AP Statistics – Unit 08
Chi Square & Regression Inference

Name Key
Period________

Multiple Choice

1. When calculating the degrees of freedom when doing inference for linear regression, we use:
   a. DF = n – 1
   b. DF = n – 2
   c. DF = n + 1
   d. DF = n + 2
   e. DF = n – 4

B

2. Looking at an Minitab output, where can you find the slope of the regression line?
   a. Coef column, Constant row
   b. Coef column, Variable row
   c. SE Coef column, Constant row
   d. SE Coef column, Variable row
   e. T column, Constant row

B

3. Looking at an Minitab output, where can you find the y-intercept of the regression line?
   a. Coef column, Constant row
   b. Coef column, Variable row
   c. SE Coef column, Constant row
   d. SE Coef column, Variable row
   e. T column, Constant row

A

4. When using the acronym LINER to check conditions for regression inference, the letters stand for:
   a. L: Linear, I: Independent, N: Normal, E: Equal sample size, R: Random
   b. L: Linear, I: Independent, N: Normal, E: Equal SD, R: Random
   c. L: Linear, I: Individual, N: Normal, E: Equal sample size, R: Random
   d. L: Linear, I: Individual, N: Normal, E: Equal SD, R: Random
   e. None of the above

B

5. When analyzing survey results from a two-way table, the main distinction between a test for independence and a test for homogeneity is
   a. How the degrees of freedom are calculated
   b. How the expected counts are calculated
   c. The number of samples obtained
   d. The number of rows in the two-way table
   e. The number of columns in the two-way table

C

Free Response

GOF

6. Biologists wish to cross pairs of tobacco plants having genetic makeup Gg, indicating that each plant has one dominant gene (G) and one recessive gene (g) for color. Each offspring plant will receive one gene for color from each parent. The Punnett square below shows the possible combinations of gene received by the offspring:

		Parent 2 passes on:	
		G	**g**
Parent 1 passes on:	**G**	GG	Gg
	g	Gg	gg

The Punnett square suggests that the expected ratio of green (GG) to yellow-green (Gg) to albino (gg) tobacco plants should be 1:2:1. In other words, the biologists predict that 25% of the offspring will be green, 50% will be yellow-green, and 25% will be albino. To test their hypothesis about the distribution of offspring, the biologists mate 84 randomly selected pairs of yellow-green parent plants. Of 84 offspring, 23 plants were green, 50 were yellow-green, and 11 were albino. Do the data provide convincing evidence at the $\alpha = 0.01$ level that the true distribution of offspring is different from what the biologists predict?

State: H_0: The stated distribution of the offspring coloring (1:2:1) is correct.

$\alpha = 0.01$

H_A: The stated distribution of the offspring coloring is not correct.

color	observed	expected
green	23	21
yg	50	42
albino	11	21

Plan: random: 84 randomly selected pairs of plants ✓

10% condition: 840 < all yellow-green pairs of plants ✓

Large counts: all expected counts (in table above) are at least 5.

Because our conditions are met, we will perform a χ^2 test for Goodness of Fit.

DO: $$\chi^2 = \frac{(23-21)^2}{21} + \frac{(50-42)^2}{42} + \frac{(11-21)^2}{21} = 6.4762$$

df = 2

p-value = 0.03924

Conclude: Because our p-value = 0.03924 is greater than our significance level $\alpha = 0.01$, we fail to reject the null. There is not convincing evidence that the stated distribution of offspring coloring (1:2:1) is incorrect.

homog.

7. A study in Charlotte, NC, 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.

(police responses)

Subsequent arrest?	Advise & Separate	Treatment Citation	Arrest	Total
Yes	187	181	175	543
No	25	43	39	107
Total	212	224	214	650

a. Explain the purpose of the random assignment in the design of this study.

Random assignment was used to create three roughly equivalent groups at the beginning of the study.

(essentially 3 separate samples now)

b. State an appropriate pair of hypotheses for performing a chi-square test in this setting.

H_0: The true proportion of spouse abusers (like the ones in this study) who will be arrested within 6 months is the same for all 3 police responses.

H_A: The true proportion of spouse abusers who will be arrested within 6 months is not the same for all 3 police responses. (we use prop. instead of dist. because it's a yes/no).

c. Assume that all of the conditions for performing the test are met. The test yields $\chi^2 = 5.063$ and a P-value of 0.0796. Interpret this P-value in context. What conclusion should we draw from the study?

*p-value represents the probability of getting differences as large as or larger than the ones observed by chance alone (like our sample), assuming H_0 is true.

Because our p-value = 0.0796 is less than our significance level $\alpha = 0.05$, we fail to reject the null. There is not convincing evidence that the true proportion of spouse abusers (like the ones in this study) who will be arrested within 6 months is not the same across all 3 police responses.

indep.

8. In the United States, there is a strong relationship between education and smoking: well-educated people are less likely to smoke. Does a similar relationship hold in France? To find out, researchers recorded the level of education and smoking status of a random sample of 459 French men aged 20 to 60 years. The two-way table below displays the data.

	Education		
Smoking Status	**Primary School**	**Secondary School**	**University**
Nonsmoker	56	37	53
Former Smoker	54	43	28
Moderate Smoker	41	27	36
Heavy Smoker	36	32	16

Is there convincing evidence of an association between smoking status and educational level among French men aged 20 to 60 years?

State: H_0: There is no association between smoking status and educational level among French men aged 20 to 60 years.

$\alpha = 0.05$

H_A: There is an association between smoking status and educational level among French men aged 20 to 60 years.

Plan: random: random sample of 459 French men ✓

10% condition: 4590 < all French men aged 20-60 years. ✓

Large counts: all expected counts are ≥ 5 ✓

status	primary	secondary	university
nonsmoker	59.481	44.214	42.305
former smoker	50.926	37.854	36.22
moderate smoker	42.37	31.495	30.135
heavy smoker	34.222	25.438	24.34

Because our conditions are met, we will perform a χ^2 test for independence.

Do:

$$\chi^2 = \frac{(56-59.481)^2}{59.481} + \frac{(37-44.214)^2}{44.214} + \frac{(53-42.305)^2}{42.305} + \dots$$

test statistic = 13.305

df = 6

p-value = 0.03844

Conclude: Because our p-value = 0.03844 is less than our significance level, we reject the null. There is convincing evidence that there is an association between smoking status and educational level of French men aged 20-60 years.

9. The student in Mrs. De Marre's class measures the arm spans and heights (in inches) of a random sample of 18 students from their large high school. Some computer output from a least-squares regression analysis on these data is shown below. Construct and interpret a 90% confidence interval for the slope of the population regression line. Assume that the conditions for performing inference are met.

Predictor	Coef	Stdev	t-ratio	p
Constant	11.547	5.600	2.06	0.056
Armspan	0.84042	0.08091	10.39	0.000

S = 1.613 R-Sq = 87.1% R-Sq(adj) = 86.3%

State: we want to estimate the slope β of the population regression line relating heights to arm spans for students in this high school with 90% confidence.

Plan: Because our conditions are met, we will use a t-interval for slope β of a regression line.

DO:

$0.84042 \pm 1.746(0.08091) =$

$(0.69915, 0.98169)$

$df = 16$

conclude: we are 90% confident that the interval from 0.69915 to 0.98169 captures the slope β of the population regression line relating heights to arm spans for students in this high school.