

## Practice Problems Section 4F

(#1-10) Use each of the following categorical association  $\chi^2$ -test statistics and the corresponding critical values to fill out the table.

	$\chi^2$ -test stat	Sentence to explain $\chi^2$ -test statistic.	Critical Value	Does the $\chi^2$ -test statistic fall in a tail determined by the critical value? (Yes or No)	Does sample data significantly disagree with $H_0$ ?
1.	+1.573		+4.117		
2.	+6.226		+5.118		
3.	+2.144		+4.121		
4.	+3.415		+5.091		
5.	+13.718		+7.189		
6.	+0.972		+4.812		
7.	+31.652		+12.557		
8.	+11.185		+5.181		
9.	+25.443		+7.008		
10.	+1.133		+8.336		

(#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.	0.263			10%			
12.	0.0042			1%			
13.	$5.22 \times 10^{-4}$			5%			
14.	0.0639			1%			
15.	0			5%			
16.	0.539			10%			
17.	0.0419			5%			
18.	0.0027			10%			
19.	$7.73 \times 10^{-8}$			1%			
20.	0.674			5%			

21. If we have two raw categorical data sets, what must we click on in Statcato to perform a categorical association test?
22. If we have summary counts organized in a contingency table, what must we click on in Statcato to perform a categorical association test?
23. What are the assumptions for a categorical association test if the data was collected from on random sample?
24. What are the assumptions for a categorical association test if the data was collected from multiple random samples?
25. How are the expected counts calculated in a categorical association test?
26. If the expected counts from the null hypothesis are significantly different from the observed sample counts, describe the effect on the Chi-Squared test statistic.
27. If the expected counts from the null hypothesis are close to the observed sample counts, describe the effect on the Chi-Squared test statistic.



(#28-31) Directions: For each of the following problems, use the Statcato printout provided to answer the following questions.

- Write the null and alternative hypothesis. Make sure to label which one is the claim.
- Check the assumptions for the categorical association test.
- What is the Chi-squared test statistic? Write a sentence to explain the test statistic.
- Does the test statistic fall in the tail determined by the critical value?
- Does the sample data significantly disagree with the null hypothesis? Explain your answer.
- Are the observed counts significantly different from the expected counts? Explain your answer.
- What is the P-value? Write a sentence to explain the P-value.
- Compare the P-value to the significance level. Should we reject the null hypothesis or fail to reject the null hypothesis? Explain your answer.
- If the null hypothesis was true, could the sample data or more extreme have occurred by sampling variability or is it unlikely to be sampling variability? Explain your answer.
- Write a conclusion for the test addressing evidence and the claim. Explain your conclusion in non-technical language.
- Are the categories related or not? Explain your answer.

28. A random sample of male college students were asked their major. Later, a random sample of female college students were asked their major. The goal of the study was to show that gender is not related to major. Use a 5% significance level and the Statcato printout below to answer the questions given above.

**Chi-Square Test: Contingency Table:**

	Business	English	History	Music	Biology	Math	Total
<b>Female</b>	89.0 (97.30) [0.71]	71.0 (62.45) [1.17]	62.0 (58.58) [0.20]	48.0 (48.89) [0.02]	56.0 (57.12) [0.02]	9.0 (10.65) [0.26]	335.0
<b>Male</b>	112.0 (103.70) [0.67]	58.0 (66.55) [1.10]	59.0 (62.42) [0.19]	53.0 (52.11) [0.02]	62.0 (60.88) [0.02]	13.0 (11.35) [0.24]	357.0
<b>Total</b>	201.0	129.0	121.0	101.0	118.0	22.0	692.0

(expected frequency), [test statistic contribution]

Significance Level	DOF	$\chi^2$	Critical value	p-Value
0.05	5	4.6014	11.0705	0.4664



29. A random sample of adults were asked their blood type and Rh status. (Blood tests were provided for those that did not know their blood type and Rh status.) The goal of the study was to show that blood type is related to Rh status (dependent). Use a 10% significance level and the Statcato printout below to answer the questions given above.

**Chi-Square Test: Contingency Table:**

	Type A	Type B	Type AB	Type O	Total
<b>Rh+</b>	35.0 (36.03) [0.03]	24.0 (23.0) [0.04]	11.0 (16.1) [1.62]	91.0 (85.87) [0.31]	161.0
<b>Rh-</b>	12.0 (10.97) [0.10]	6.0 (7.0) [0.14]	10.0 (4.9) [5.31]	21.0 (26.13) [1.01]	49.0
<b>Total</b>	47.0	30.0	21.0	112.0	210.0

(expected frequency), [test statistic contribution]

Significance Level	DOF	$\chi^2$	Critical value	p-Value
0.10	3	8.5522	6.2514	0.0359

30. A hospital wanted to determine if the age of a patient is not related to what part of the hospital they were in. They took a random sample of patients that have visited their hospital and determined both their age and the part of the hospital. The ages were broken up into age groups. Use a 1% significance level and the Statcato printout below to answer the questions given above.

**Chi-Square Test: Contingency Table:**

	Med/Surg	ICU	SDS	ER	Total
<b>18-35 years old</b>	19.0 (19.12) [7.98 · 10 <sup>-4</sup> ]	4.0 (11.47) [4.87]	25.0 (17.85) [2.87]	16.0 (15.55) [0.01]	64.0
<b>36-49 years old</b>	27.0 (19.42) [2.96]	7.0 (11.65) [1.86]	22.0 (18.13) [0.83]	9.0 (15.80) [2.92]	65.0
<b>50-64 years old</b>	17.0 (18.53) [0.13]	13.0 (11.12) [0.32]	15.0 (17.29) [0.30]	17.0 (15.07) [0.25]	62.0
<b>65+ years old</b>	12.0 (17.93) [1.96]	21.0 (10.76) [9.75]	8.0 (16.73) [4.56]	19.0 (14.58) [1.34]	60.0
<b>Total</b>	75.0	45.0	70.0	61.0	251.0

(expected frequency), [test statistic contribution]

Significance Level	DOF	$\chi^2$	Critical value	p-Value
0.01	9	34.9208	21.666	6.153 · 10 <sup>-5</sup>



31. A random sample of American adults was taken and their health and education status obtained. Test to test the claim that health and education are related. Use a 5% significance level and the Statcato printout below to answer the questions given above.

**Chi-Square Test: Contingency Table:**

	Excellent Health	Good Health	Fair Health	Poor Health	Total
<b>Less Than High School</b>	72.0 (148.64) [39.51]	202.0 (249.76) [9.13]	199.0 (106.91) [79.33]	62.0 (29.70) [35.14]	535.0
<b>High School Diploma</b>	465.0 (502.31) [2.77]	877.0 (844.04) [1.29]	358.0 (361.29) [0.03]	108.0 (100.36) [0.58]	1808.0
<b>Some College / Associates Degree</b>	80.0 (77.24) [0.10]	138.0 (129.78) [0.52]	49.0 (55.55) [0.77]	11.0 (15.43) [1.27]	278.0
<b>Bachelor's Degree</b>	229.0 (161.42) [28.30]	276.0 (271.23) [0.08]	64.0 (116.10) [23.38]	12.0 (32.25) [12.72]	581.0
<b>Graduate Degree</b>	130.0 (86.40) [22.00]	147.0 (145.19) [0.02]	32.0 (62.15) [14.62]	2.0 (17.26) [13.49]	311.0
<b>Total</b>	976.0	1640.0	702.0	195.0	3513.0

(expected frequency), [test statistic contribution]

Significance Level	DOF	$\chi^2$	Critical value	p-Value
0.05	12	285.0610	21.0261	0

(#32-35) Directions: Use StatKey at [www.lock5stat.com](http://www.lock5stat.com) to simulate the following chi-squared categorical association tests. Go to the "More Advanced Randomization Tests" menu at the bottom of the StatKey page. Click on the button that says, " $\chi^2$  Test for Association". Click on "Edit Data" and type in the contingency table provided. Click on "Generate 1000 Samples" a few times to create the simulated sampling distribution and answer the following questions.

- Write the null and alternative hypothesis. Make sure to label which one is the claim.
- Check the assumptions for the categorical association test. Assume the data was collected randomly. Under "Original Sample", click on "Show Details" to see the expected counts.
- Use the formula  $df = (r - 1)(c - 1)$  to calculate the degrees of freedom. "r" is the number of rows and "c" is the number of columns not counting the totals.
- What is the Chi-squared test statistic? Write a sentence to explain the test statistic.
- Put the significance level proportion in the right tail proportion to calculate the critical value. What is the critical value? (Answers will vary slightly.) Does the original sample  $\chi^2$  test statistic fall in the tail determined by the critical value?
- Does the sample data significantly disagree with the null hypothesis? Explain your answer.
- Are the observed counts significantly different from the expected counts? Explain your answer.
- Put the original sample test  $\chi^2$  test statistic in the bottom box in the simulation to calculate the P-value. What is the P-value? (Answers will vary slightly.) Write a sentence to explain the P-value.



- i) Compare the  $P$ -value to the significance level. Should we reject the null hypothesis or fail to reject the null hypothesis? Explain your answer.
- j) If the null hypothesis was true, could the sample data or more extreme have occurred by sampling variability or is it unlikely to be sampling variability? Explain your answer.
- k) Write a conclusion for the test addressing evidence and the claim. Explain your conclusion in non-technical language.
- l) Are the categories related or not? Explain your answer.

32. We want to know if the state a home is built in is related to the size of the home. A random sample of homes in the U.S was taken. Click on "Edit Data" in StatKey and type in the following contingency table. Do not forget to include a space after the commas. Use a 5% significance level and randomized simulation to test the claim that the state is not related to size of the home.

[blank], CA, NJ, NY, PA

Large, 7, 6, 7, 3

Small, 23, 24, 23, 27

33. Open the "Car Data" at [www.matt-teachout.org](http://www.matt-teachout.org). Copy and paste the "Country" and "Cylinders" columns next to each other in a new Excel spreadsheet. Then copy the two columns together. Click on "Edit Data" in StatKey and paste the two columns into StatKey. Use a 1% significance level to test the claim that the country a car is made in is related to the cylinders. Answer the questions above.

34. We want to show that gender is related to getting an award. A random sample of people that won famous awards in the Olympic, Academia, and Nobel was taken and their gender was noted. Click on "Edit Data" in StatKey and type in the following contingency table. Do not forget to include a space after the commas. Use a 10% significance level and randomized simulation to test the claim that awards are related to gender.

[blank], Olympic, Academy, Nobel

Male, 109, 11, 73

Female, 73, 20, 76

35. Open the "Math 140 Fall 2015 Survey Data" at [www.matt-teachout.org](http://www.matt-teachout.org). Copy and paste the "Tattoo" and "Favorite Social Media" columns next to each other in a new Excel spreadsheet. Then copy the two columns together. Click on "Edit Data" in StatKey and paste the two columns into StatKey. Use a 5% significance level to test the claim that having a tattoo or not is not related to social media. Answer the questions above.

---

