AP STATISTICS **Name:** _____ **Block:** ____

Review Unit IX – Chi Squared

MULTIPLE CHOICE Write the letter corresponding to the best answer in the given blank.

_____ 1. A survey is conducted on a group of 90 students. The data is sorted into a 2-way table with 6 rows and 5 columns. If χ^2 is 36.1, which of the following intervals contains the *p*-value?

A) $p < 0.01$ B) $0.01 < p < 0.02$ C) $0.02 < p < 0.025$ D) $0.025 < p < 0.05$ E) $p > 0.05$

_____ 2. For the survey described in the previous problem, the totals for each category of response are summarized in the following table:

	Chocolate	**Vanilla**	**Strawberry**	**Totals**
Male				*50*
Female				*40*
Totals	*37*	*22*	*41*	*90*

The expected number of females who prefer vanilla flavor ice cream is:

A) $\frac{(50)(22)}{90}$ B) $\frac{(40)(22)}{90}$ C) $\frac{90}{(40)(22)}$ D) $\frac{40}{(22)(90)}$ E) $\frac{22}{(40)(90)}$

_____3. For the scenario in the previous two problems: To see if there is statistical evidence that gender is related to preference in ice cream flavor, which of the following is the correct null hypothesis?

A) the distribution of sample proportions for ice cream preference is the same for males and females.
B) the distribution of sample proportions for ice cream preference is not the same for males and females.
C) there is an association between gender and preference in ice cream flavor.
D) there is no association between gender and preference in ice cream flavor.
E) the distribution of proportions of ice cream preference matches the advertised relative distribution.

_____ 4. Pig dice are little plastic pig-shaped figures that can be tossed and rolled just like dice! One of the six positions that pig dice may land on is called "razorback", in which the pig lands on its back with all 4 legs in the air. Based on empirical data, the probability that a pig will land on this position is about 20%.

In a game of chance, you get to roll a pig dice a certain number of times, and if more than 50% of your rolls land "razorback", then you win a large prize! From the following three choices, you get the pick how many times to roll the pig dice:

n=10, n=20, and n=100,

Which value of n should the player choose in order to maximize the probability of winning a prize?

A) n=10 only B) n=20 only C) n=100 only D) all three probabilities are equal

_____ 5. The distribution of SAT Math scores of students taking Statistics at a very large university is *skewed right* with a mean of 625 and a standard deviation of 44.5. If random samples of 100 students are repeatedly taken, which give the sampling model, which statement best describes this sampling model of sample mean scores?

A) Approximately normal with a mean of 625 and standard deviation of 44.5.
B) Approximately normal with a mean of 625 and standard deviation of 4.45.
C) Skewed to the right with a mean of 625 and standard deviation of 44.5.
D) Skewed to the right with a mean of 625 and standard deviation of 4.45.
E) No conclusion can be drawn since the population is not normally distributed.

6. In today's age of increasing competitiveness, parents are looking for any edge that they can get in helping their children achieve success in school. One particular study focuses on whether parents are more likely to enroll their children in supplemental educational programs if they are told that the teaching methods used during the programs originated overseas (versus if they are told that the teaching methods were developed somewhere in the United States). A group of 153 parents volunteered for a study, in which all of the parents were looking to enroll their children in some sort of supplemental education program. All of the 153 parents were offered the same program for their children, however – unknown to the parents – 73 were randomly assigned to be told that the teaching methods for this program were developed in Finland, while the other 80 parents were told that the teaching methods for this program originated in Louisiana. Each of the parents was then given the choice of whether or not to enroll their children in this program. The results of this experiment are summarized in the following table.

		Told that teaching methods were developed in Finland	Told that teaching methods originated in Louisiana	Total
Enrolled children in this program	Yes	36	25	61
	No	37	55	92
	Total	73	80	153

The researchers for this study wish to conduct a significance test to determine whether there is a difference in the proportions of parents who would enroll their children in a supplemental educational program based on the purported location of origin (Finland versus Louisiana).

a) Suppose the researchers use the data to perform a chi-squared test of homogeneity. **State the hypotheses, and calculate the test statistic and *p*-value for the test** (you may assume that the conditions for inference have been checked and verified).

b) Now suppose the researchers wish to use the data to conduct a 2-sample *z*-test for a difference between two proportions (aka, a 2-proportion *z*-test). **State the appropriate hypotheses and calculate the test-statistic and p-value for the test.** *(Again, assume the conditions for inference have been verified)*

c) What do you notice about the *p*-values for the two tests?

d) Suppose the researchers now wish to use the data to test whether parents are more likely to enroll their children in a program if the teaching methods are purported to have originated in Finland than if they originated in Louisiana. Could this still be accomplished by using a chi-squared test of homogeneity? What about a 2-proportion *z*-test? Explain.

7. Young magic students at a large school of magic are sorted by a "sorting hat" into one of four houses, which serve as the learning and living communities for its students. A random sample of 86 magicians was taken from all students entering the school in the year 1971, and a separate random sample of 72 magicians was taken from all students entering the school in year 1991. For the students in these random samples, the distribution of the number of students observed in the four houses in each year is shown in the following table:

	House			
	G	H	R	S
1971	22	13	30	21
1991	26	21	16	9

At the 10 percent level of significance, is there evidence of a difference in the relative distribution of students among the four houses between 1971 and 1991?

8. The table below shows the political party registration by gender of 500 randomly selected registered voters in Franklin Township.

PARTY REGISTRATION—FRANKLIN TOWNSHIP

	Party W	Party X	Party Y
Female	60	120	120
Male	28	124	48

a) Among registered voters in Franklin Township, is there evidence of an association between gender and party registration? Give appropriate statistical evidence to support your conclusion.

b) Suppose the conclusion that you made in part (a) was incorrect. What type of error (type I or type II) would this be? Explain this error in context.

9. The department of parks and recreation of a certain city conducts summer programs for residents of its five districts. The summer programs include operating and maintaining community swimming pools in each of the districts as well as offering sports and recreational programs for school-age children, young adults, and older adults.

The proportion of total households in each of the districts was determined for this study, and the proportions are displayed in the middle column in the table below. For instance, 32% of the city's households are located in district A, 12% of the city's households are located in district B, and so on.

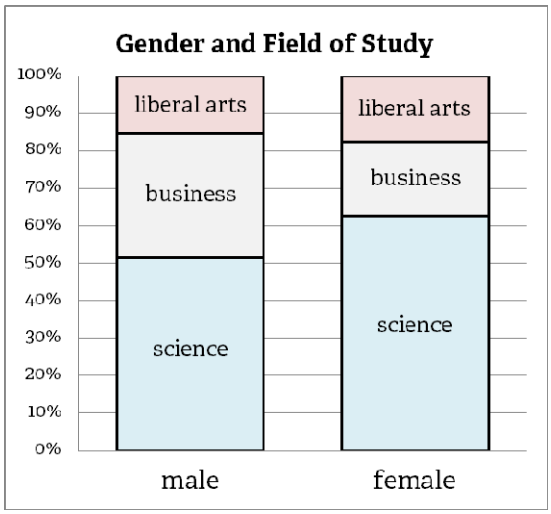
A group of statisticians collected data from a simple random sample this past summer, and the <u>number</u> of households that participated in summer programs from each district is shown in the right-hand column below.

District	Proportion of Total Households	Number of Households that Participated in Summer Programs
A	0.32	100
B	0.12	35
C	0.10	40
D	0.27	45
E	0.19	51
Total	1.000	271

a) The department of parks and recreation expect the number of households that participate in each district to be proportional to the percentage of the city's households that are located in each district. Are the data consistent with this expectation? Conduct an appropriate statistical test to support your conclusion. **Assume the conditions for inference are met.**

b) Relative to the proportion of total households, in which district is participation in summer programs the lowest? Explain.

10. Parker and Katrina wish to investigate whether there is a relationship between gender and intended field of study for the students who currently attend their high school. For their end-of-year project, they surveyed a random sample of 117 of their classmates, and the results are summarized in the segmented bar graph below.



a) For this group of people, does there appear to be an association between "gender" and "field of study"? Write a few sentences describing what this graph reveals about this association.

b) Suppose Parker and Katrina wish to perform a significance test to determine whether there is an association between gender and field of study. Name the appropriate inference procedure, along with the degrees of freedom (if appropriate).

c). Suppose that their hypothesis test yields a test statistic of $\chi^2 = 2.94$.
If they were testing at a significance level of $\alpha = 0.05$, what would they conclude for their hypothesis test?
(Note: For this problem, you must first use either your calculator or the χ^2 table to find a p-value...)

11. A group of psychologists from Podunk University (our favorite place!) conducted a study to investigate whether there is an association between being happy with your grades in school and eating breakfast on a regular basis for the students who attend the local Podunk High School.

The Podunkian psychologists used two independent random samples of students from Podunk High School: 34 who reported that they choose to eat breakfast on a daily basis, and 33 who reported that they choose not to eat breakfast regularly (if at all). 19 of the 34 that ate breakfast regularly stated that they were satisfied with their grades, while 10 of the 33 that did not eat breakfast regularly reported satisfaction with their school grades.

a) Suppose you wish to perform a test to see if these samples provide evidence that the proportion of students who are satisfied with their school grades is greater for those who eat breakfast regularly than for those who do not. **What type of test would be appropriate for this problem? Write the hypotheses for this test, then check the appropriate conditions for inference (but DO NOT perform the hypothesis test).**

b) Suppose that the probability of rejecting H0 when H0 is true is 0.10, and that the probability of failing to reject H0 when H0 is false is 0.38. Calculate the power of this test, and explain clearly the meaning of this value in context.

c) Suppose that the results of the hypothesis test that you performed in part (a) were statistically significant (in other words, that the proportion of all Podunk students that are satisfied with their grades is higher for those who eat breakfast regularly than for those who do not eat breakfast regularly). Based on this study, can we state that eating breakfast **causes** a greater likelihood that a Podunk High student will be satisfied with their grades? Clearly explain why or why not.

AP Statistics
Unit IX Review – Inference with Chi-Squared

****ANSWERS ONLY**** *(for explanations, please come in for tutorials)*

1. B
2. B
3. D
4. A
5. B

6. a) χ^2 =5.196, df = 1, p-value = 0.0226
   b) z = 2.279, p-value = 0.0226
   c) They are equal!
   d) Since this is now a 1-sided test, only a 2-proportion z-test can be used.

7. χ^2 test of homogeneity, df = (4 – 1)(2 – 1) = 3

   Ho: The distribution of proportions (or relative distribution) of students in the four houses are equal for 1971 and 1991.
   Ha: The distribution of proportions of students in the four houses are NOT equal for 1971 and 1991.

   χ^2 = 10.115, p-value = 0.0176

   Since p-value < a, we reject H0. We have evidence that at least one of the proportions for the four houses has changed between 1971 and 1991.

8. a) χ^2 test of independence (or association), df = 2
   Ho: There is NO association between gender and party registration
   Ha: There IS an association between gender and party registration

   χ^2 = 23.499, p-value = 7.89×10^{-6}

   b) Type I (since you should have rejected H0 in part a).
   There is no association between gender and party registration, but we concluded that there was.

9. a) χ^2 Goodness of Fit
   χ^2 = 19.21, p-value = 0.00071…
   b) District D. This is the district that had the largest positive difference between the observed and expected number of participants.

10. a) Yes… a greater proportion of females intend to study science compared to males (about 62% to 52%), while males are more likely to study business than females (about 32% to 20%). Similar proportions of males and females intend to study liberal arts (about 17-19%).
    b) χ^2 test of independence, df = 2 (be sure to show the calculation for the df)
    c) Table: p-value is between 0.20 and 0.25. Calculator: p-value = 0.2299.
    Since this p-value is > alpha, we fail to reject H0, and lack evidence of an association between…

11. a) 2-proportion z-test. Ho: $p_{\text{breakfast}} = p_{\text{non-breakfast}}$ Ha: $p_{\text{breakfast}} > p_{\text{non-breakfast}}$
    b) Power = $1 - \beta$ = 0.62. This is the probability of correctly rejecting H0 (be sure to put this in context!)
    c) No… *(check your unit 1 notes if you forgot what you need to show cause and effect…)*