

Statistics
Chi-Squared Tests for Homogeneity

1. It's common folk wisdom that drinking cranberry juice can help prevent urinary tract infections in women. In 2001 the *British Medical Journal* reported the results of a Finnish study in which three groups of 50 women were monitored for these infections over 6 months. One group drank cranberry juice daily, another group drank a lactobacillus drink, and the third drank neither of those beverages, serving as a control group. In the control group, 18 women developed at least one infection compared with 20 of those who consumed the lactobacillus drink and only 8 of those who drank cranberry juice. Does this study provide supporting evidence for the value of cranberry juice in warding off urinary tract infections?

- a) Is this a survey, a retrospective study, a prospective study, or an experiment?
- b) Will you test goodness-of-fit, homogeneity, or independence? Why?
- c) Fill in the tables below. You may need to answer the questions to help you determine the percentages for the expected values in the 2nd table.

	Cranberry	Lactobacillus	Neither
Infection(s)			
None			

- o How many women got infection(s)? _____ What overall percent? _____
- o How many women did not get infection(s)? _____ What overall percent? _____

(Use the percents you just calculated to determine the expected number for each category)

	Cranberry		Lactobacillus		Neither	
	Observed	Expected	Observed	Expected	Observed	Expected
Infection(s)						
None						

- d) Determine the degrees of freedom: _____
- e) Determine the P-value: _____ (Hint: L₁ is Observed and L₂ is Expected)
- f) What does that P-value tell you?

e) Create a table with the standardized residuals to look closer at the calculations
(Hint: have the calculator store it into L₃)

	Cranberry	Lactobacillus	Neither
Infection(s)			
None			

Standardize Residuals

$$c = \frac{Obs - Exp}{\sqrt{Exp}}$$

f) What do you recognize from the table?

2. In July 1991 and again in April 2001, the Gallup Poll asked random samples of 1015 adults about their opinions on working parents. The table summarizes responses to the question, "Considering the needs of both parents and children, which of the following do you see as the ideal family in today's society?"

	1991	2001
Both work full time	142	131
One works full time, other part time	274	244
One works, other works at home	152	173
One works, other stays home for kids	396	416
No opinion	51	51

	1991		2001	
	Observed	Expected	Observed	Expected
Both work full time	142		131	
One works full time, other part time	274		244	
One works, other works at home	152		173	
One works, other stays home for kids	396		416	
No opinion	51		51	

- Is this a survey, a retrospective study, a prospective study, or an experiment?
 - Will you use a goodness-of fit, homogeneity, or independence? Explain.
 - Fill in the Observed and Expected table above.
 - Determine the P-value and Chi-squared value.
 - Based on these results, do you think there was a change in people's attitudes during the 10 years between these polls?
3. Two different professors teach an introductory Statistics course. The table shows the distribution of final grades they reported. We wonder whether one of these professors is an "easier" grade.

	Prof. Alpha	Prof. Beta
A	3	9
B	11	12
C	14	8
D	9	2
F	3	1

	Prof. Alpha		Prof. Beta	
	Observed	Expected	Observed	Expected
A	3		9	
B	11		12	
C	14		8	
D	9		2	
F	3		1	

- Will you test a goodness-of-fit, homogeneity, or independence? Why?
- Write appropriate Hypotheses
- Find the expected counts for each cell, and explain why the chi-square procedures are not appropriate for this table.

4. Some people believe that a full moon elicits unusual behavior in people. The table shows the number of arrests made in a small town during weeks of six full moons and six other randomly selected weeks during the same year. We wonder if there is evidence of a difference in types of illegal activity that take place.

	Full Moon	Not Full
Violent (murder, assault, rape, etc.)	2	3
Property (burglary, vandalism, etc.)	17	21
Drugs/Alcohol	27	19
Domestic Abuse	11	14
Other Offenses	9	6

	Full Moon		Not Full	
	Observed	Expected	Observed	Expected
Violent (murder, assault, rape, etc.)	2		3	
Property (burglary, vandalism, etc.)	17		21	
Drugs/Alcohol	27		19	
Domestic Abuse	11		14	
Other Offenses	9		6	

- Will you test goodness-of-fit, homogeneity, or independence? Why?
- Write appropriate Hypotheses
- Find the expected counts for each cell, and explain why the chi-square procedures are not appropriate for this table.

5. In the problem above you found that you could not proceed because it did not satisfy the expected count condition. Determine a sensible way to combine some cells that will make the expected counts acceptable.

	Full Moon	Not Full

	Full Moon		Not Full	
	Observed	Expected	Observed	Expected

- Complete both tables
- Test a hypothesis about the full moon and state your conclusion
- Create a table with the standardized residuals to look closer at the calculations
(Hint: have the calculator store it into L_3)

	Full Moon	Not Full

Standardize Residuals

$$c = \frac{Obs - Exp}{\sqrt{Exp}}$$

- What do you recognize from the table?