Problems Section 4D (updated version)

(#1-10) Use the number of groups (k) to calculate the degrees of freedom (k – 1). Go to StatKey at www.lock5stat.com. Under the "Theoretical Distributions" menu, click on χ^2 . Enter the degrees of freedom (k – 1) and click "Right Tail" and type the significance level in the upper right box. The lower right box will now be the critical value and show where the right tail begins. Use the given χ^2 Test Statistic and the calculated Critical Value to fill out the table. Assume the test statistics came from unbiased data passing conditions.

	Type of Test	χ^2-test stat	Number of Groups (k)	Degrees of Freedom (k – 1)	Significance Level Proportion (put in upper right tail)	Critical Value? (Bottom Right Tail box in StatKey)	Does the χ^2-test statistic fall in the right tail determined by the critical value?	Do the observed sample counts significantly disagree with the expected counts from the null hypothesis?
1.	Right Tail	+28.573	10		0.1			
2.	Right Tail	+1.226	3		0.01			
3.	Right Tail	+2.137	4		0.05			
4.	Right Tail	+14.415	8		0.01			
5.	Right Tail	+3.718	5		0.05			
6.	Right Tail	+0.891	6		0.1			
7.	Right Tail	+51.652	14		0.05			
8.	Right Tail	+1.185	3		0.1			
9.	Right Tail	+2.442	5		0.01			
10.	Right Tail	+14.133	7		0.05			

(#11-20) Use the number of groups (k) to calculate the degrees of freedom (k – 1). Go to StatKey at www.lock5stat.com. Under the "Theoretical Distributions" menu, click on χ^2 . Enter the degrees of freedom (k – 1) and click "Right Tail" and type the χ^2 test statistic in the bottom right box. The upper right box will now be the P-value proportion. Use the calculated P-values and corresponding significance levels to fill out the table. Assume the P-values came from unbiased data passing conditions.

	χ^2-test stat	Number of Groups (k)	Degrees of Freedom (k – 1)	P-value Proportion? (Top RightTail box in StatKey)	P-value %	Significance Level %	Is the P-value lower than sig. level or higher?	If H_0 is true, could the test stat or more extreme occur by sampling variability or is it unlikely?	Reject H_0 or Fail to reject H_0?
11.	+28.573	7				10%			
12.	+1.226	3				1%			
13.	+2.137	4				5%			
14.	+14.415	8				1%			
15.	+3.718	5				5%			
16.	+0.891	6				10%			
17.	+51.652	4				5%			
18.	+1.185	3				10%			
19.	+2.442	5				1%			
20.	+14.133	4				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. What is the χ^2-test statistic formula?

24. Write the χ^2-test statistic definition sentence.

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

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

(#27-29) Use the following Statcato Goodness of Fit Test printouts to answer the questions and perform the Goodness of Fit Test.

27. 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. Assume the people in the sample were not related to each other. 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

a) Write the null and alternative hypothesis.

b) How many total groups (k) does this data have?

c) What is the degrees of freedom = k – 1

d) What is the Chi-square test statistic?

e) What are the expected counts? Are they all greater than 5?

f) Assuming the data was collected randomly and individuals were not related, does the data pass all of the conditions for the Goodness of Fit test?

g) What is the critical value (where the right tail begins)?

h) Does the Chi-square test statistic fall in the right tail beginning at the critical value?

i) Are the observed counts in the sample data significantly different from the expected counts from the null hypothesis? Explain why.

j) What was the P-value proportion? Convert the P-value into a percentage.

k) Is the P-value lower or higher than the significance level?

l) Could the chi-square test statistic or more extreme have occurred because of sampling variability or is that unlikely? Explain how you know.

m) Should we reject the null hypothesis (significant evidence) or fail to reject the null hypothesis (not significant evidence)? Explain your answer.

n) Write the formal conclusion for the hypothesis test.

28. We want to test the claim that the day of the week is not related to having a fatal car accident? To test this claim, use a 1% significance level and a Goodness of Fit test to determine if the probability of having a fatal car accident is the same on each day of the week. 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.

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:
nput: 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

a) Write the null and alternative hypothesis.

b) How many total groups (k) does this data have?

c) What is the degrees of freedom = k – 1

d) What is the Chi-square test statistic?

e) What are the expected counts? Are they all greater than 5?

f) Assuming the data was collected randomly and individuals were not related, does the data pass all of the conditions for the Goodness of Fit test?

g) What is the critical value (where the right tail begins)?

h) Does the Chi-square test statistic fall in the right tail beginning at the critical value?

i) Are the observed counts in the sample data significantly different from the expected counts from the null hypothesis? Explain why.

j) What was the P-value proportion? Convert the P-value into a percentage.

k) Is the P-value lower or higher than the significance level?

l) Could the chi-square test statistic or more extreme have occurred because of sampling variability or is that unlikely? Explain how you know.

m) Should we reject the null hypothesis (significant evidence) or fail to reject the null hypothesis (not significant evidence)? Explain your answer.

n) Write the formal conclusion for the hypothesis test.

29. The National Highway Traffic Safety Administration (NHTSA) publishes reports about motorcycle accidents and helmet use. The following distribution shows the proportion of injuries by location 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 motorcycle accidents from riders that were not wearing a helmet. Use a 1% significance level to test the claim that the distribution for the sample does not match the proportions given by the NHTSA.

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

a) Write the null and alternative hypothesis.

b) How many total groups (k) does this data have?

c) What is the degrees of freedom = k – 1

d) What is the Chi-square test statistic?

e) What are the expected counts? Are they all greater than 5?

f) Assuming the data was collected randomly and individuals were not related, does the data pass all of the conditions for the Goodness of Fit test?

g) What is the critical value (where the right tail begins)?

h) Does the Chi-square test statistic fall in the right tail beginning at the critical value?

i) Are the observed counts in the sample data significantly different from the expected counts from the null hypothesis? Explain why.

j) What was the P-value proportion? Convert the P-value into a percentage.

k) Is the P-value lower or higher than the significance level?

l) Could the chi-square test statistic or more extreme have occurred because of sampling variability or is that unlikely? Explain how you know.

m) Should we reject the null hypothesis (significant evidence) or fail to reject the null hypothesis (not significant evidence)? Explain your answer.

n) Write the formal conclusion for the hypothesis test.

o) Which injury location has the largest discrepancy between the observed and expected value? What does this tell us about the importance of wearing helmets when riding motorcycles?

(#30-33) Directions: Use StatKey at www.lock5stat.com to perform a Chi-square Goodness of Fit test.

30. 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. Assume the data was collected randomly and individuals were not related.

a) Write the null and alternative hypothesis.

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", copy and paste the given sample data and click "Ok". Under "Original Sample" write down the χ^2 test statistic. Also make a note of how many groups (k) you have. Next to "Original Sample" click on "Show Details". Write down the middle value expected counts in green.

Letter, Count
A, 85
B, 90
C, 79
D, 78
E, 68

b) How many total groups (k) does this data have?

c) Calculate the degrees of freedom = k – 1

d) What is the Chi-square test statistic?

e) What are the expected counts? Are they all greater than 5?

f) Assuming the data was collected randomly and individuals were not related, does the data pass all of the conditions for the Goodness of Fit test?

Click on "χ^2" under the "Theoretical Distributions" menu in StatKey. Type in the degrees of freedom and click on "Ok". Click on "Right Tail". Type in the significance level proportion into the top box in the right tail. The lower box in the right tail is now the critical value.

g) What is the critical value (where the right tail begins)?

h) Does the Chi-square test statistic fall in the right tail beginning at the critical value?

i) Are the observed counts in the sample data significantly different from the expected counts from the null hypothesis? Explain why.

Click on "χ^2" under the "Theoretical Distributions" menu in StatKey. Type in the degrees of freedom and click on "Ok". Click on "Right Tail". Type in the chi-square test statistic into the bottom box in the right tail. The upper box in the right tail is now the P-value.

j) What was the P-value proportion? Convert the P-value into a percentage.

k) Is the P-value lower or higher than the significance level?

l) Could the chi-square test statistic or more extreme have occurred because of sampling variability or is that unlikely? Explain how you know.

m) Should we reject the null hypothesis (significant evidence) or fail to reject the null hypothesis (not significant evidence)? Explain your answer.

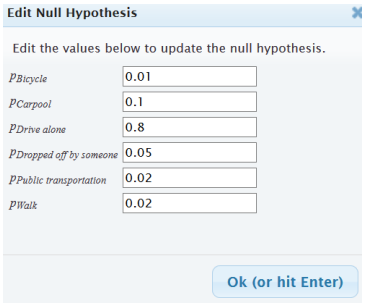
n) Write the formal conclusion for the hypothesis test.

31. We collected sample data from statistics students in the fall 2015 semester. A person that works at COC thinks that 80% of stat 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, StatKey, and the following sample data to test the claim that these percentages are correct. Assume the data represents the population and individuals are not related.

a) Write the null and alternative hypothesis.

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", copy and paste the given sample data and click "Ok". Type in the null hypothesis proportions under "Null Hypothesis" in StatKey. Under "Original Sample" write down the χ^2 test statistic. Also make a note of how many groups (k) you have. Next to "Original Sample" click on "Show Details". Write down the middle value expected counts in green.

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



b) How many total groups (k) does this data have?

c) Calculate the degrees of freedom = k – 1

d) What is the Chi-square test statistic?

e) What are the expected counts? Are they all greater than 5?

f) Assuming the data represents the population and individuals were not related, does the data pass all of the conditions for the Goodness of Fit test?

Click on "χ^2" under the "Theoretical Distributions" menu in StatKey. Type in the degrees of freedom and click on "Ok". Click on "Right Tail". Type in the significance level proportion into the top box in the right tail. The lower box in the right tail is now the critical value.

g) What is the critical value (where the right tail begins)?

h) Does the Chi-square test statistic fall in the right tail beginning at the critical value?

i) Are the observed counts in the sample data significantly different from the expected counts from the null hypothesis? Explain why.

Click on "χ^2" under the "Theoretical Distributions" menu in StatKey. Type in the degrees of freedom and click on "Ok". Click on "Right Tail". Type in the chi-square test statistic into the bottom box in the right tail. The upper box in the right tail is now the P-value.

j) What was the P-value proportion? Convert the P-value into a percentage.

k) Is the P-value lower or higher than the significance level?

l) Could the chi-square test statistic or more extreme have occurred because of sampling variability or is that unlikely? Explain how you know.

m) Should we reject the null hypothesis (significant evidence) or fail to reject the null hypothesis (not significant evidence)? Explain your answer.

n) Write the formal conclusion for the hypothesis test.

32. We collected sample data from statistics students in the fall 2015 semester. Use StatKey, a 5% significance level, and the following sample data to test the claim that at least one of the population percentages for the political affiliations are different. You can assume that the data represents the population and individuals are independent.

a) Write the null and alternative hypothesis.

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", copy and paste the given sample data and click "Ok". Under "Original Sample" write down the χ^2 test statistic. Also make a note of how many groups (k) you have. Next to "Original Sample" click on "Show Details". Write down the middle value expected counts in green.

Political Affiliation, Count
Democratic, 110
Republican, 63
Independent, 65
Other, 90

b) How many total groups (k) does this data have?

c) Calculate the degrees of freedom = k – 1

d) What is the Chi-square test statistic?

e) What are the expected counts? Are they all greater than 5?

f) Assuming the data represents the population and individuals were not related, does the data pass all of the conditions for the Goodness of Fit test?

Click on "χ^2" under the "Theoretical Distributions" menu in StatKey. Type in the degrees of freedom and click on "Ok". Click on "Right Tail". Type in the significance level proportion into the top box in the right tail. The lower box in the right tail is now the critical value.

g) What is the critical value (where the right tail begins)?

h) Does the Chi-square test statistic fall in the right tail beginning at the critical value?

i) Are the observed counts in the sample data significantly different from the expected counts from the null hypothesis? Explain why.

Click on "χ^2" under the "Theoretical Distributions" menu in StatKey. Type in the degrees of freedom and click on "Ok". Click on "Right Tail". Type in the chi-square test statistic into the bottom box in the right tail. The upper box in the right tail is now the P-value.

j) What was the P-value proportion? Convert the P-value into a percentage.

k) Is the P-value lower or higher than the significance level?

l) Could the chi-square test statistic or more extreme have occurred because of sampling variability or is that unlikely? Explain how you know.

m) Should we reject the null hypothesis (significant evidence) or fail to reject the null hypothesis (not significant evidence)? Explain your answer.

n) Write the formal conclusion for the hypothesis test.

33. 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 StatKey, 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.

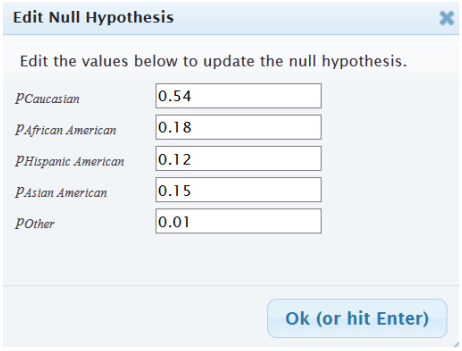
a) Write the null and alternative hypothesis.

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", copy and paste the given sample data below and click "Ok". Type in the null hypothesis proportions given below under "Null Hypothesis" in StatKey. Under "Original Sample" write down the χ^2 test statistic. Also make a note of how many groups (k) you have. Next to "Original Sample" click on "Show Details". Write down the middle value expected counts in green.

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.



b) How many total groups (k) does this data have?

c) Calculate the degrees of freedom = k – 1

d) What is the Chi-square test statistic?

e) What are the expected counts? Are they all greater than 5?

f) Assuming the data represents the population and individuals were not related, does the data pass all of the conditions for the Goodness of Fit test?

Click on "χ^2" under the "Theoretical Distributions" menu in StatKey. Type in the degrees of freedom and click on "Ok". Click on "Right Tail". Type in the significance level proportion into the top box in the right tail. The lower box in the right tail is now the critical value.

g) What is the critical value (where the right tail begins)?

h) Does the Chi-square test statistic fall in the right tail beginning at the critical value?

i) Are the observed counts in the sample data significantly different from the expected counts from the null hypothesis? Explain why.

Click on "χ^2" under the "Theoretical Distributions" menu in StatKey. Type in the degrees of freedom and click on "Ok". Click on "Right Tail". Type in the chi-square test statistic into the bottom box in the right tail. The upper box in the right tail is now the P-value.

j) What was the P-value proportion? Convert the P-value into a percentage.

k) Is the P-value lower or higher than the significance level?

l) Could the chi-square test statistic or more extreme have occurred because of sampling variability or is that unlikely? Explain how you know.

m) Should we reject the null hypothesis (significant evidence) or fail to reject the null hypothesis (not significant evidence)? Explain your answer.

n) Write the formal conclusion for the hypothesis test.
