Problems Section 4D

(#1-10) Use each of the following Goodness of Fit χ^2-test statistics and the corresponding critical values to fill out the table.

	χ^2-test stat	Sentence to explain χ^2-test statistic.	Critical Value	Does the χ^2-test statistic fall in a tail determined by the critical value? (Yes or No)	Does sample data significantly disagree with H_0?
1.	+28.573		+9.117		
2.	+1.226		+7.113		
3.	+2.137		+5.521		
4.	+14.415		+6.114		
5.	+3.718		+7.182		
6.	+0.891		+3.994		
7.	+51.652		+14.881		
8.	+1.185		+4.181		
9.	+2.442		+8.619		
10.	+14.133		+10.336		

(#11-20) Use each of the following P-values and corresponding significance levels to fill out the table.

	P-value Proportion	P-value %	Sentence to explain the P-value	Significance Level %	Significance level Proportion	If H_0 is true, could the sample data occur by random chance or is it unlikely?	Reject H_0 or Fail to reject H_0?
11.	0.0006			10%			
12.	0.042			1%			
13.	9.16×10^{-7}			5%			
14.	0.739			1%			
15.	0.0035			5%			
16.	0			10%			
17.	0.419			5%			
18.	0.0274			10%			
19.	3.77×10^{-5}			1%			
20.	0.067			5%			

21. How is the degrees of freedom calculated in a Goodness of Fit test?

22. The χ^2-test statistic compares the observed sample counts to the expected counts from H_0. Explain how the expected counts are calculated.

23. Explain how the χ^2-test statistic is calculated from the observed and expected counts.

24. If the observed sample counts were significantly different from the expected counts, would the χ^2-test statistic be large or small? Explain why.

25. If the observed sample counts were close to the expected counts, would the χ^2-test statistic be large or small? Explain why.

(#26-29) Directions: Use StatKey at www.lock5stat.com to simulate the following Chi-squared Goodness of Fit tests. Go to "more advanced randomization tests" at the bottom of the StatKey page. Click on the button that says "χ^2 Goodness of Fit". Under "Edit Data", type in the given sample data. Create a randomized simulation of the null hypothesis to answer the following questions.

a) Write the null and alternative hypothesis. Include relationship implications.

b) What is the degrees of freedom?

c) What is the Chi-squared test statistic? Write a sentence to explain the test statistic.

d) Adjust the right tail of your simulation to reflect the significance level. Did the Chi-squared test statistic fall in the tail?

e) Does the sample data significantly disagrees with the null hypothesis? Explain your answer.

f) Are the observed counts in the sample data significantly different from the expected counts from the null hypothesis? Explain your answer.

g) Put the Chi-squared test statistic into the bottom box in the right tail of your simulation in order to calculate the P-value. What was the P-value? (Answers will vary.) Write a sentence to explain the P-value.

h) Use the P-value and significance level to determine if the sample data could have occurred by random chance (sampling variability) or is it unlikely to random chance? Explain your answer.

i) Should we reject the null hypothesis or fail to reject the null hypothesis? Explain your answer.

j) Write a conclusion for the hypothesis test. Explain your conclusion in plain language.

k) Is the population proportion related to the categorical variable or not? Explain your answer.

26. It is a big job to write and grade the AP-statistics exam for high school students each year. It is a difficult multiple-choice exam. All questions have five possible answers A-E. Use a 5% significance level and the following sample data to test the claim that percent of A answers is the same as the percent of B answers which is the same as C, D and E. This would indicate that the letter of the answer is not related to the percentage of times it happens. You can assume that the sample data meets the assumptions. Type the following sample data under the "Edit Data" menu of StatKey.
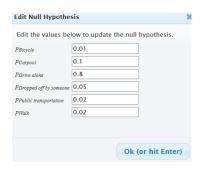
Choice, Count
A, 85
B, 90
C, 79
D, 78
E, 68

27. We collected data from all of the math 140 statistics students in the fall 2015 semester. A person that works at COC thinks that 80% of COC students drive alone, 10% carpool, 5% are dropped off by someone, 2% walk, 1% bike, and 2% use public transportation. Use a 5% significance level and the following sample data to test the claim that these percentages are wrong. You can assume that the data meets the assumptions for inference. Type in the proportions under "Null Hypothesis" in StatKey. Under "Edit Data" type in the following sample data from the fall 2015 survey data.

Choice, Count
Bicycle, 1
Carpool, 30
Drive Alone, 267
Dropped Off, 18
Public Transportation, 6
Walk, 10

Edit Null Hypothesis

Edit the values below to update the null hypothesis.

P_Bicycle	0.01
P_Carpool	0.1
P_Drive alone	0.8
P_Dropped off by someone	0.05
P_Public transportation	0.02
P_Walk	0.02

Ok (or hit Enter)

28. We collected data from all of the math 140 statistics students in the fall 2015 semester. Use a randomized simulation in StatKey, a 5% significance level, and the following sample data to test the claim that the population percentages for the different political parties are different. This would indicate that the political party is related to the population percentages. You can assume that the data meets the assumptions for inference. Under "Edit Data", type in the following sample data from the fall 2015 survey data.

Choice, Count
Democratic, 110
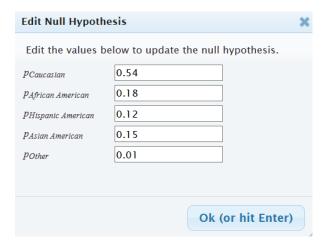Republican, 63
Independent, 65
Other, 90

29. Juries are required to meet the racial demographic of the county they represent. Here is the racial demographic for Alameda county: 54% Caucasian, 18% African American, 12% Hispanic American, 15% Asian American, and 1% other. We are worried that the juries in Alameda County may not be representing these percentages. Use randomized simulation, a 1% significance level, and the following observed sample counts to test the claim that the juries do not represent the demographic of the county. Under the "Edit Data" menu in StatKey, type in the following sample counts.

Jury Sample Data Observed Counts

Race, Count
Caucasian, 780
African American, 117
Hispanic American, 114
Asian American, 384
Other, 58

Under the "Null Hypothesis" menu, type in the following.

Edit Null Hypothesis ✖

Edit the values below to update the null hypothesis.

PCaucasian	0.54
PAfrican American	0.18
PHispanic American	0.12
PAsian American	0.15
POther	0.01

Ok (or hit Enter)

(#30-32) Directions: Use the following Statcato printouts to answer the following questions.

a) Write the null and alternative hypothesis. Include relationship implications.

b) Check the assumptions for a Goodness of Fit test.

c) What is the Chi-squared test statistic? Write a sentence to explain the test statistic.

d) Did the Chi-squared test statistic fall in the tail determined by the critical value?

e) Does the sample data significantly disagrees with the null hypothesis? Explain your answer.

f) Are the observed counts in the sample data significantly different from the expected counts from the null hypothesis? Explain your answer.

g) What was the P-value? Write a sentence to explain the P-value.

h) Use the P-value and significance level to determine if the sample data could have occurred by random chance (sampling variability) or is it unlikely to random chance? Explain your answer.

i) Should we reject the null hypothesis or fail to reject the null hypothesis? Explain your answer.

j) Write a conclusion for the hypothesis test. Explain your conclusion in plain language.

k) Is the population proportion related to the categorical variable or not? Explain your answer.

30. An online sports magazine wrote an article about the favorite sports in America. It said that 43% of Americans prefer Football, 23% of Americans prefer Baseball, 20% of Americans prefer Basketball, 8% of Americans prefer Hockey, and 6% of Americans prefer Soccer. When 130 randomly selected adults were asked their favorite sport, we found the following: 44 said Football, 26 said Baseball, 29 said Basketball, 13 said Hockey, and 18 said Soccer. Use a 5% significance level to test the claim that the proportions match the distribution claimed in the magazine article.

Chi-Square Goodness-of-Fit Test:
Input: C1 Observed Counts#4
Expected probabilities in C2 Null Hypothesis

Category	Observed Frequency	Expected Frequency	Contribution to x^2
0	44.0	55.9	2.5333
1	26.0	29.9000	0.5087
2	29.0	26.0	0.3462
3	13.0	10.4	0.6500
4	18.0	7.8	13.3385

N	Number of Categories	DOF	Significance	Critical Value	Test statistics	p-Value
130.0	5	4	0.05	9.4878	17.3766	0.0016

31. Thousands of people die from car accidents across the U.S. every year, but is the day of the week related to the probability of having a fatal car accident? To test this claim, use a 1% significance level and a Goodness of Fit test to determine if the probabilities of a fatal car accident are significantly different. The following random sample data summary gives the observed number of the number of deaths from car accidents in the U.S. for each day of a randomly selected week. The total number of deaths for the week was 805.

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Number of Fatal Car Accidents	106	104	103	113	130	132	117

Chi-Square Goodness-of-Fit Test:
Input: C5 Observed #5
Expected frequency = 115.0

Category	Observed Frequency	Expected Frequency	Contribution to x^2
0	106.0	115.0	0.7043
1	104.0	115.0	1.0522
2	103.0	115.0	1.2522
3	113.0	115.0	0.0348
4	130.0	115.0	1.9565
5	132.0	115.0	2.5130
6	117.0	115.0	0.0348

N	Number of Categories	DOF	Significance	Critical Value	Test statistics	p-Value
805.0	7	6	0.01	16.8119	7.5478	0.2731

32. The National Highway Traffic Safety Administration (NHTSA) publishes reports about motorcycle fatalities and helmet use. The following distribution shows the proportion of fatalities by location of injury for motorcycle accidents.

Location of Injury	Multiple Locations	Head	Neck	Thorax	Abdomen/Spine
Proportion	0.57	0.31	0.03	0.06	0.03

The random sample data below shows the distribution of 2068 randomly selected fatalities from riders that were not wearing a helmet. Use a 0.01 significance level to test the claim that the distribution for the sample does not match the proportions given by the NHTSA. Where is the largest discrepancy between the observed and expected value? What does this tell us about the importance of wearing helmets?

Location of Injury	Multiple Locations	Head	Neck	Thorax	Abdomen/Spine
Number of Deaths	1036	864	38	83	47

Chi-Square Goodness-of-Fit Test:
Input: C3 observed#6
Expected probabilities in C4 Ho#6

Category	Observed Frequency	Expected Frequency	Contribution to X^2
0	1036.0	1178.76	17.2897
1	864.0	641.08	77.5150
2	38.0	62.04	9.3153
3	83.0	124.08	13.6006
4	47.0	62.04	3.6461

N	Number of Categories	DOF	Significance	Critical Value	Test statistics	p-Value
2068.0	5	4	0.01	13.2767	121.3667	0