Test 11C AP Statistics Name:

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

Use the following for questions 1 – 2:

A statistically-minded toll collector wonders if drivers are equally likely to choose each of the three lanes at his toll booth. He selects a random sample from all the cars that approach the booth when all three lanes are empty, so that the driver's choice isn't influenced by the cars already at the booth. Here are his results:

Lane	Left	Center	Right
Number of drivers	137	159	169

1. Which of the following is the appropriate *alternative* hypothesis for addressing this question?
(a) The observed number of cars choosing each lane is equal.
(b) The observed number of cars choosing each lane is different from the expected number of cars.
(c) The proportions of cars choosing each of the three lanes are equal.
(d) The proportions of cars choosing at least one of the lanes is different from the proportion choosing the other two lanes.
(e) The proportions of cars choosing each of the three lanes are all different.

2. Which of the following is the correct expression for the chi-square goodness-of-fit test in this situation?

(a) $\chi^2 = \left(\frac{(137-155)}{155} + \frac{(159-155)}{155} + \frac{(169-155)}{155} \right)^2$

(b) $\chi^2 = \left(\frac{(137-155)}{137} + \frac{(159-155)}{159} + \frac{(169-155)}{169} \right)^2$

(c) $\chi^2 = \left(\frac{(137-155)^2}{155} + \frac{(159-155)^2}{155} + \frac{(169-155)^2}{155} \right)$

(d) $\chi^2 = \left(\frac{(137-155)^2}{137} + \frac{(159-155)^2}{159} + \frac{(169-155)^2}{169} \right)$

(e) $\chi^2 = \left(\frac{(137-155)^2}{465} + \frac{(159-155)^2}{465} + \frac{(169-155)^2}{465} \right)$

3. Which of the following is a condition that must be met in order to carry out a chi-square goodness-of-fit test?
(a) The population must be normally distributed, or the sample size must be greater than 30.
(b) The cell counts for our sample have to be approximately the same as the expected counts.
(c) All observed cell counts must be greater than 5.
(d) All expected counts must be greater than 5.
(e) More than one of these conditions must be met.

Use the following for questions 4 – 7:

A company manufactures special plastic containers that hold a specified volume of hazardous material. Some containers do not meet specifications as required by the company's customer because they are either too small or too large. Are there differences between the sizes of containers produced by the three shifts? Here are the results of inspecting 500 randomly-selected containers from each of the shifts.

		Conformance to specifications			
		Too small	Within specs.	Too large	Total
Shift	8 am	36	452	12	500
	4 pm	24	443	33	500
	Midnight	12	438	50	500
	Total	72	1333	95	1500

4. Which of the following is the appropriate null hypothesis for this test?
(a) The expected counts for each cell are equal.
(b) The observed counts in each cell are equal to the expected counts.
(c) The distribution of the variable "conformance to specifications" is the same for all three shifts.
(d) Conformance to specifications is associated with shift.
(e) The proportion of containers that conform to specification is the same for all three shifts.

5 Which of the following are the degrees of freedom for the distribution of the test statistic in this situation?
(a) 9 (b) 8 (c) 6 (d) 4 (e) 2

6. Which of the following represents the expected count for the cell "Too large/midnight?"
(a) 50 (b) $\frac{1500}{9}$ (c) $\frac{(50)(500)}{1500}$ (d) $\frac{(95)(500)}{50}$ (e) $\frac{(95)(500)}{1500}$

7. A test of the hypothesis in question 4 is statistically significant at the $\alpha = 0.005$ level. Below are the individual components of the chi-square statistic for these data:

		Conformance to specifications		
		Too small	Within specs.	Too large
Shift	8 am	6.000	0.132	12.214
	4 pm	0	0.004	0.056
	Midnight	6.000	0.090	10.614

Based on these components and original two-way table, which of the following statements is true?
(a) The observed count of containers produced by the midnight that were within specifications was very close to the expected number.
(b) The observed count of containers produced by the 8 am shift that were within specifications was much higher than the expected number.
(c) The 4 pm shift has the smallest proportion of containers that are either too large or too small.
(d) The 4 pm shift does not produce any containers that are too small.
(e) Most of the containers produced by the 8 am shift that were "within specifications" were closer to being too small than too large.

Use the following for questions 8 – 9:

Are all employees equally prone to having accidents? To investigate this hypothesis, a researcher looked at a light manufacturing plant and classified a random sample of accidents by type and by age of the employee.

		Accident Type		
		Sprain	Burn	Cut
Age	Under 25	9	17	5
	25 or over	61	13	12

8. Here are the expected counts for this table under the null hypothesis that tests the question stated above:

		Accident Type		
		Sprain	Burn	Cut
Age	Under 25	18.5	7.9	4.5
	25 or over	51.5	22.1	12.5

What is the individual component of the chi-square statistic for the cell "Under 25/Burn?"

(a) 1.15 (b) 4.87 (c) 5.64 (d) 9.10 (e) 10.48

9. A chi-square test for the above data gave a test statistic of 20.78. If we test at $\alpha = 0.05$ which of the following is an appropriate conclusion?

(a) Reject H_0; the test does not provide convincing evidence of an association between accident type and age.

(b) Reject H_0; the test provides convincing evidence that accident type is not independent of age.

(c) Reject H_0; the test provides convincing evidence that age is independent of accident type.

(d) Fail to reject H_0; the test provides convincing evidence that there is a correlation between age and accident type.

(e) Fail to reject H_0; the test provides convincing evidence that the proportion of sprain, cuts, and burns seems to be similar for both age classes.

10. A survey was conducted to investigate whether alcohol consumption and smoking are related. In an SRS of 300 smokers, 196 said they had consumed alcohol at least once in the past week. In an independent SRS of 300 non-smokers, 159 said they had consumed alcohol in the past week. If p_s is the proportion of smokers in the population who have had a drink in the past week and p_{ns} is the corresponding proportion of non-smokers, then a test of the hypotheses $H_0 : p_s - p_{ns} = 0$ against the two-sided alternative produces a test statistic of $z = 3.07$ and a P-value of 0.002. If we had instead analyzed these results with a chi-square test of homogeneity, which of the following would be the test statistic and P-value?

(a) $\chi^2 = 9.42;\ P - \text{value} = 0.002$

(b) $\chi^2 = 9.42;\ P - \text{value} = 0.004$

(c) $\chi^2 = 3.07;\ P - \text{value} = 0.004$

(d) $\chi^2 = 1.75;\ P - \text{value} = 0.002$

(e) $\chi^2 = 1.75;\ P - \text{value} = 0.004$

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. Are you more likely to be allergic to house dust mites if you are born in a certain season? Recent research suggests this might be true. Below are the birth seasons of 500 randomly selected people who are allergic to house dust mites, along with the expected proportion of people born in each season, based on data from the entire population. Do these data provide evidence that people who suffer from this allergy have a different birth season distribution? Support your conclusion with the appropriate statistical test.

Birth Season	Number of allergy sufferers	Proportion of births in general population
Winter	117	0.30
Spring	105	0.22
Summer	145	0.24
Fall	133	0.24

12. Which modern electronic devices do you care most about? A simple random sample of 120 people were asked two questions: their age, and "If you had to give up one of the following three electronic services for three months, which would you choose to give up?" Only people who actually owned all three devices were interviewed. Here are the results:

		Which service would you give up?			
		Cable TV	Home internet	Cell phone	Total
Age Group	Under 25	24	5	3	32
	25 – 50	20	13	6	39
	Over 50	7	8	14	29
	Total	51	26	23	100

(a) In the blank table below, provide the conditional distributions that best describe the relationship between age group and which device people would be willing to give up, based on the data above. Summarize the relationship in a sentence or two.

		Which service would you give up?		
		Cable TV	Home internet	Cell phone
Age Group	Under 25			
	25 – 50			
	Over 50			

(b) Provide the appropriate statistical test to support your conclusion in part (a).

Test 11D (Cumulative) AP Statistics Name:

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

Use the following for questions 1 – 3:

A recent poll by Pew research asked 856 randomly selected adults about their education level and whether or not they have a Facebook account. Here are the results.

	Facebook account?		
Education Level	Yes	No	Total
High school or less	157	104	261
Some college	183	67	250
College degree or more	235	110	345
Total	575	281	856

1. If we are interested in whether there is a relationship between education level and use of Facebook, which of the following is the appropriate null hypothesis?

(a) $H_0 : p_{1,1} = p_{1,2} = p_{1,3} = p_{2,1} = p_{2,2} = p_{2,3}$

(b) The distribution of Education level is the same for both people who have Facebook accounts and people who don't.

(c) Sample proportions for the variable "Facebook account" are the same for all education levels.

(d) The variables "Education level" and "Facebook account" are independent.

(e) There is an association between the variables "Education level" and "Facebook account."

2. The two-way table below gives the chi-square components for the appropriate test on these data.

	Facebook account?	
Education Level	Yes	No
High school or less	1.92	3.92
Some college	1.35	2.77
College degree or more	0.05	0.09

Based on this information, which of the following statements are valid?

I. The observed counts for the "College degree or more" row are farther from their corresponding expected counts than the observed counts for "High School or less" and "Some college."

II. "High school or less" individuals differ the most from the sample as a whole.

III. The observed count for "Some college/No Facebook" is considerably higher than we would expect if the null hypothesis were true.

(a) I only

(b) II only

(c) III only

(d) II and III only

(e) I, II, and III are valid.

3. The χ^2 statistic for this test is 10.09. Which of the following gives the correct interval for the *P*-value of the test?

(a) *P*-value < 0.01
(b) 0.01 < *P*-value < 0.02
(c) 0.02 < *P*-value < 0.05
(d) 0.05 < *P*-value < 0.1
(c) 0.15 < *P*-value < 0.10

4. Which one of the following statements about χ^2 distributions is true?

(a) A χ^2 distribution with 20 degrees of freedom is more strongly skewed right than a χ^2 distribution with 2 degrees of freedom.

(b) For any positive number C, $P\left(\chi^2 > C\right) < P\left(\chi^2 > 2C\right)$

(c) $P\left(\chi^2 > 0\right) = P\left(\chi^2 < 0\right)$

(d) In any χ^2 distribution, the median is greater than the mean.

(e) For any χ^2 distribution, $P\left(\chi^2 \geq 0\right) = 1$

5. A consumer electronics store sells a particular brand of protective case for smartphones in four different colors. Below is data on color preference for 76 randomly selected sales of these cases in a three-month period.

Red	Blue	Green	Yellow
23	20	16	17

For the null hypothesis that the proportion of total sales for each color is the same, what is the expected cell count for "Green"?

(a) 20 (b) 19 (c) 16 (d) 0.25 (e) 0.21

6. A psychology professor at a large university asks 100 randomly selected students majoring in the humanities and 100 randomly selected math/science majors which of two brands of smartphone they prefer, "Brand A" or "Brand S." Here are her results:

	Brand A	Brand S
Humanities	58	42
Math/Science	47	53

If the professor wants to determine if humanities and math/science majors differ significantly in their preferences for these two brands, which of the following tests should she use?

(a) A z-test for the difference of two proportions
(b) A chi-square goodness of fit test.
(c) A chi-square test for homogeneity.
(d) A chi-square test for independence.
(e) Either (a) or (c).

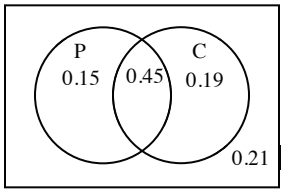
7. The mean score on Mr. Groucher's AP Calculus test was $\mu = 68.5$ and the standard deviation was $\sigma = 15.4$. The scores were distinctly skewed left. Ace's test score was 58. Which of the following is closest to Ace's z-score?

(a) -0.68
(b) -0.15
(c) 0.15
(d) 0.68
(e) It is not appropriate to calculate a z-score for a distribution that is non-Normal.

8. Which one of the following potential problems is addressed by performing a blind study?

(a) Confounding
(b) The placebo effect
(c) Bias associated with voluntary response
(d) Error arising from random assignment
(e) Type II error

9. Choose a random shopper at a local grocery store. Let Event P = The shopper is buying some kind of fresh produce. Let Event C = The shopper is paying with a credit card. Below is a Venn Diagram of the probabilities associated with these two events.



Which one of the two-way tables below conveys the same information?

(a)

	P	**P^c**
C	**0.45**	**0.64**
C^c	**0.60**	**0.21**

(b)

	P	**P^c**
C	**0.45**	**0.19**
C^c	**0.15**	**0.21**

(c)

	P	**P^c**
C	**0.79**	**0.19**
C^c	**0.15**	**0.21**

(d)

	P	**P^c**
C	**0.15**	**0.19**
C^c	**0.21**	**0.45**

(e)

	P	**P^c**
C	**0.15**	**0.21**
C^c	**0.19**	**0.45**

10. Suppose you are conducting a test of significance for which $H_0 : p = 0.7$. Which of the following tests against a specific alternative hypothesis has the greatest power?

(a) $H_a : p = 0.68,$ with $n = 40$ and $\alpha = 0.05$
(b) $H_a : p = 0.68,$ with $n = 80$ and $\alpha = 0.01$
(c) $H_a : p = 0.68,$ with $n = 80$ and $\alpha = 0.005$
(d) $H_a : p = 0.73,$ with $n = 80$ and $\alpha = 0.05$
(e) $H_a : p = 0.73,$ with $n = 40$ and $\alpha = 0.01$

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. A recent study published in the journal *Pediatrics* examined research on emergency room visits by randomly selected pediatric patients suffering from abdominal pain. The following data was collected on the ethnicity of each patient and whether or not they were given medication for pain relief.

	Given Medication?		
	Yes	No	**Total**
White	**330**	**888**	1218
Black	**87**	**465**	552
Hispanic	**91**	**392**	483
Other	**3**	**42**	45
Total	511	1787	2298

(a) Suppose one of these patients were selected at random. what is the probability that she was provided with pain medication, given that she was Hispanic?

(b) In the blank table at right, given the conditional distributions that best illustrate the impact that ethnicity has on treatment for abdominal pain, based on the data above.

	Given Medication?	
	Yes	No
White		
Black		
Hispanic		
Other		

(c) Do these data provide evidence that pediatric patients from different ethnic groups received different treatments for abdominal pain? The χ^2 statistic for the appropriate test is 39.53 (df = 3) and the *P*-value is nearly 0. Provide the rest of the required information for a complete significance test to support your answer.

12. In a large national survey, middle school students were asked which one of the following they would prefer to be later in life: happy, healthy, famous, or rich. 49% said they would prefer to be happy, 16% said healthy, 15% famous, and 20% rich. Let's assume these percentages represent parameter values for the entire population of U.S. middle school students.

(a) The principle of a large middle school (1000 students) wonders if the preferences of his students differ from the national distribution of preferences. Describe a method by which the principle could select a simple random sample of 80 of his students to survey about their preferences.

(b) The principle employs the methods you described in (a). Here are his results:

Happy	Healthy	Famous	Rich
48	15	9	8

Does this provide evidence that this school's preferences are different from national preferences? Support your answer with appropriate statistical methods.

(c) You had advised the principal to randomly arrange the four options on each of the questionnaires, so that "Happy" only appeared first about one-fourth of the time. He chose not to do this, and asked in the order the data has been presented. Describe one source of bias that might arise from his method and include the direction of that bias.
