Part 1: Multiple Choice. *Circle the letter corresponding to the best answer.*

Use the following for questions 1 – 3:

A well-known chewing gum maker wants to determine if any of its four flavors of gum are more popular than the others. A random sample of 80 people who say they chew gum regularly is asked to identify their favorite flavor of gum. Here are the results:

Flavor	Peppermint	Cinnamon	Wintergreen	Spearmint
Frequency	25	19	22	14

1. Which of the following would be an appropriate null hypothesis for the company to test?
    (a) $\mu_1 = \mu_2 = \mu_3 = \mu_4$

    (b) The observed counts are all equal to 20.
    (c) Flavor preferences for the population are evenly distributed across the four flavors.
    (d) At least one of the four flavor preferences in the population is different from the other three.
    (e) The observed counts are equal to the expected counts.

2. Which of the following are conditions that must be met in order to test this hypothesis using a chi-square test?
    I.   If p = proportion of gum-chewers in the population, then $np \geq 10$ and $n(1-p) \geq 10$.
    II.   All expected cell counts are at least 5.
    III. The sample size is no more than 10% of the population size.
    (a) I and II only
    (b) II and III only
    (c) I and III only
    (d) II only
    (e) I, II, and III

3. Which of the following represents the component of the chi-square statistic for Wintergreen?
    (a) 22
    (b) $\sqrt{\dfrac{(22-20)^2}{20}}$
    (c) $\sqrt{\dfrac{(22-20)^2}{4}}$
    (d) $\dfrac{(22-20)^2}{22}$
    (e) $\dfrac{(22-20)^2}{20}$

Use the following for questions 4 – 6:

Do male and female children respond differently to colors? A study of color association in children asked separate random samples of male and female fourth-graders what emotion they associated with the color red. Here are the results for each group:

Emotion

	Anger	Happiness	Love	Pain	Total
Female	27	19	39	17	102
Male	34	12	38	28	112
Total	61	31	77	45	214

4. Which of the following would be the appropriate null hypothesis for this test?
   (a) The distribution of emotional associations with the color red is the same for male and female fourth-graders.
   (b) Gender is dependent upon emotional association with the color red.
   (c) Emotional associations with the color red are independent of gender.
   (d) The number of observations in each cell is the same for each emotional association.
   (e) 25% of all fourth graders associate the color red with each of the four listed emotions.

5. Under the assumption that the null hypothesis is true, which of the following represents the expected count for female children who associate the color red with love?
   (a) 39
   (b) $\dfrac{(77)(214)}{102}$
   (c) $\dfrac{(77)(102)}{214}$
   (d) $\dfrac{(39)(102)}{77}$
   (e) $\dfrac{(39)^2}{214}$

6. The chi-square statistic for these data is $X^2 = 4.629$. Which of the following intervals contains the P-value for this test?
   (a) $0.005 \leq P-\text{value} \leq 0.01$
   (b) $0.01 \leq P-\text{value} \leq 0.025$
   (c) $0.025 \leq P-\text{value} \leq 0.05$
   (d) $0.05 \leq P-\text{value} \leq 0.1$
   (e) $P-\text{value} \geq 0.1$

Use the following for questions 7 – 8:

State traffic engineers want to characterize the types of vehicles found on three state roads. They take a random sample of vehicles on each road over a two-week period and get the results in the table for the number of vehicles of each type on each road. The engineers perform a chi-square test of homogeneity, using the null hypothesis that there is no difference in distribution of vehicles types on the four roads.

	Vehicle type		
	Cars	Light trucks/SUVs	Heavy trucks/trailers
Route 9	126	42	16
Route 47	216	51	35
Route 116	271	41	56
Route 176	413	37	24

7. For this chi-square test, what are the correct degrees of freedom?
   (a) 3   (b) 5   (c) 6   (d) 11   (e) 12

8. Below are the individual components for the chi-square statistic for this test:

	Cars	Light trucks/SUVs	Heavy trucks/trailers
Route 9	1.8	14.1	0.3
Route 47	1.3	3.8	0.9
Route 116	0.6	0.9	10.7
Route 176	6.0	9.5	11.1

Based on the original data and the components, which of the following statements is true?
(a) The observed count of heavy trucks/trailers on Route 176 is much higher than the expected count.
(b) There are many more light trucks on Route 9 than we would expect if the null hypothesis were true.
(c) The number of observed cars on Route 116 is much lower than we would expect if the null hypothesis were true.
(d) The greatest difference between observed and expected counts is for heavy trucks/trailers on Route 9.
(e) The chi-square statistic for this test is less than 30.

9. Which of the following statements about chi-square distributions are true?
   I.   A chi-square distribution with fewer than 10 degrees of freedom is roughly symmetric.
   II. The more degrees of freedom a chi-square distribution has, the larger the median of the distribution.
   III. For all chi-square distributions, $P(\chi^2 \geq 0) = 1$

   (a) I only (b)
   II only (c)
   III only (d)
   II and III
   (e) All three statements are true.

10. Is the accident rate for some car colors different than for other car colors? An insurance company selects a random sample of cars that it insures and records their color (using five categories: white, silver, black, red, or "all others") and whether or not they have been involved in an accident in the last three years. They perform a chi-square test of association and obtain a test statistics of $\chi^2 = 8.474$, which yields a P-value of 0.0758. Using a significance level of $\alpha = 0.05$, which of the following is the appropriate conclusion for this test?

(a) Reject H_0: there is convincing evidence of an association between car color and proportion of cars involved in accidents.

(b) Accept H_0: there is convincing evidence that car color and proportion of cars involved in accidents are independent.

(c) Reject H_0: there is insufficient evidence to establish an association between car color and proportion of cars involved in accidents.

(d) Fail to reject H_0: there is insufficient evidence to establish an association between car color and proportion of cars involved in accidents.

(e) Fail to reject H_0: there is convincing evidence that car color and proportion of cars involved in accidents are independent.

Part 2: Free Response

Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

11. Big Box Electronics, a large national chain store, has one store in the city of Kingston. One factor in deciding whether to build a second store in the city is whether the current store is serving all residents equally well, or whether unequal proportions of residents from different parts of town are using the store because it's located on one side of town. The national managers of Big Box divide Kingston into four geographical regions and determine the percentage of residents who live in each region. Here's what they find:

Region	North	South	East	West
Percentage of population	40%	24%	22%	14%

Then the managers take a simple random sample of 250 shoppers at Kingston's Big Box store and determine which part of town they come from by asking for their zip code when they are checking out:

Region	North	South	East	West
Number of shoppers	120	48	62	20

Is Kingston's only Big Box store used by a higher proportion of the residents in some parts of town than others? Support your conclusion with an appropriate statistical test.

12. A few weeks before the senatorial election between incumbent Senator Smirk and his challenger, former Governor Graft, the senator's polling organization wants to know where he should concentrate his campaigning. They take simple random samples of potential voters in the southern and northern portions of the state, and ask them if they have decided who to vote for or are still undecided. Here are the results:

		Decided on a candidate	Still undecided	Total
Region	North	116	60	176
	South	148	52	200
	Total	264	112	376

(a) Do these data provide convincing evidence that there is a difference in the distribution of voters who have decided or are still undecided in the two regions? Use a chi-square test to support your conclusion.

(b) The pollsters are concerned that while all 200 people in the "South" sample responded, 24 people (out or the original SRS of 200) in the "North" sample did not respond. Is it possible that the opinions of these people would change the pollsters' conclusions? Explain.