Name: Key (2015)

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*. Fourty-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

expected counts

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?

P = .0057

9 < 15.5 ✗ a. 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.

✗ b. χ^2 is not significant at the 0.05 level.

P is lower than $\alpha = .01$ ✗ c. 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.

d. 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.

✗ e. 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?

a. χ^2 goodness-of-fit test
b. Inference for regression *t*-test
c. Two-sample *t*-test
d. Matched-pairs *t*-test
e. 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
Expected	14.7	18	47.7	69.75

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

$X^2 = 2.156$

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

$15.49 \times 100 = 1549$

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?

a. $\frac{130 \times 172}{523}$

b. $\frac{75 \times 122}{523}$

c. $\frac{130 \times 75}{523}$

d. $\frac{130 \times 122}{523}$

e. $\frac{130 \times 172}{122}$

$\frac{130 \times 122}{523}$

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?

a. $3.47812 \pm 1.725 \times 0.294$

b. $3.47812 \pm 2.086 \times 0.3153$

c. $3.47812 \pm 1.96 \times 0.6174$

d. $3.47812 \pm 2.086 \times 0.294$

e. $3.47812 \pm 1.725 \times 0.3153$

$b_1 \pm t^*_{20}\ s.e.(b_1)$

$3.47812 \pm 2.08596\,(0.294)$

invT (.975, 20)

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	16053.1	1714	9.37	≤ 0.0001
Diameter	-3.48571	1.054	-3.31	0.0028

Is there evidence of a relationship between diameter and distance (could the slope be zero)

a. There is evidence of a relationship between diameter and distance at the 0.01 level but not at the 0.001 level. $.001 < .0028 < .01$
b. There is evidence of a relationship between diameter and distance at the 0.05 level but not at the 0.01 level.
c. There is no evidence of a relationship between diameter and distance at any of the commonly accepted levels.
d. There is evidence of a relationship between diameter and distance at the 0.10 level but not at the 0.05 level.
e. 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

a. testing the question of whether 14 genetic traits are equally distributed in a population. no $n-1 = 13$.
b. a comparison of the production percentage distribution of 7 car model colors with the statistically determined national preference for those colors. no $n-1 = 6$
c. testing whether the choice to smoke cigarettes or not to smoke cigarettes is independent of ethnicity among 7 different ethnic groups.
d. testing whether choice of color is independent of age among 3 age groups and 5 color choices. $(3-1)(5-1) = 8$
e. a comparison of the equality of proportions of 8 sports activities for 3 hig hschool grade levels. $(8-1)(3-1) = 14$

____ 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	1.575e4	1.73	0.0956
Age	244.283	337.7	0.723	0.4763

Is there a significant relationship between age and income? (could the slope be zero?)

a. A significant relationship exists between age and income at the α =0.05 level but not at the α =0.01 level.

b. A significant relationship between age and income does not exist and any of the comonly accepted levels. .4763 > .1 (the highest comonly used α level)

c. A significant relationship exists between age and income at the α =0.001 level.

d. A significant relationship exists between age and income at the α =0.10 level but not at the α =0.05 level.

e. A significant relationship exists between age and income at the α =0.01 level but not at the α =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?

a. 0.959388 ± ~~1.729~~(0.1326)

b. ~~309.914~~ ± 2.093(626.4)

c. 0.959388 ± ~~1.96~~(0.1326)

d. 0.959388 ± 2.093(0.1326)

e. ~~309.914~~ ± 1.729(146.7)

$b_1 \pm t^*_{19} \, se(b_1)$

$0.959388 \pm 2.093\,(0.1326)$

invT (.975, 19)

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

110

90

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?

a. 75
b. 49
c. 40
d. 25
e. 60

$$\frac{110 \times 90}{200} = 49.5$$

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

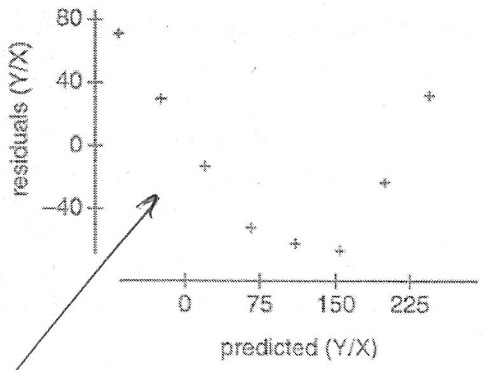
Dependent variable is: Y
No Selector
R squared = 83.0% R squared (adjusted) = 80.5%
s = 59.67 with 9 - 2 = 7 degrees of freedom

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	-113.855	43.35	-2.61	0.0350
X	44.9667	7.703	5.84	0.0005

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
IntQRange	97.3157



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

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

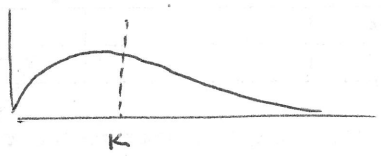
It should be a random scattering.

____ 13. Which of the following statements is (are) true about the χ^2-distribution with k degrees of freedom?

I. The χ^2-distribution is left-skewed. (right skewed)

II. The χ^2-distribution with k degrees of freedom is more skewed than the χ^2-distribution with $k+1$ degrees of freedom. as the df increases, it becomes less skewed

III. The mean of the χ^2-distribution with k degrees of freedom equals k. This is true



a. I only
b. I and II
c. III only
d. II only
e. 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

Time = 29.3 + 5.42 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?

a. 5.4211 ± ~~3.045~~ × 0.7573
b. 5.4211 ± 2.052 × 0.7573
c. 5.4211 ± ~~2.160~~ × 5.219
d. ~~29.284~~ ± 2.052 × 4.273
e. ~~29.284~~ ± ~~1.703~~ × 4.273

$b_1 \pm t^*_{n-2} \, se(b_1)$

5.4211 ± 2.052

invT(.975, 27)

____ 15. High school students on a closed campus recently retitioned 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?

a. 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.
b. 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.
c. The conditions for conducting an appropriate test of homogeneity between students and teachers on the issue have not been met.
d. 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.
e. There is insufficient evidence that students and teachers have different opinions on the issue.

Although the p-value is .00223, the expected counts in each cell were not at least 5. So

____ 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-sqare* test to be valid?

a. All observed counts are at least 5.
b. The number of categories is at least 5.
c. The total number of observations is at least 5.
d. All expected counts are at least 5.
e. 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

not data

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 if takes to read the cards?

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

$(10-1)(2-1)$
$9 \times 1 = 9$

____ 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
Expected	364.24	202.22	56.52	5.024

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?

This is more than 5, so the conditions are satisfied

a. There is evidence of a significant differece 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 differece in distributions at the 0.01 level but not at the 0.001 level.
d. There is evidence of a significant differece in distributions at the 0.05 level but not at the 0.01 level.
e. There is evidence of a significant differece in distributions at the 0.10 level but not at the 0.05 level.

$p = 7.45 \times 10^{-9} < .001$

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

$r^2 = .795$

$r = \sqrt{.795}$

$= \pm .8916$

because $b_1 > 0$

$r = +.8916$

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?

a. 0.892
b. Not enough information is given to calculate the correlation coefficient.
c. 0.778
d. 0.795
e. 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.

review question

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

$\bar{x} \pm t^*_{n-1} \frac{se(\bar{x})}{\sqrt{n}}$

$23.7 \pm 2.228 \times \frac{4.8}{\sqrt{11}}$

↖ invT(.975, 10)