

# Chapter 11

## Tests of Homogeneity and Independence

### 11.1 Study Suggestions

Chapter 11 presents the chi-squared test of homogeneity (or of independence) and gives a few examples of its use. Following the approach of Chapter 7, emphasis is placed on the type of study, and it is argued that the chi-squared test of homogeneity is appropriate in three different situations.

You should understand the rationale behind the formula for the expected count for any cell of the table, namely the product of the row total and the column total divided by the sample size. You also should understand why

$$\frac{(O - E)^2}{E}$$

is a reasonable way to measure the amount of evidence in a particular cell in support of the alternative hypothesis (the argument appeared in Chapter 10). The computation of the test statistic for the chi-squared test of homogeneity is too tedious to perform by hand. I strongly recommend the use of a computer!

Make sure you can read computer output of the type given in the text for the test of homogeneity, and that, given the value of the test statistic, you can find the P-value for the test of homogeneity from the table of the chi-squared distributions.

If the study has three or more treatments or populations, and the test of homogeneity yields a P-value less than or equal to 0.05, then you must understand that you only can conclude that the treatments (populations) are not identical. You cannot conclude which treatments (populations) differ without further testing. The text provides one strategy for testing,

but many statisticians believe this problem requires a more sophisticated approach than I advocate.

### 11.2 Solutions to Odd-Numbered Exercises

5. (c)  $E = 132(178)/449 = 52.3$ .  
(d)

$$\frac{(O - E)^2}{E} = \frac{(71 - 52.3)^2}{52.3} = 6.69.$$

- (f) For a  $3 \times 2$  contingency table, the number of degrees of freedom of the reference chi-squared curve is  $2(1) = 2$ . The approximate P-value is the area to the right of 35.063 under the chi-squared curve with two degrees of freedom. This area is smaller than 0.01. The data are highly statistically significant, and the researcher should reject the null hypothesis that the three populations are identical. Next, the researcher will want to investigate where the differences lie and can do this by making pairwise comparisons of the populations.
- (g) First, I will compare the population corresponding to no injections with the population corresponding to an average of 1–100 injections per month. The standardized value of the test statistic for Fisher's test is

$$z = \frac{\sqrt{316}[9(144) - 66(98)]}{\sqrt{75(242)(107)(210)}} = -4.56.$$

The approximate P-value for the third alternative is twice the area to the right of 4.56 under the standard normal curve. This area is smaller than 0.0004. The difference between the two populations is highly statistically significant and the researcher can conclude that there is a real difference between them.

Next, I will compare the population corresponding to no injections with the population corresponding to an average of more than 100 injections per month. The standardized value of the test statistic for Fisher's test is

$$z = \frac{\sqrt{206}[9(61) - 66(71)]}{\sqrt{75(132)(80)(127)}} = -5.93.$$

The approximate P-value for the third alternative is twice the area to the right of 5.93 under the standard normal curve. This area is smaller than 0.0004. The difference between the two populations is highly statistically significant and the researcher can conclude that there is a real difference between them.

Finally, I will compare the population corresponding to an average of 1–100 injections per month with the population corresponding to an average of more than 100 injections per month. The standardized value of the test statistic for Fisher's test is

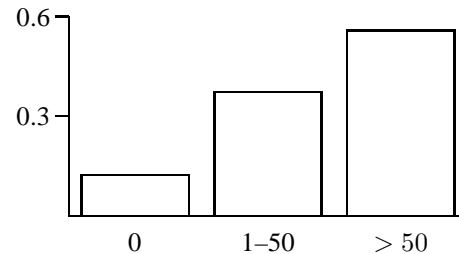
$$z = \frac{\sqrt{373}[98(61) - 144(71)]}{\sqrt{242(132)(169)(205)}} = -2.47.$$

The approximate P-value for the third alternative is twice the area to the right of 2.47 under the standard normal curve. This area equals  $2(0.0068) = 0.0136$ . The difference between the two populations is statistically significant, but following the cautious advice given in the text, I am reluctant to conclude that the difference between the populations is real.

In summary, the no injections group had a significantly lower rate of HIV infection either of the injection groups. Among

those who inject drugs, the group with more than 100 injections per month may have a higher rate of HIV infection than the group with fewer than 100 injections per month, but the data are subject to interpretation.

7. (a) The bar chart of the proportions of HIV positive subjects is below.



(c)  $E = 158(273)/451 = 95.6$ .

(d)

$$\frac{(O - E)^2}{E} = \frac{(70 - 95.6)^2}{95.6} = 6.86.$$

- (b) For a  $3 \times 2$  contingency table, the number of degrees of freedom of the reference chi-squared curve is  $2(1) = 2$ . The approximate P-value is the area to the right of 41.590 under the chi-squared curve with two degrees of freedom. This area is smaller than 0.01. The data are highly statistically significant and the researcher should reject the null hypothesis that the three populations are identical. Next, the researcher will want to investigate where the differences lie and can do this by making pairwise comparisons of the populations.

- (c) First, I will compare the population corresponding to no IV drug use with the population corresponding to a duration of 1–50 months use. The standardized value of the test statistic for Fisher's test is

$$z = \frac{\sqrt{292}[9(137) - 66(81)]}{\sqrt{75(218)(90)(203)}} = -4.07.$$

The approximate P-value for the third alternative is twice the area to the right

of 4.07 under the standard normal curve. This area is smaller than 0.0004. The difference between the two populations is highly statistically significant, and the researcher can conclude that there is a real difference between them.

Next, I will compare the population corresponding to no IV drug use with the population corresponding to a duration of more than 50 months use. The standardized value of the test statistic for Fisher's test is

$$z = \frac{\sqrt{232}[9(70) - 66(88)]}{\sqrt{75(158)(97)(136)}} = -6.32.$$

The approximate P-value for the third alternative is twice the area to the right of 6.32 under the standard normal curve. This area is smaller than 0.0004. The difference between the two populations is highly statistically significant, and the researcher can conclude that there is a real difference between them.

Finally, I will compare the population corresponding to a duration of 1–50 months of IV drug use with the population corresponding to a duration of use of more than 50 months use. The standardized value of the test statistic for Fisher's test is

$$z = \frac{\sqrt{375}[81(70) - 137(88)]}{\sqrt{218(158)(169)(207)}} = -3.57.$$

The approximate P-value for the third alternative is twice the area to the right of 3.57 under the standard normal curve. This area equals  $2(0.0002) = 0.0004$ . The difference between the two populations is highly statistically significant, and the researcher can conclude that there is a real difference between them.

9. For a  $3 \times 2$  contingency table, the number of degrees of freedom of the reference chi-squared curve is  $2(1) = 2$ . The approximate P-value is the area to the right of 5.548 under the chi-squared curve with two degrees of freedom. This area is between 0.05 and 0.10. The data are

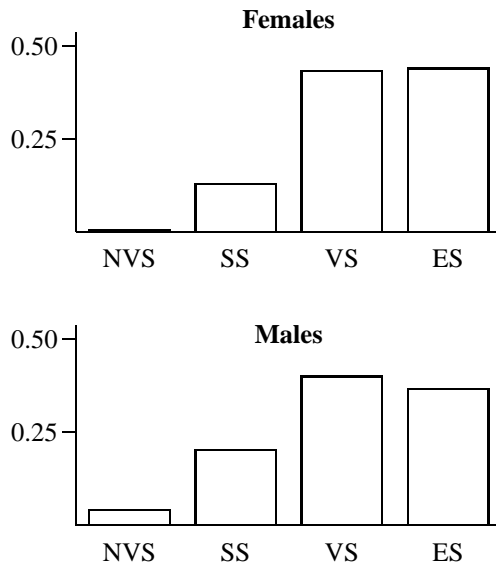
not statistically significant, and the researcher should not reject the null hypothesis that the three treatments are identical. The researcher should not perform pairwise comparisons of the treatments because the conclusion of the test is that no significant differences exist.

11. For a  $3 \times 2$  contingency table, the number of degrees of freedom of the reference chi-squared curve is  $2(1) = 2$ . The approximate P-value is the area to the right of 14.210 under the chi-squared curve with two degrees of freedom. This area is smaller than 0.01. The data are highly statistically significant, and the researcher should reject the null hypothesis that the three treatments are identical. The researcher should perform pairwise comparisons of the treatments to investigate which treatments differ.
13. The table of row proportions is below. In order to be able to fit the table into this column, I have used several abbreviations:

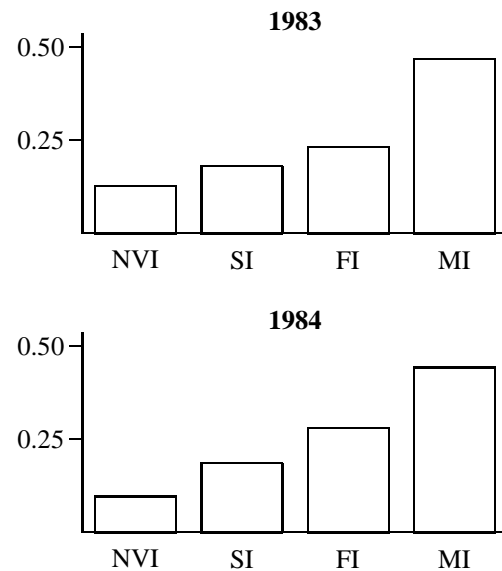
- G: Gender
- NVS: Not very serious
- SS: Somewhat serious
- VS: Very serious
- ES: Extremely serious
- T: Total
- F: Female
- M: Male

		<u>Attitude</u>			
G	NVS	SS	VS	ES	T
F	0.00	0.13	0.43	0.44	1.00
M	0.04	0.20	0.40	0.36	1.00

The bar charts for females and males are below.



The bar charts for 1983 and 1984 are below.



The table of row proportions and the bar charts both reveal that women view drunk driving to be a more serious problem than do men.

For a  $2 \times 4$  contingency table, the number of degrees of freedom of the reference chi-squared curve is  $1(3) = 3$ . The approximate P-value is the area to the right of 62.907 under the chi-squared curve with three degrees of freedom. This area is smaller than 0.01. The data are highly statistically significant and the researcher should reject the null hypothesis that the two populations are identical.

15. The table of row proportions is below. In order to be able to fit the table into this column, I have used several abbreviations:

- NVI: Not very important
- SI: Somewhat important
- FI: Fairly important
- MI: Most important
- T: Total

Year	Attitude				T
	NVI	SI	FI	MI	
1983	0.13	0.18	0.23	0.47	1.00
1984	0.10	0.18	0.28	0.44	1.00

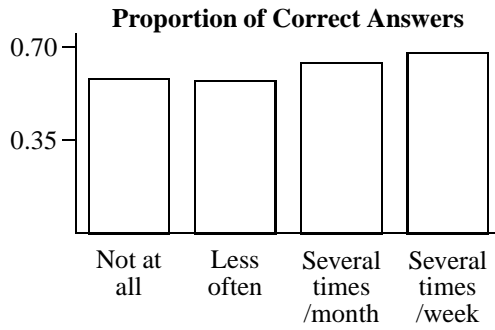
The table of row proportions and the bar charts both reveal that the responses were more extreme in 1983 than in 1984, but the difference is small.

For a  $2 \times 4$  contingency table, the number of degrees of freedom of the reference chi-squared curve is  $1(3) = 3$ . The approximate P-value is the area to the right of 13.001 under the chi-squared curve with three degrees of freedom. This area is smaller than 0.01. The data are highly statistically significant and the researcher should reject the null hypothesis that the two populations are identical.

17. The proportions of correct answers for the four drinking frequency groups are in the table below.

Drinking Frequency	Proportion Correct
Not at all	0.577
Less often	0.570
Several times per month	0.637
Several times per week	0.676

A bar chart of the four proportions of correct answers is below.



The above table and picture reveal that the two groups with the lowest drinking frequency have approximately the same level of knowledge. Those who drink several times per month have a substantially higher level of knowledge than either of these groups, and those who drink several times per week have the highest level of knowledge.

For a  $4 \times 2$  contingency table, the number of degrees of freedom of the reference chi-squared curve is  $3(1) = 3$ . The approximate P-value is the area to the right of 16.745 under the chi-squared curve with three degrees of freedom. This area is smaller than 0.01. The data are highly statistically significant and the researcher should reject the null hypothesis that the four populations are identical. Next, the researcher will want to investigate where the differences lie and can do this by making pairwise comparisons of the populations.

Each pairwise comparison involves performing Fisher's test. For simplicity of exposition,

- Group 1 denotes "not at all."
- Group 2 denotes "less often."
- Group 3 denotes "several times per month."
- Group 4 denotes "several times per week."

For the comparison of groups 1 and 2, the standardized value of the test statistic for Fisher's test is

$$z = \frac{\sqrt{1,471}[290(417) - 213(552)]}{\sqrt{503(969)(842)(630)}} = 0.25.$$

The approximate P-value for the third alternative is twice the area to the right of 0.25 under the standard normal curve. This area equals  $2(0.4013) = 0.8026$ .

For the comparison of groups 1 and 3, the standardized value of the test statistic for Fisher's test is

$$z = \frac{\sqrt{1,132}[290(229) - 213(401)]}{\sqrt{503(630)(691)(442)}} = -2.06.$$

The approximate P-value for the third alternative is twice the area to the right of 2.06 under the standard normal curve. This area equals  $2(0.0197) = 0.0394$ .

For the comparison of groups 1 and 4, the standardized value of the test statistic for Fisher's test is

$$z = \frac{\sqrt{851}[290(113) - 213(236)]}{\sqrt{503(349)(526)(326)}} = -2.94.$$

The approximate P-value for the third alternative is twice the area to the right of 2.94 under the standard normal curve. This area equals  $2(0.0016) = 0.0032$ .

For the comparison of groups 2 and 3, the standardized value of the test statistic for Fisher's test is

$$z = \frac{\sqrt{1,598}[552(229) - 417(401)]}{\sqrt{969(630)(953)(646)}} = -2.66.$$

The approximate P-value for the third alternative is twice the area to the right of 2.66 under the standard normal curve. This area equals  $2(0.0039) = 0.0078$ .

For the comparison of groups 2 and 4, the standardized value of the test statistic for Fisher's test is

$$z = \frac{\sqrt{1,317}[552(113) - 417(236)]}{\sqrt{969(349)(788)(530)}} = -3.48.$$

The approximate P-value for the third alternative is twice the area to the right of 3.48 under the standard normal curve. This area equals  $2(0.0003) = 0.0006$ .

Finally, for the comparison of groups 3 and 4, the standardized value of the test statistic for Fisher's test is

$$z = \frac{\sqrt{976}[401(113) - 229(236)]}{\sqrt{630(349)(637)(342)}} = -1.25.$$

The approximate P-value for the third alternative is twice the area to the right of 1.25 under the standard normal curve. This area equals  $2(0.1056) = 0.2112$ .

To summarize, the differences between groups 1 and 2, groups 1 and 3, and groups 3 and 4 reasonably can be attributed to chance, although one might argue with my conclusion for groups 1 and 3 since the approximate P-value is 0.0394. I conclude that group 4 had a higher level of knowledge than either groups 1 or 2, and that group 3 had a higher level of knowledge than group 2.

### 11.3 Exam Questions

1. A CRD is performed to compare four treatments. There are three possible responses: low, medium, and high. Suppose the sample data are presented in a contingency table as described in Chapter 11 of the text; what are the dimensions of the table?
2. Consider the following  $2 \times 3$  contingency table.

Treatment				Total
1	20	20	10	50
2	40	60	50	150
Total	60	80	60	200

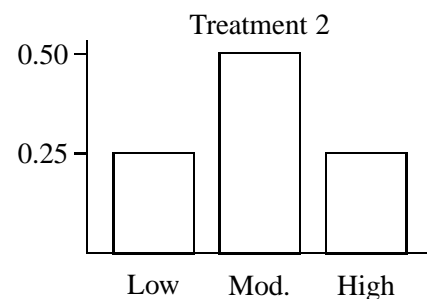
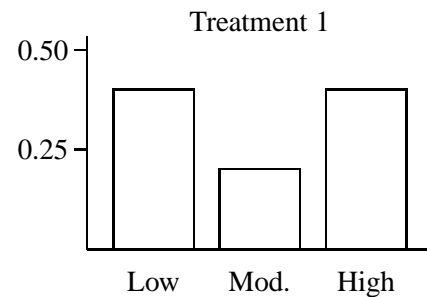
For the test of homogeneity, what is the expected count ( $E$ ) for the cell with observed count equal to 40?

3. A CRD with three treatments yields the following data.

Treatment	S	F	Total
1	116	84	200
2	75	75	150
3	42	58	100
Total	233	217	450

Compute the value of  $(O - E)^2/E$  for the cell for treatment 1, failure.

4. A CRD is performed with two treatments and a three-category response. The response is ordered with categories low, moderate, and high. Below are the bar charts of the response by treatment.



Which of the following statements best describes the pattern revealed in these bar charts?

- (a) Treatment 1 tends to give larger responses than treatment 2.
  - (b) Treatment 1 tends to give smaller responses than treatment 2.
  - (c) Treatment 1 tends to give more extreme responses than treatment 2.
  - (d) Treatment 1 tends to give less extreme responses than treatment 2.
5. A five-category response is obtained from each subject in a four-treatment CRD. The value of the test statistic for the chi-squared test of homogeneity equals 27.00.
    - (a) Compute the approximate P-value for the chi-squared test of homogeneity.

- (b) Should the analyst perform any additional chi-squared tests of homogeneity on these data? Explain your answer.
6. A CRD is performed with six treatments, and the response variable has four categories. The value of the test statistic for the chi-squared test of homogeneity equals 21.00.
- (a) Compute the approximate P-value for the chi-squared test of homogeneity.
- (b) Should the analyst perform any additional chi-squared tests of homogeneity on these data? Explain your answer.
7. Teri selects independent random samples from three populations and obtains the value of a dichotomous response from each subject in her study. The test of homogeneity yields an approximate P-value of 0.0000. Her sample proportions of successes are

$$\hat{p}_1 = 0.70, \hat{p}_2 = 0.68, \text{ and } \hat{p}_3 = 0.40.$$

Pairwise comparisons of the populations (Fisher's test for the third alternative) yield the approximate P-values displayed in the table below.

Populations Compared	Approximate P-values
1 and 2	0.7592
1 and 3	0.0000
2 and 3	0.0001

Briefly summarize the conclusions Teri should draw from the test of homogeneity and the pairwise comparisons.

8. Sandy selects independent random samples from three populations and obtains the value of a dichotomous response from each subject in her study. The test of homogeneity yields an approximate P-value of 0.0155. Her sample proportions of successes are

$$\hat{p}_1 = 0.70, \hat{p}_2 = 0.60, \text{ and } \hat{p}_3 = 0.50.$$

Pairwise comparisons of the populations (Fisher's test for the third alternative) yield the approximate P-values displayed in the table below.

Populations Compared	Approximate P-values
1 and 2	0.1382
1 and 3	0.0039
2 and 3	0.1552

Briefly summarize the conclusions Sandy should draw from the test of homogeneity and the pairwise comparisons.

9. Rebecca selects independent random samples from five populations and obtains the value of a dichotomous response from each subject in her study. The test of homogeneity yields an approximate P-value of 0.0916. Her sample proportions of successes are

$$\hat{p}_1 = 0.60, \hat{p}_2 = \hat{p}_3 = \hat{p}_4 = 0.50, \text{ and } \hat{p}_5 = 0.40.$$

Pairwise comparisons of the populations (Fisher's test for the third alternative) yield an approximate P-value of 0.0046 for the comparison of populations 1 and 5. Each other comparisons yields a P-value that is larger than 0.05.

Briefly summarize the conclusions Rebecca should draw from the test of homogeneity and the pairwise comparisons.

## 11.4 Solutions to Exam Questions

- $4 \times 3$ .
- $E = (150(60))/200 = 45$ .
- 1.605.
- (c).
- (a)  $\mathbf{P} < 0.01$ .  
(b) Yes. The analyst has concluded that the four treatments do not have an identical effect, but does not know which ones differ.
- (a)  $0.10 < \mathbf{P} < 0.50$ .  
(b) No. The analyst has concluded that there is insufficient evidence to reject the notion that the treatments are identical. Hence,

there is no reason to continue to look for differences.

7. Teri should reject the null hypothesis that the three populations have the same proportion of successes. The pairwise comparisons allow Teri to conclude populations 1 and 2 are both better than population 3, but there is insufficient evidence to conclude that a difference exists between populations 1 and 2.
8. Sandy should reject the null hypothesis that the M three populations have the same proportion of successes. The pairwise comparisons allow Sandy to conclude that population 1 is better than population 3, but there is insufficient evidence to conclude that a difference exists between populations 1 and 2 or populations 2 and 3.
9. Rebecca should not reject the null hypothesis that the the five populations have the same proportion of successes. Thus, she should not perform pairwise comparisons of the populations.