Review χ^2 and inference for regression

1. A chi-square goodness-of-fit test is used to test whether a 0 to 9 is "fair" (that is, the outcomes are all equally likely). The spinner is spun 100 times, and the results are recorded. The degrees of freedom for the test will be
   a) 8 b) 9 c) 10 d) 99 e) None of these

Refer 2-3 to the following setting: Recent revenue shortfalls in a midwestern state led to a reduction in the state budget for higher education. To offset the reduction, the largest state university proposed a 25% tuition increase. It was determined that such an increase was needed simply to compensate for the lost support from the state. Separate random samples of 50 freshmen, 50 sophomores, 50 juniors, and 50 seniors from the university were asked whether or not they were strongly opposed to the increase, given that it was the minimum increase necessary to maintain the university's budget at current levels. The results are given in the following table.

Strongly	**Year**			
Opposed?	**Freshmen**	**Sophomores**	**Juniors**	**Seniors**
Yes	39	36	29	18
No	11	14	21	32

2. Which hypotheses would be appropriate for performing a chi-squared test?
   a) The null hypothesis is that the closer students get to graduation, the less likely they are opposed to tuition increases. The alternative is that how close students are to graduation makes no difference in their opinion.
   b) The null hypothesis is that the mean number of students who are strongly opposed is the same for each of the four years. The alternative is that the mean is different for at least two of the four years.
   c) The null hypothesis is that the distribution of student opinion about the proposed tuition increase is the same for each of the four years at this university. The alternative is that the distribution is different for at least two of the four years.
   d) The null is that year in school and student opinion about tuition increase in the sample is independent. The alternative is that these variables are dependent.
   e) The null hypothesis is that there is an association between year in school and opinion about the tuition increase at this university. The alternative hypothesis is that these variables are not associated.

3. The conditions for carrying out the chi-square test above are
   I. Separate random samples from the population of interest.
   II. Expected counts are large enough.
   III. The samples themselves and the individual observations in each sample are independent.
   Which of the conditions is (are) satisfied in this case?
   a) I only b) II only c) I and II only d) II and III only e) I, II, and III

Refer 4-6 to the following setting: A random sample of traffic tickets given to motorists in a large city is examined. The tickets are classified according to the race of the driver. The results are summarized in the following table:

Race:	White	Black	Hispanic	Other
# of tickets:	69	52	18	9

The proportion of this city's population in each of the racial categories listed above is as follows:

Race:	White	Black	Hispanic	Other
Proportion:	0.55	0.3	0.08	0.07

We wish to test H_o: The racial distribution of traffic tickets in the city is the same as the racial distribution of the city's population.

4. Assuming H_o is true, the expected number of Hispanic drivers who would receive a ticket is
   a) 8 b) 10.36 c) 11 d) 11.84 e) 12

5. We compute the value of the chi-square statistic to be 6.58. Assuming that the conditions for inference are met, the p-value of our test is

a) greater than 0.20 b) between 0.10 and 0.20 c) between 0.05 and 0.10
d) between 0.01 and 0.05 e) less than 0.01

6. The category that contributes the largest component to the chi-square statistic is

a) White b) Black c) Hispanic d) Other e) The answer cannot be determined

Refer 7-10 to the following setting: All current-carrying wires produce electromagnetic (EM) radiation, including the electrical wiring running into, through, and out of our homes. High-frequency EM radiation is thought to be a cause of cancer. The lower frequencies associated with household current are generally assumed to be harmless. To investigate this, researchers visited the addresses of a random sample of children who had died of some form of cancer (leukemia, lymphoma, or some other type) and classified the wiring configuration outside the dwelling as either a high-current configuration (HCC) or a low-current configuration (LCC). Here are the data:

	Leukemia	**Lymphoma**	**Other cancers**
HCC	52	10	17
LCC	84	21	31

Computer software was used to analyze the data. The output is given below. It includes the cell counts, some of the expected counts, and the value of the χ^2 statistic. In the table, expected counts are printed below observed counts and enclosed within parentheses.

	Leukemia	**Lymphoma**	**Other cancers**	**Total**
HCC	52	10	17	79
	(49.97)		(17.64)	
LCC	84	21	31	136
	(86.03)		(30.36)	
Total	136	31	48	215

χ^2 = 0.082 + 0.170 + 0.023 + 0.048 + 0.099 + 0.013 = 0.435

7. The appropriate degrees of freedom for the χ^2 statistic is

a) 1 b) 2 c) 3 d) 4 e) 5

8. The expected count of cases with lymphoma in homes with an HCC is

a) $\frac{79 \cdot 31}{215}$ b) $\frac{10 \cdot 31}{215}$ c) $\frac{79 \cdot 31}{10}$ d) $\frac{136 \cdot 31}{215}$ e) None of these.

9. Which of the following may we conclude, based on the results?

a) There is strong evidence of an association between wiring configuration and the chance that a child will develop some form of cancer.
b) HCC either causes cancer directly or is a major contributing factor to the development of cancer in children.
c) Leukemia is the most common type of cancer among children.
d) There is not much evidence of an association between wiring configuration and the type of cancer that caused the deaths of children in this study.
e) There is weak evidence that HCC causes cancer in children.

10. A Type I error would occur if we conclude that

a) HCC wiring caused cancer when it actually did not.
b) HCC wiring did not cause cancer when it actually did.
c) there is no association between type of wiring and the form of cancer when there actually is an association.
d) there is an association between type of wiring and the form of cancer when there actually is no association.
e) the type of wiring and the form of cancer have a positive correlation when they actually do not.

FREE RESPONSE

11. A large distributor of gasoline claims that 60% of all cars stopping at their service stations choose regular unleaded gas and that premium and supreme are each selected 20% of the time. To investigate this claim, researchers collected data from a random sample of drivers who put gas in their vehicles at the distributor's service stations in a large city. The results were as follows:

Regular	Premium	Supreme
261	51	88

Is there evidence that the distributor's claim is correct? Support your answer with statistical evidence.

12. A study conducted in Charlotte, North Carolina, tested the effectiveness of three police responses to spouse abuse: (1) advise and possibly separate the couple, (2) issue a citation to the offender and, (3) arrest the offender. Police officers were trained to recognize eligible cases. When presented with an eligible case, a police officer called the dispatcher, who would randomly assign one of the three available treatments to be administered. There were a total of 650 cases in the study. Each case was classified according to whether the abuser was subsequently arrested within six months of the original incident.

Treatment	No subsequent arrest	Subsequent arrest
1. Advise and separate	187	25
2. Citation	181	43
3. Arrest	175	39

a) Explain the purpose of the random assignment in the design of the study.
b) Construct a well-labeled graph that is suitable for comparing the effectiveness of the three treatments.
c) Is there evidence that there is a difference in the proportions of spouse abusers that would be arrested within six months after receiving their respective treatments? Carry out an appropriate test.
d) Interpret the p-value in context.

Refer 13-18 to the following setting: To determine property taxes, Florida reappraises real estate every year, and the county appraiser's Web site lists the current "fair market value" of each piece of property. Property usually sells for somewhat more than the appraised market value. Data was collected on the appraised market values x and actual selling prices y (in thousands of dollars) of a random sample of 16 condominium units. A scatterplot of the data shows a fairly strong positive linear relationship between the variables. Here is part of the Minitab output from a least-squares regression analysis using these data.

Predictor	Coef	SE Coef	T	P
Constant	127.27	79.49	1.60	0.132
appraisal	1.0446	0.1126	9.29	0.000
S = 69.7299	R-Sq -= 86.1%	R-Sq (adj) = 85.1%		

13. The equation of the least-squares regression line for predicting selling price from appraised value is
a) Price = 79.49 + 0.1126(appraised value)
b) Price = 0.1126 + 1.0466(appraised value)
c) Price = 127.27 + 1.0466(appraised value)
d) Price = 1.0466 + 127.27(appraised value)
e) Price = 1.0466 + 69.7299(appraised value)

14. What is the correlation between selling price and appraised value?
a) 0.1126 b) -0.861 c) -0.928 d) 0.861 e) 0.928

15. The slope β of the population regression line describes
a) the exact increase in the selling price of an individual unit when its appraised value increases by $1000
b) the average increase in the appraised value in a population of units when selling price increases by $1000.
c) the average increase in selling price in a population of units when appraised value increases by $1000.
d) the average selling price in a population of units when a unit's appraised value is 0.
e) the average increase in appraised value in a sample of 16 units when selling price increases ny $1000.

16. Is there significant evidence that selling price increases as appraised value increases? To answer this question, test the hypotheses

a) H_o: $\beta = 0$ versus H_a: $\beta > 0$ b) H_o: $\beta = 0$ versus H_a: $\beta < 0$

c) H_o: $\beta = 0$ versus H_a: $\beta \neq 0$ d) H_o: $\beta > 0$ versus H_a: $\beta = 0$

e) H_o: $\beta = 1$ versus H_a: $\beta > 1$

17. Confidence intervals and tests for these data use the *t* distribution with degrees of freedom

a) 9.29 b) 14 c) 15 d) 16 e) 30

18. A 95% confidence interval for the population slope β is

a) 1.0466 ± 149.5706 b) 1.0466 ± 0.2415 c) 1.0466 ± 0.2387

d) 1.0466 ± 0.1983 e) 1.0466 ± 001126

FREE RESPONSE

Lamb's-quarter is a common weed that interferes with the growth of corn. An agriculture researcher planted corn at the same rate in 16 small plots of ground and then weeded the plots by hand to allow a fixed number of lamb's-quarter plants to grow in each meter of corn row. The decision of how many of these plants to leave in each plot was made at random. No other weeds were allowed to grow. Here are the yields of corn (bushels per acre) in each of the plots:

Weeds per meter	Corn yield	Weeds per meter	Corn yield
0	166.7	3	158.6
0	172.2	3	176.4
0	165.0	3	153.1
0	176.9	3	156.0
1	166.2	9	162.8
1	157.3	9	142.4
1	166.7	9	162.8
1	161.1	9	162.4

a) Draw a scatterplot of the data. Describe what the graph tells you about the relationship between these two variables.

Minitab output from a linear regression on these data is shown below.

Predictor	Coef	SE Coef	T	P
Constant	166.483	2.725	61.11	0.000
Weeds per meter	-1.0987	0.5712	****	****
S = 7.97665	R-Sq -= 20.9%	R-Sq (adj) = 15.3%		

b) What is the equation of the least-squares regression line for predicting corn yield from the number of lamb's-quarter plants per meter?

c) Interpret the slope and y-intercept in context.

d) Interpret S and SE_b.

e) Do these data provide convincing evidence that more weeds reduce corn yield? Support your answer with statistical evidence.