Review Questions for Topic 7—Inference When Variables Are Related

Multiple Choice

1. A χ^2 test of independence is appropriate when we want to test the hypothesis that
   (A) two quantitative variables measured on the same subjects are related.
   (B) two categorical variables measured on the same subjects are related.
   (C) a distribution of counts in one categorical variable matches the distribution predicted by a model.
   (D) the distribution of counts for two or more groups on the same categorical variable is the same.
   (E) the means of two distinct populations are statistically significantly different.

2. Below appears a portion of the table showing who survived the sinking of the *Titanic*, based on whether they were crew members or passengers booked in first-, second-, or third-class staterooms.

	Crew	**First**	**Second**	**Third**	Total
Alive	212				710
Dead				528	
Total		325	285		2201

The expected value for the number of first-class passengers who died on the *Titanic* is

(A) 104.8
(B) 162.5
(C) 220.2
(D) 285.5
(E) There is not enough information to determine the expected value.

3. The table shows the number of students absent for the past school year at Bell Baxter High School, broken down by the day of the week. The dean of students would like to know if student absenteeism is related to the day of the week.

Monday	Tuesday	Wednesday	Thursday	Friday
98	84	92	86	76

The p-value for this test is

(A) p-value < 0.01
(B) $0.01 < p$-value < 0.05
(C) $0.05 < p$-value < 0.10
(D) $0.10 < p$-value < 0.15
(E) p-value > 0.15

Regression analysis for average SAT Math scores versus per-student spending for 22 school districts in Montgomery County, Pennsylvania, is given in the table below. Use the table to answer questions 4–6.*

Dependent variable is **SAT-Math**
No Selector
R squared = 21.9% R squared (adjusted) = 18.0%
s = 27.99 with 22 - 2 = 20 degrees of freedom

Source	Sum of Squares	*df*	Mean Square	*F*-ratio
Regression	4401.65	1	4401.65	5.62
Residual	15666.9	20	783.347	

Variable	Coefficient	s.e. of Coeff	*t*-ratio	prob
Constant	446.990	37.66	11.9	≤ 0.0001
Per Student	7.33145e-3	0.0031	2.37	0.0279

4. The Pennsylvania Department of Education is interested in the relationship between per-student spending and average SAT* Math scores. A scatterplot of the data suggests average SAT* Math scores tend to increase as per-student spending in a school district increases. The Department of Education is interested in whether or not this increase is significant. Assuming the conditions for inference are satisfied, the p-value for the linear regression t-test is approximately

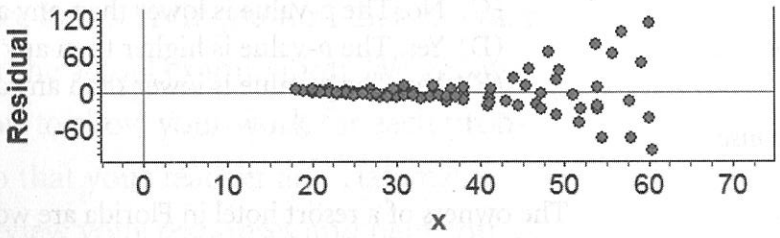
(A) 0.0001
(B) 0.0031
(C) 0.0140

(D) 0.0279
(E) 0.2799

5. Construct a 95% confidence interval for the slope of the regression line in question 4.
   (A) $0.00733 \pm 1.96(0.0031)$
   (B) $0.00733 \pm 2.086(0.0031)$
   (C) $0.00733 \pm 2.086\left(\frac{27.99}{\sqrt{22}}\right)$
   (D) $7.33145 \pm 1.96\left(\frac{27.99}{\sqrt{22}}\right)$
   (E) $7.33145 \pm 2.086(0.0031)$

6. Which of the following gives the appropriate hypotheses for the test of significance in question 4?
   (A) $H_0 : b = 0; H_A : b > 0$
   (B) $H_0 : \mu = 0; H_A : \mu > 0$
   (C) $H_0 : \alpha = 0; H_A : \alpha > 0$
   (D) $H_0 : \beta = 0; H_A : \beta > 0$
   (E) None of the above.

7. A least squares regression line for y versus x is calculated, and the corresponding residual graph is shown below.



   Which of the following is a correct statement?
   (A) The model is more accurate for smaller values of x than for larger values.
   (B) The condition of constant variance of the residuals is violated.
   (C) A linear model is appropriate.
   (D) Both A and B are correct.
   (E) Both A and C are correct.

Faculty and administrators in a school district were surveyed regarding proposed changes in the school's new health care program. A summary of their responses to the question "Are the proposed changes in the health care program acceptable?" is given in the following table. Use the table to answer questions 8–10.

	Agree	Neutral	Disagree	Total
Faculty	22	15	88	125
Administrators	22	10	1	33
Total	44	25	89	158

8. An appropriate hypothesis test to determine whether there is sufficient statistical evidence to conclude that the faculty and administrators differ in their opinions about the proposed new health care program is a
   (A) two-proportion z-test.
   (B) linear regression t-test.
   (C) χ^2 test of independence.
   (D) χ^2 test of goodness-of-fit.
   (E) χ^2 test of homogeneity (equal proportions).

9. The number of degrees of freedom for the appropriate test of significance in question 8 is
   (A) 2
   (B) 4
   (C) 12
   (D) 147
   (E) not applicable; the appropriate test does not require degrees of freedom.

10. Is there sufficient statistical evidence that the faculty and administrators differ in their opinions about the proposed new health care program?
   (A) No. The same number of faculty and administrators (22) agree that the proposed health care program is acceptable.
   (B) No. The p-value is higher than any acceptable α-level.
   (C) No. The p-value is lower than any acceptable α-level.
   (D) Yes. The p-value is higher than any acceptable α-level.
   (E) Yes. The p-value is lower than any acceptable α-level.

Free Response

The owners of a resort hotel in Florida are wondering whether their customers' preferred method of booking their stays during the peak season has changed significantly for the given time periods. Records for peak season rentals for 1995, 2000, and 2005 appear below.

	1995	2000	2005
Travel Agent	112	103	57
Book Online	42	77	123
Other	56	30	30

Is there evidence to suggest that customers' booking preferences for these time periods have changed?

Answers for Topic 7—Inference When Variables Are Related—begin on page 275.