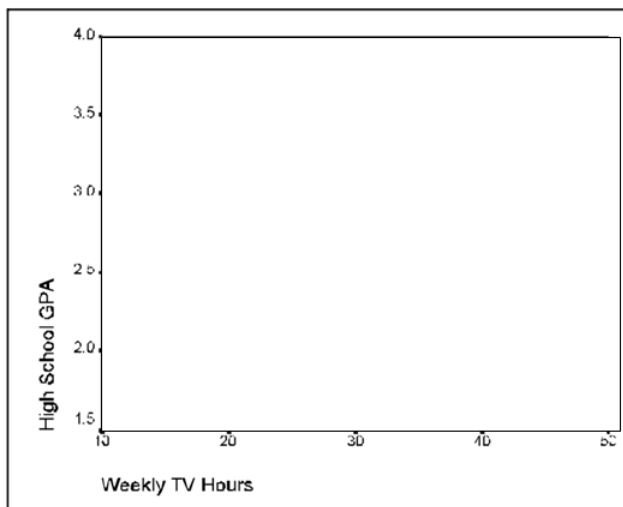
- (a) Identify the response variable Y and the explanatory variable X .

Weekly TV Hours	Grade Point Average
29	3.11
21	3.25
48	2.08
30	3.31
28	2.79
25	3.40
11	3.95
36	2.23
14	2.84
39	2.57
41	2.46
20	3.42
13	3.36
38	2.94
24	2.58
30	2.88
23	2.60
34	2.25
40	2.01
33	2.53
17	3.56
44	1.97
33	2.99
18	3.74
22	3.21
24	2.55
27	3.20

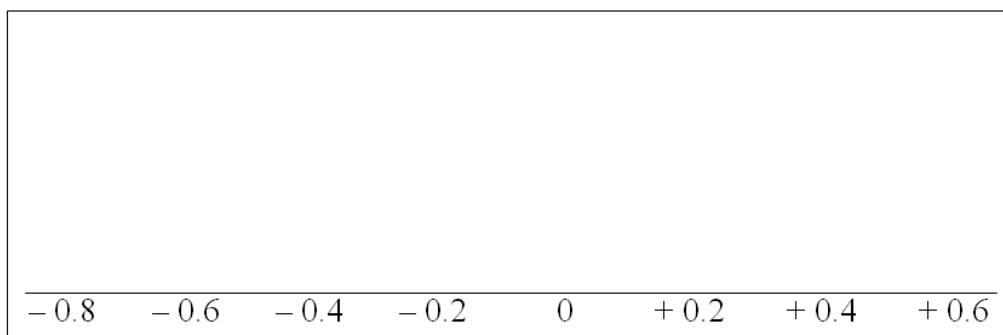
31-3 - continued

- (b) Find the correlation between the variables X and Y .
- (c) Verify that the least squares line to predict high school grade point average (gpa) from weekly TV hours (tvh) is $gpa = 4.081 - 0.04254(tvh)$, and write a one sentence interpretation of the slope of the least squares line.

- (d) Complete the construction of the scatter plot of the data, graph the least squares line on the scatter plot, and decide whether you think the relationship should be considered linear or non-linear.

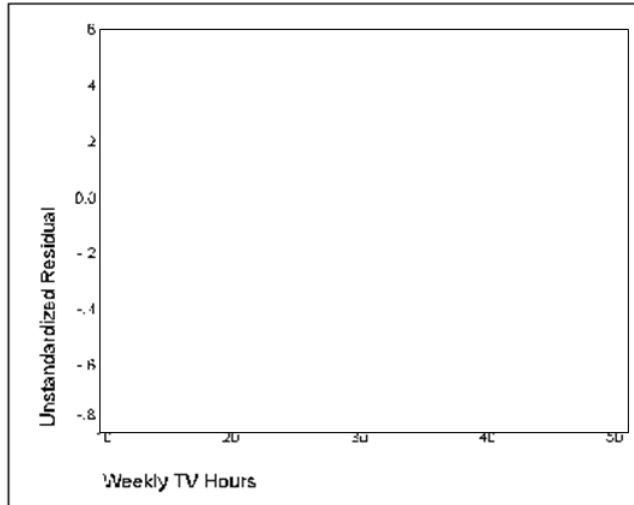


- (e) Find the five-number summary for the residuals, and find the interquartile range for the residuals.
- (f) Decide whether there appear to be any candidates for outliers, and complete the construction of the modified box plot of the residuals.



31-3 - continued

(g) Complete the construction of the residual plot; then, decide whether or not the linearity assumption appears to be reasonable, and state why or not.



(h) A 0.05 significance level is chosen for a hypothesis test to see if there is any evidence that the linear relationship between grade point average and weekly TV hours is significant, or in other words, evidence that the slope in the linear regression of grade point average on weekly TV hours is different from zero. Complete the four steps of the hypothesis test below. You should find that the t test statistic is $t_{25} = -6.182$.

Step 1 H_0 :

H_1 :

$\alpha =$

Step 2

Step 3

Step 4

(i) Considering the results of the hypothesis test, decide which of the Type I or Type II errors is possible, and describe this error.

(j) Decide whether H_0 would have been rejected or would not have been rejected with each of the following significance levels: (i) $\alpha = 0.01$, (ii) $\alpha = 0.10$.

31-3 - *continued*

- (k) Use the least squares line to predict the grade point average of a high school student who watches 25 hours of TV weekly.
- (l) Use the least squares line to estimate the mean grade point average among high school students who watch 40 hours of TV weekly.
- (m) Why would it not be appropriate to use the least squares line to estimate the grade point average of a high school student who watches 5 hours of TV weekly?
- (n) On average, what is the change in grade point average with an increase of 5 hours in weekly TV time?
- (o) Use the least squares line to approximate the weekly TV hours that would correspond to a high school grade point average of 2.5.
- (p) Do the values for the explanatory variable in the data look like they were controlled by the experimenter, or do they look random?

31-4 The relationship between high school grade point average and the weekly studying time for high school students in a particular state is being studied, with regard to the prediction of grade point average from weekly study hours. The individuals selected for the STUDENT DATA, displayed as Data Set 23-2 at the end of Unit 23, are treated as comprising a simple random sample. The portion of the data to be used is displayed on the right.

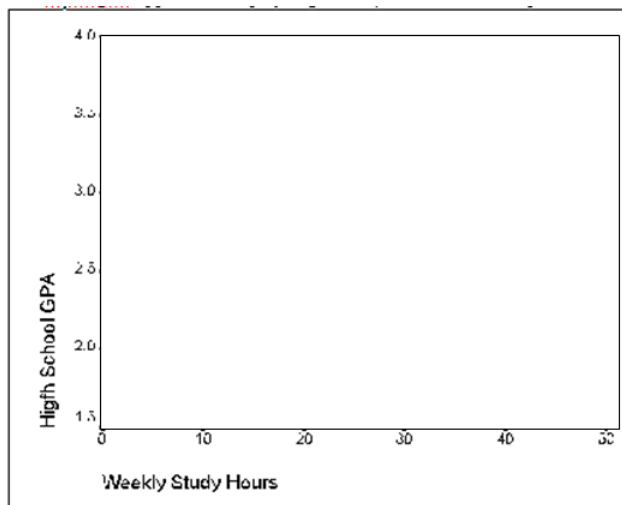
- (a) Identify the response variable Y and the explanatory variable X .

Weekly TV Study Hours	Grade Point Average
17	3.11
27	3.25
9	2.08
22	3.31
28	2.79
27	3.40
46	3.95
16	2.23
32	2.84
14	2.57
2	2.46
35	3.42
41	3.36
17	2.94
21	2.58
29	2.88
26	2.60
21	2.25
12	2.01
19	2.53
29	3.56
10	1.97
19	2.99
40	3.74
37	3.21
31	2.55
21	3.20

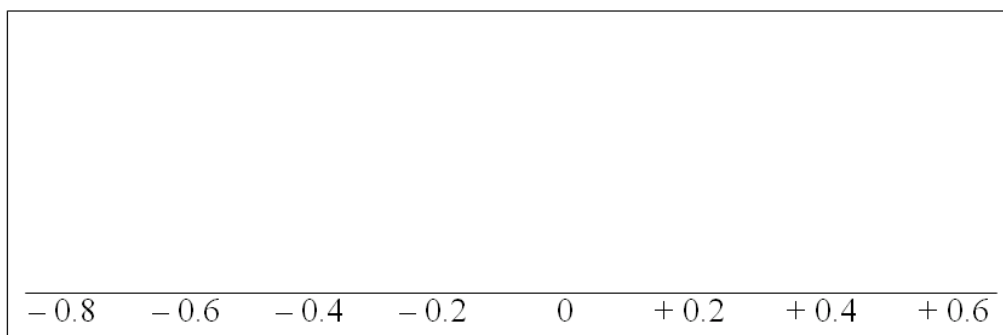
31-4 - continued

- (b) Find the correlation between the variables X and Y .
- (c) Verify that the least squares line to predict high school grade point average (gpa) from weekly study hours (std) is $gpa = 1.965 + 0.03814(std)$, and write a one sentence interpretation of the slope of the least squares line.

- (d) Complete the construction of the scatter plot of the data, graph the least squares line on the scatter plot, and decide whether you think the relationship should be considered linear or non-linear.

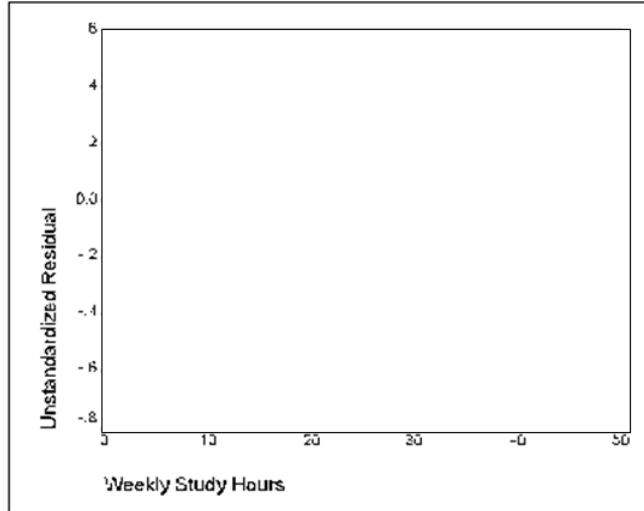


- (e) Find the five-number summary for the residuals, and find the interquartile range for the residuals.
- (f) Decide whether there appear to be any candidates for outliers, and complete the construction of the modified box plot of the residuals.



31-4 - continued

(g) Complete the construction of the residual plot; then, decide whether or not the linearity assumption appears to be reasonable, and state why or not.



(h) A 0.05 significance level is chosen for a hypothesis test to see if there is any evidence that the linear relationship between grade point average and weekly study hours is significant, or in other words, evidence that the slope in the linear regression of grade point average on weekly study hours is different from zero. Complete the four steps of the hypothesis test below. You should find that the t test statistic is $t_{25} = + 5.725$.

Step 1 H_0 :

H_1 :

$\alpha =$

Step 2

Step 3

Step 4

(i) Considering the results of the hypothesis test, decide which of the Type I or Type II errors is possible, and describe this error.

(j) Decide whether H_0 would have been rejected or would not have been rejected with each of the following significance levels: (i) $\alpha = 0.01$, (ii) $\alpha = 0.10$.

31-4 - continued

- (k) Use the least squares line to predict the grade point average of a high school student who studies 35 hours weekly.

- (l) Use the least squares line to estimate the mean grade point average among high school students who study 10 hours weekly.

- (m) Why would it not be appropriate to use the least squares line to estimate the grade point average of a high school student who studies 60 hours weekly?

- (n) On average, what is the change in grade point average with an increase of 3 hours in weekly study time?

- (o) Use the least squares line to approximate the weekly study hours that would correspond to a high school grade point average of 3.5.

- (p) Do the values for the explanatory variable in the data look like they were controlled by the experimenter, or do they look random?

31-5 A standard treatment for an ailment is being compared with two new treatments labeled "New A," and "New B." A 0.05 significance level is chosen for a hypothesis test to see if there is any evidence of a difference in the proportion of complete cures. The data observed on a group of subjects treated as a simple random sample have been organized into the contingency table displayed on the right.

<u>Cure</u>	<u>Type of Major</u>		
	<i>Standard</i>	<i>New A</i>	<i>New B</i>
<i>Complete</i>	52	12	49
<i>Not Complete</i>	48	63	26

- (a) Circle the labels of each correct way to complete the statement below.
 Looking for evidence of a difference in the proportion of complete cures among the "Standard," "New A," and "New B," treatments is the same as looking for evidence of a
- (i) relationship among the three treatments.
 - (ii) difference in the distribution of the three treatments between complete cures and incomplete cures.
 - (iii) difference between cure and treatment.
 - (iv) relationship between cure and treatment.
- (b) Explain how the data for this hypothesis test is appropriate for a chi-square test concerning independence.

(c) Complete the four steps of the hypothesis test below. You should find that $\chi^2_2 = 39.958$.

Step 1 H_0 :

H_1 :

$\alpha =$

Step 2

Step 3

Step 4

31-6 A 0.05 significance level is chosen for a hypothesis test to see if there is any evidence of a relationship between religious affiliation and opinion on a proposed bill to ban smoking in public restaurants in a city. A simple random sample of residents in the city were polled to obtain their religious affiliation and their opinion on the proposed bill.

<u>Religion</u>	<u>No Smoking Bill</u>		
	<i>Yes</i>	<i>Neutral</i>	<i>No</i>
<i>Protestant</i>	64	37	94
<i>Catholic</i>	68	33	109
<i>Jewish</i>	10	10	26
<i>Other</i>	10	19	20

The data are organized into the contingency table displayed on the right.

- (a) Circle the labels of each correct way to complete the statement below.
 Looking for evidence of a relationship between religious affiliation and opinion on the bill is the same as looking for evidence of a
- (i) relationship among the religious affiliations.
 - (ii) difference in the distribution of opinions about the bill among religious affiliations.
 - (iii) difference in the distribution of religious affiliation among the different opinions about the bill.
 - (iv) difference between opinion about the bill and religious affiliation.
- (b) Explain how the data for this hypothesis test is appropriate for a chi-square test concerning independence.

- (c) Complete the four steps of the hypothesis test below. You should find that $\chi^2_6 = 15.725$.

Step 1 H_0 :

H_1 :

$\alpha =$

Step 2

Step 3

Step 4

31-6 - *continued*

- (d) Construct an appropriate graphical display. Then, describe the relationship which appears to exist, if necessary; if this is not necessary, say why not.
- (e) Verify that the sample size is sufficiently large for the χ^2 statistic to be appropriate.
- (f) Considering the results of the hypothesis test, decide which of the Type I or Type II errors is possible, and describe this error.
- (g) Decide whether H_0 would have been rejected or would not have been rejected with each of the following significance levels: (i) $\alpha = 0.01$, (ii) $\alpha = 0.10$.

31-7 Knowledge about good nutrition habits is to be compared among three different high schools: Central high school, Grandview high school, and Sandler high school. A 30-item questionnaire is administered to randomly selected students at each school, and the number of correct responses for each student is displayed in the table on the right. A 0.10 significance level is chosen for a hypothesis test to see if there is any evidence of a difference in mean score among the high schools.

Central:	16	21	10	12	
Grandview:	23	10	15		
Sandler:	23	28	18	29	26

- (a) Explain how the data for this hypothesis test is appropriate for a one-way ANOVA.
- (b) Complete the four steps of the hypothesis test below. As part of the second step, complete the construction of the ANOVA table below, where you should find that $SSB = 266.70$, $SSE = 235.55$, and Fisher's f statistic is $f_{2,9} = 5.10$.

Step 1 H_0 :
 H_1 :
 $\alpha =$

Step 2

Step 3

Step 4

<i>Source</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>f</i>	<i>P-value</i>
Error					
Total					

31-7 - continued

- (c) If multiple comparison is necessary, apply Scheffe's method and state the results; if multiple comparison is not necessary, explain why not.
- (d) In the list below, circle the best graphical display for this data and say why.
(i) multiple pie charts (ii) scatter plot (iii) multiple box plots
- (e) Considering the results of the hypothesis test, decide which of the Type I or Type II errors is possible, and describe this error.
- (f) Decide whether H_0 would have been rejected or would not have been rejected with each of the following significance levels: (i) $\alpha = 0.01$, (ii) $\alpha = 0.05$.
- (g) What would the presence of one or more outliers in the data suggest about using the f statistic?

31-8 Three chemical sprays, named Buzzoff, Nopest, and Flynot are designed to kill flies, and are being compared. A 0.10 significance level is chosen for a hypothesis test to see if there is any evidence of a difference in mean number of flies killed. Each spray is applied five times to 100 randomly selected live flies, and the number of flies killed is recorded with results displayed in the table on the right.

Buzzoff:	62	72	65	75	67
Nopest:	53	64	49	47	62
Flynot:	74	58	64	51	61

- (a) Explain how the data for this hypothesis test is appropriate for a one-way ANOVA.
- (b) Complete the four steps of the hypothesis test below. As part of the second step, complete the construction of the ANOVA table below, where you should find that $SSB = 435.6$, $SSE = 630.0$, and Fisher's f statistic is $f_{2,12} = 4.15$.

Step 1 H_0 :
 H_1 :
 $\alpha =$

Step 2

Step 3

Step 4

<i>Source</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>f</i>	<i>P-value</i>
Error					
Total					

31-8 - *continued*

- (c) If multiple comparison is necessary, apply Scheffe's method and state the results; if multiple comparison is not necessary, explain why not.
- (d) In the list below, circle the best graphical display for this data and say why.
(i) multiple pie charts (ii) scatter plot (iii) multiple box plots
- (e) Considering the results of the hypothesis test, decide which of the Type I or Type II errors is possible, and describe this error.
- (f) Decide whether H_0 would have been rejected or would not have been rejected with each of the following significance levels: (i) $\alpha = 0.01$, (ii) $\alpha = 0.05$.
- (g) What would the presence of one or more outliers in the data suggest about using the f statistic?