

Name: _____

STATISTICS
PART7 PRACTICE TEST
Time - 1 hour and 30 minutes
Number of multiple choice questions - 20
Number of free response questions - 3

1. An investigator was studying a territorial species of Central American termites, *Nasutitermes corniger*. Forty-nine termite pairs were randomly selected; both members of each of these pairs were from the same colony. Fifty-five additional termite pairs were randomly selected; the two members in each of these pairs were from different colonies. The pairs were placed in petri-dishes and observed to see whether they exhibited aggressive behavior. The results are shown in the table below.

	Aggressive	Nonaggressive	Total
Same colony	40 (33.5)	9 (15.5)	49
Different colonies	31 (37.5)	24 (17.5)	55
Total	71	33	104

A Chi-squared test for homogeneity was conducted, resulting in $\chi^2 = 7.638$. The expected counts are shown in parentheses in the table. Which of the following sets of statements follows from these results?

- χ^2 is significant, $p < 0.01$; the counts in the table suggest that termite pairs from the same colony are less likely to be aggressive than termite pairs from difference colonies.
- χ^2 is not significant at the 0.05 level.
- χ^2 is significant, $0.01 < p < 0.05$; the counts in the table suggest that termite pairs from different colonies are less likely to be aggressive than termite pairs from the same colony.
- χ^2 is significant, $p < 0.01$; the counts in the table suggest that termite pairs from different colonies are less likely to be aggressive than termite pairs from the same colony.
- χ^2 is significant, $0.01 < p < 0.05$; the counts in the table suggest that termite pairs from the same colony are less likely to be aggressive than termite pairs from different colonies.

2. According to the U.S. Burous of the Census, in 2000 the breakdown of the U.S. population (in %) by age is as follows.

Age	<5	5-14	15-24	25-34	35-44	45-54	55-64	65-74	75-84	>85
Percent	6.8	14.6	13.9	14.2	16	13.4	8.6	6.5	4.4	1.5

In order to determine if the distribution of the ages in a city are comparable to those of the country, a random sample of 200 ages are found. What test should be conducted?

- χ^2 goodness-of-fit test
- Inference for regression t -test
- Two-sample t -test
- Matched-pairs t -test
- One-sample t -test

3. The following table gives the percentage of nursing home residents by age.

Age	under 65	65-74	75-84	85 and older
Percentage	9.8	12	31.8	46.5

In a random sample of 150 families with a relative in a nursing home, the following was the distribution of the relatives' ages: 10 under 65, 20 ages 65-74, 52 ages 75-85, and 68 ages 85 and older. If a goodness-of-fit test were performed, what would be the value of the χ^2 statistic?

- a. 0.541
- b. 2.810
- c. 28.110
- d. 2.156
- e. 0.707

4. Two measures, x and y , are taken on numerous subjects, and a least squares regression equation is computed. The resulting equation is $\hat{y} = 358.52 - 15.49x$. A correct interpretation of the slope and intercept is

- a. for every increase of 100 units in x , y increases approximately 35852 units; when $x = 0$, y is predicted to be 15.49.
- b. for every increase of 100 units in x , y decreases approximately 1549 units; when $x = 0$, y is predicted to be 358.52.
- c. for every increase of 100 units in x , y increases approximately 1549 units; when $x = 0$, y is predicted to be 358.52.
- d. for every increase of 100 units in x , y decreases approximately 35852 units; when $x = 0$, y is predicted to be 15.49.
- e. for every increase of 100 units in x , y increases approximately 35852 units; when $x = 0$, y is predicted to be -15.49 .

5. A random sample of students was surveyed and each was asked to provide their grade level and in which location they eat lunch each day. The results are summarized in the table below.

	Large Cafeteria	Small Cafeteria	Student Center	Total
Freshman	22	87	16	125
Sophomore	30	75	25	130
Junior	75	10	45	130
Senior	110	0	36	146
Total	237	172	122	523

If a chi-squared test is performed to determine if grade level is independent of eating location, which of the following is the expected cell count for juniors that eat in the student center?

- $\frac{130 \times 172}{523}$
- $\frac{75 \times 122}{523}$
- $\frac{130 \times 75}{523}$
- $\frac{130 \times 122}{523}$
- $\frac{130 \times 172}{122}$

6. In a study of the performance of a computer printer, the size (in kilobits) 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 regression 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	0.294	11.8	≤ 0.0001
R squared = 87.5% R squared (adjusted) = 86.9%				
s = 0.6174 with 22 - 2 = 20 degrees of freedom				

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

- $3.47812 \pm 1.725 \times 0.294$
- $3.47812 \pm 2.086 \times 0.3153$
- $3.47812 \pm 1.96 \times 0.6174$
- $3.47812 \pm 2.086 \times 0.294$
- $3.47812 \pm 1.725 \times 0.3153$

7. There are 39 known moons orbiting the planet Jupiter. The diameter (in kilometers) and distance from Jupiter (in thousands of kilometers) are recorded for the moons discovered prior to May 2002. The conditions for regression analysis were met, and the results follow.

Dependent variable is **Distance (in 1000 km)**

No Selector

R squared = 29.6% R squared (adjusted) = 26.9%

s = 8358 with 28 - 2 = 26 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	763.35e6	1	763.35e6	10.9
Residual	1.81616e9	26	69.8524e6	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	15053.1	1714	9.37	≤ 0.0001
Diameter	-3.48571	1.054	-3.31	0.0028

Is there evidence of a relationship between diameter and distance

- There is evidence of a relationship between diameter and distance at the 0.01 level but not at the 0.001 level.
- There is evidence of a relationship between diameter and distance at the 0.05 level but not at the 0.01 level.
- There is no evidence of a relationship between diameter and distance at any of the commonly accepted levels.
- There is evidence of a relationship between diameter and distance at the 0.10 level but not at the 0.05 level.
- There is evidence of a relationship between diameter and distance at the 0.001 level.

8. A χ^2 -distribution with 14 degrees of freedom is a correct model for

- testing the question of whether 14 genetic traits are equally distributed in a population.
- a comparison of the production percentage distribution of 7 car model colors with the statistically determined national preference for those colors.
- testing whether the choice to smoke cigarettes or not to smoke cigarettes is independent of ethnicity among 7 different ethnic groups.
- testing whether choice of color is independent of age among 3 age groups and 5 color choices.
- a comparison of the equality of proportions of 8 sports activities for 3 high school grade levels.

9. A random sample of 27 individuals is selected, and the age and income of each individual is recorded. Regression analysis is performed, with the following results.

Dependent variable is: **Income**

No Selector

R squared = 2.0% R squared (adjusted) = -1.9%

s = 2.734e4 with 27 - 2 = 25 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	3.90728e8	1	3.90728e8	0.523
Residual	1.86828e10	25	7.4731e8	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	27300.4	1.576e4	1.73	0.0956
Age	244.203	337.7	0.723	0.4763

Is there a significant relationship between age and income?

- A significant relationship exists between age and income at the $\alpha = 0.05$ level but not at the $\alpha = 0.01$ level.
- A significant relationship between age and income does not exist at any of the commonly accepted levels.
- A significant relationship exists between age and income at the $\alpha = 0.001$ level.
- A significant relationship exists between age and income at the $\alpha = 0.10$ level but not at the $\alpha = 0.05$ level.
- A significant relationship exists between age and income at the $\alpha = 0.01$ level but not at the $\alpha = 0.001$ level.

10. Boiling and melting points (in degrees Celsius) are recorded for selected substances, and regression analysis is used to describe the relationship between them. The results of the analysis are shown below:

Dependent variable is: **Boiling Point**

No Selector

R squared = 73.4% R squared (adjusted) = 72.0%

s = 626.4 with 21 - 2 = 19 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	20.5469e6	1	20.5469e6	52.4
Residual	7.45573e6	19	392407	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	309.914	146.7	2.11	0.0481
Melting Point	0.959388	0.1326	7.24	≤ 0.0001

Assuming all of the conditions for regression have been met, which of the following gives the 95% confidence interval for the slope of the regression line?

- $0.959388 \pm 1.729(0.1326)$
- $309.914 \pm 2.093(626.4)$
- $0.959388 \pm 1.96(0.1326)$
- $0.959388 \pm 2.093(0.1326)$
- $309.914 \pm 1.729(146.7)$

11. Two hundred students were classified by gender and hostility level (low, medium, high), as measured by a HLT-test. The result were the following:

	Hostility Level		
	Low	Medium	High
Male	38	44	8
Female	59	46	5

- 75
 - 49
 - 40
 - 25
 - 60
- If the hostility level among students were independent of their gender, then how many female students would we expect to show the medium HLT score?

12. Regression analysis is performed on two variables, X and Y . A residual plot is graphed, and summary statistics are found for the residuals.

Dependent variable is: Y

No Selector

R squared = 83.0% R squared (adjusted) = 80.5%

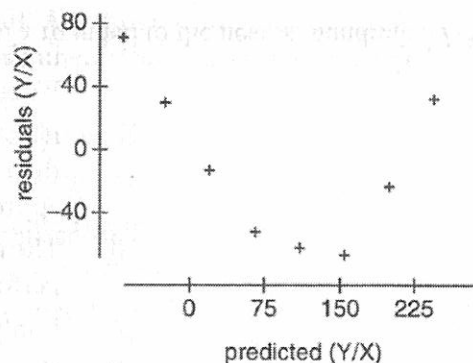
s = 59.67 with 9 - 2 = 7 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	121320	1	121320	34.1
Residual	24921.5	7	3560.21	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	-113.056	43.35	-2.61	0.0350
X	44.9667	7.703	5.84	0.0006

Summary of residuals(Y/X)
No Selector

Count	9
Mean	-42.6326e-15
Median	-11.8444
MidRange	5.80556
StdDev	55.8139
Range	145.1
IntQRRange	97.3167



What tells you there is a better fit for the data?

- The linear regression model for X and Y is appropriate; a better fit does not exist.
- There is a low correlation between X and Y .
- The mean residual value is 0.
- There is a pattern in the residual plot.
- A significant relationship does not exist for X and Y .

13. Which of the following statements is (are) true about the χ^2 -distribution with k degrees of freedom?
- The χ^2 -distribution is left-skewed.
 - The χ^2 -distribution with k degrees of freedom is more skewed than the χ^2 -distribution with $k + 1$ degrees of freedom.
 - The mean of the χ^2 -distribution with k degrees of freedom equals k .
- I only
 - I and II
 - III only
 - II only
 - II and III

14. In a study of how fast students can write their names, 29 students recorded the length of each of their first names (in letters) and the time (in seconds) that it took them to write 20 repetitions of their first name. A line satisfactorily represents the relationship between name length and writing time. The results of the regression analysis are given below.

Regression Analysis

The regression equation is

$$\text{Time} = 29.3 + 5.42 \text{ NameLength}$$

Predictor	Coef	Stdev	t-ratio	p
Constant	29.284	4.273	6.85	0.000
NameLength	5.4211	0.7573	7.16	0.000
s = 5.219		R-sq = 64.7%		R-sq(adj) = 63.4%

Which of the following should be used to compute a 95% confidence interval for the slope of the least-squares regression line?

- $5.4211 \pm 2.045 \times 0.7573$
- $5.4211 \pm 2.052 \times 0.7573$
- $5.4211 \pm 7.160 \times 5.219$
- $29.284 \pm 2.052 \times 4.273$
- $29.284 \pm 1.703 \times 4.273$

15. High school students on a closed campus recently petitioned their school board to allow students to leave campus for lunch. In order to support their opinion, the students randomly polled students and teachers with the following question: "Do you think that students should be allowed to leave campus for lunch?" The results are as follows:

	Agree	Disagree	No Opinion	Total
Students	123	36	6	165
Teachers	13	5	5	23
Total	136	41	11	188

Which of the following best describes the responses of students and teachers on the issue?

- There is evidence that students and teachers have different opinions on the issue at the 0.10 level, but not at the 0.05 level.
- There is evidence that students and teachers have different opinions on the issue at the 0.05 level but not at the 0.01 level.
- The conditions for conducting an appropriate test of homogeneity between students and teachers on the issue have not been met.
- There is evidence that students and teachers have different opinions on the issue at the 0.01 level but not at the 0.001 level.
- There is insufficient evidence that students and teachers have different opinions on the issue.

16. A test of independence is to be performed on a 3 x 4 contingency table. Which of the following assumptions needs to be satisfied for the *chi-square* test to be valid?

- All observed counts are at least 5.
- The number of categories is at least 5.
- The total number of observations is at least 5.
- All expected counts are at least 5.
- The *chi-square* test statistics is at least 5.

17. The Stroop Effect is a psychological phenomenon where naming the physical color of a word, the word itself can “interfere” with the process of naming the color of the word. For example, the word “red” is colored blue and a person asked to describe the color of the word will often say “red.” An experiment was done with 10 people comparing the time (in seconds) taken to say colors of words that were colored differently than the names of the words, and the time taken to say the colors of colored rectangles.

Person	Rectangles	Words
1	5	6
2	3	5
3	3	7
4	4	6
5	6	11
6	5	9
7	3	6
8	4	5
9	5	12
10	6	9

What is the number of degrees of freedom that would be used for the significance test to see if there is a difference in the mean time it takes to read the cards?

- a. 16
- b. 20
- c. 9
- d. 10
- e. 8

18. According to the U.S. Bureau of the Census, the distribution of ages for senior citizens is as follows:

Age	65-74	75-84	85-94	95 and over
Percentage	58.0	32.2	9.0	0.8

Of a random sample of 628 senior residents from a large community, 301 are ages 65-74, 232 are ages 75-84, 93 are ages 85-94, and 2 are over 95. Is there evidence to show that the distribution of residents in the community is different from the distribution of senior citizens in the nation?

- a. There is evidence of a significant difference in distributions at the 0.001 level.
- b. There is insufficient evidence of a significant difference in distributions at any of the commonly accepted levels.
- c. There is evidence of a significant difference in distributions at the 0.01 level but not at the 0.001 level.
- d. There is evidence of a significant difference in distributions at the 0.05 level but not at the 0.01 level.
- e. There is evidence of a significant difference in distributions at the 0.10 level but not at the 0.05 level.

19. The following information resulted from regression analysis for the percentage of white children under the age of 18 versus the percentage of black children under the age of 18 who live in poverty in several large cities.

Dependent variable is: **Black**

No Selector

R squared = 79.5% R squared (adjusted) = 77.8%

s = 1.072 with 14 - 2 = 12 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	53.5882	1	53.5882	46.6
Residual	13.7889	12	1.14907	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	30.4954	1.942	15.7	≤ 0.0001
White	0.911941	0.1335	6.83	≤ 0.0001

If the data were reevaluated using the percentage of white children as the dependent variable, the correlation coefficient would be which of the following?

- 0.892
- Not enough information is given to calculate the correlation coefficient.
- 0.778
- 0.795
- 0.882

20. A test engineer wants to estimate the mean gas mileage μ (in miles per gallon) for a particular model of automobile. Eleven of these cars are subjected to a road test, and the gas mileage is computed for each car.

A dotplot of the 11 gas-mileage values is roughly symmetrical and has no outliers. The mean and standard deviations of these values are 23.7 and 4.8, respectively. Assuming that these 11 automobiles can be a simple random sample of cars of this model, which of the following is a correct statement?

- A 95% confidence interval for μ is $23.7 \pm 2.228 \times \frac{4.8}{\sqrt{10}}$
- A 95% confidence interval for μ is $23.7 \pm 2.201 \times \frac{4.8}{\sqrt{11}}$
- A 95% confidence interval for μ is $23.7 \pm 2.201 \times \frac{4.8}{\sqrt{10}}$
- A 95% confidence interval for μ is $23.7 \pm 2.228 \times \frac{4.8}{\sqrt{11}}$
- The results cannot be trusted; the sample is too small.

FREE RESPONSE**Questions 1-3**

Spend about 45 minutes on this part of the exam.

2003 AP[®] STATISTICS FREE-RESPONSE QUESTIONS

5. A random sample of 200 students was selected from a large college in the United States. Each selected student was asked to give his or her opinion about the following statement.

“The most important quality of a person who aspires to be the President of the United States is a knowledge of foreign affairs.”

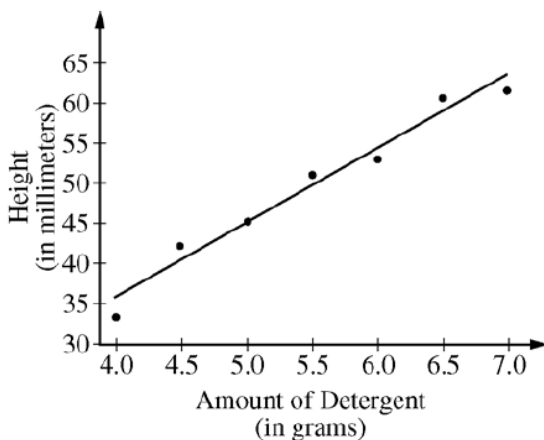
Each response was recorded in one of five categories. The gender of each selected student was noted. The data are summarized in the table below.

	Response Category				
	Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
Male	10	15	15	25	25
Female	20	25	25	25	15

Is there sufficient evidence to indicate that the response is dependent on gender? Provide statistical evidence to support your conclusion.

2. A manufacturer of dish detergent believes the height of soapsuds in the dishpan depends on the amount of detergent used. A study of the suds' heights for a new dish detergent was conducted. Seven pans of water were prepared. All pans were of the same size and type and contained the same amount of water. The temperature of the water was the same for each pan. An amount of dish detergent was assigned at random to each pan, and that amount of detergent was added to the pan. Then the water in the dishpan was agitated for a set amount of time, and the height of the resulting suds was measured.

A plot of the data and the computer output from fitting a least squares regression line to the data are shown below.



Predictor	Coef	SE Coef	T	P
Constant	-2.679	4.222	-0.63	0.554
Amount	9.5000	0.7553	12.58	0.000

S = 1.99821 R-Sq = 96.9% R-Sq(adj) = 96.3%

- (a) Write the equation of the fitted regression line. Define any variables used in this equation.
- (b) Note that $s = 1.99821$ in the computer output. Interpret this value in the context of this study.
- (c) Identify and interpret the standard error of the slope.

2007 Form A Question 6

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.

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

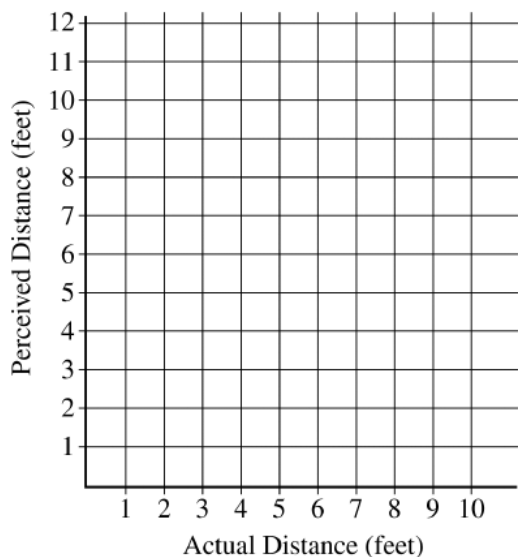
(c) Using Model 2, test the researcher's hypothesis that in dim light participants overestimate the distance, with the overestimate increasing as the actual distance increases. (Assume appropriate conditions for inference are met.)

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.

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

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.



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