

Chi-squared tests for 2-way tables

Stat 145

Chi-squared tests for 2-way tables

There is another situation that uses the same Chi-squared test statistic. If we have two categorical variables we can display a summary of the data in a 2-way table.

For example, here is a summary of a study that surveyed men who had played elite level soccer, had played soccer but not at the elite level, or had not played soccer. They were asked whether they had been diagnosed with arthritis in the hip or knee.

	elite	non-elite	no soccer
arthritis	10	9	24
no arthritis	61	206	548

We can enter this summary table into R as follows.¹

```
arthritis <- rbind(c(10, 9, 24), c(61, 206, 548)) # r for row-wise
arthritis

##      [,1] [,2] [,3]
## [1,]  10   9  24
## [2,]  61 206 548
```

We can use the same test statistic as for the goodness of fit test, but we need to adjust

- the **null hypothesis** – the null hypothesis in this situation is one of the following:
 - there is no association between the column variable and the row variable (they are independent), or
 - the column-proportions are the same in each row
 - the row-proportions are the same in each column
- how we compute the **expected counts**
- how we determine the **degrees of freedom**

¹If you want to be fancy, you can add row and column labels. You can also use `pander()` to print the table fancier in your document.

```
rownames(arthritis) <- c("arthritis", "no arthritis")
colnames(arthritis) <- c("elite", "non-elite", "no soccer")
library(pander)
arthritis %>% pander()
```

	elite	non-elite	no soccer
arthritis	10	9	24
no arthritis	61	206	548

Null Hypothesis

In our example, the null hypothesis is that having arthritis is independent of the level of soccer someone played. We could also express this by saying that the proportion of people with arthritis is the same for elite soccer players, non-elite soccer players, and non-soccer players. (So it is like a 3-proportion test.)

Expected Counts

Our null hypothesis doesn't say just what the proportions should be in each cell, only that the proportion of people that have arthritis should be the same in each of three columns. In other words, we should get the cell proportion by multiplying the row proportion by the column proportion.

Let's begin by adding row and column totals to our table.

	elite	non-elite	no soccer	total
arthritis	10	9	24	43
no arthritis	61	206	548	815
total	71	215	572	858

For the top left cell, we can compute row and column proportions using these totals:

```
# row proportion
43/858
```

```
## [1] 0.0501
```

```
# column proportion
71/858
```

```
## [1] 0.0828
```

From this we can get the expected proportion in the top left cell:

```
# expected proportion in top left cell
43/858 * 71/858
```

```
## [1] 0.00415
```

To get the expected count, we need to multiply the expected proportion by sample size:

```
# expected count in top left cell
43/858 * 71/858 * 858
```

```
## [1] 3.56
```

```
# this should be the same
43 * 71 / 858
```

```
## [1] 3.56
```

So we have

$$\text{expected count} = \frac{\text{row total} \cdot \text{column total}}{\text{grand total}}$$

Degrees of Freedom

The degrees of freedom is given by

$$\text{degrees of freedom} = (\text{number of rows} - 1)(\text{number of columns} - 1)$$

In our example, that would be $(2 - 1)(3 - 1) = 1 \cdot 2 = 2$ degrees of freedom. (If you like to think about this visually: Cross off one row and one column and count how many cells remain.)

Finishing up

We can now compute the p-value for this example using these steps:

- determine the expected count in each cell
- compute the chi-squared test statistic
- use `pchisq()` to get the p-value
- interpret the p-value

Of course, we can use `chisq.test()` or `xchisq.test()` to automate the whole thing.

```
xchisq.test(arthritis)

## Warning in chisq.test(x = x, y = y, correct = correct, p = p, rescale.p =
## rescale.p, : Chi-squared approximation may be incorrect

##
## Pearson's Chi-squared test
##
## data:  x
## X-squared = 13, df = 2, p-value = 0.001
##
##   10.00    9.00   24.00
## (  3.56) ( 10.78) ( 28.67)
## [11.662] [ 0.292] [ 0.760]
## < 3.41> <-0.54> <-0.87>
##
##   61.00   206.00  548.00
## ( 67.44) (204.22) (543.33)
## [ 0.615] [ 0.015] [ 0.040]
## <-0.78> < 0.12> < 0.20>
##
## key:
## observed
## (expected)
## [contribution to X-squared]
## <Pearson residual>
```

Notice that the expected cell count in the top left cell matches what we calculated previously.

What is that warning? R is warning us that the chi-squared distribution might not be a very good approximation in this situation. The small expected count in the top left cell ($3.56 < 5$) is triggering the warning. Ideally, we'd like these expected counts to be at least 5.

When we get that warning, we should do a randomization test instead.

```
# B = number of replicates -- not sure why they called it B
chisq.test(arthritis, simulate.p.value = TRUE, B = 5000)

##
## Pearson's Chi-squared test with simulated p-value (based on 5000
## replicates)
##
## data:  arthritis
## X-squared = 13, df = NA, p-value = 0.002
```

In this case, the p-value changes a bit, but the conclusion remains the same. We can reject the null hypothesis. It appears that the prevalence of arthritis is not the same for our three groups.

How do the groups differ? The Chi-squared test indicates that the groups don't all have the same rate of arthritis, but how do they differ?

If we look at the contributions to Chi-squared, we see that the upper left cell is contributing almost all of it. This is because we expected only 3.5 cases there and observed 10. The other cells have roughly what we would expect. This is an indication that elite soccer players are more likely to have arthritis than people in the other groups are.

We can also compute group-wise proportions and compare them.

```
c(10/71, 9/215, 24/572) %>% round(3)
```

```
## [1] 0.141 0.042 0.042
```

As we see, the arthritis rate for elite soccer players is much higher than for the other two groups.

When we have the raw data

In the example above, we have used summary tables rather than raw data. Now let's do an example using raw data. We can use `df_stats()` or `tally()` to create the tables as before, or we can just give the formula to `xchisq.test()` (but not to `chisq.test()`).

Avian malaria

In an experiment to see if laying eggs makes birds more susceptible to malaria, researchers found 65 great tit nests and randomly selected some for the removal of two eggs. This causes the female to lay an additional egg – perhaps at the cost of being less resistant to malaria.



Fourteen days after the eggs had hatched, blood samples were taken to test for malaria in the mother birds. The data are available in `GreatTitMalaria` in the `abd` package.

```
library(abd)      # analysis of biological data
library(pander)  # to pretty-print the table
names(GreatTitMalaria)
```

```
## [1] "treatment" "response"
```

```
tally(response ~ treatment, data = GreatTitMalaria) %>% pander()
```

	Control	Egg removal
Malaria	7	15
No Malaria	28	15

1. Use `chisq.test()` or `xchisq.test()` to assess the data. What is the null hypothesis? What do you conclude?
2. In this situation, we could do this another way (not using a Chi-squared test at all).
 - a. Do it the other way.

- b. In what situations will we have these two options? Does one method have any advantages over the other?

More Examples

Smoking and Diet

To test for a potential confounding variable in a study of the health effects of a “Mediterranean diet”, researchers looked to see if there was an association between diet and smoking. Diet was categorized as low, medium, or high. Here is their data in tabular form.

	low med. diet	medium med. diet	high med. diet
never smoked	2516	2920	2417
former smoker	3657	4653	3449
current smoker	2012	1627	1294

3. What are the degrees of freedom for this test? Why?
4. What should the researchers conclude?

Physicians Health Study

The Physicians Health Study is a famous example of a prospective, double-blind randomized clinical trial. In one part of the study, doctors were given either aspirin or a placebo to take daily to see how that would affect the rate of heart attacks. Over 22,000 male doctors participated in this part of the study.²

5. What does it mean that the study was “randomized”?
6. What does it mean that the study was double blind?
7. What does it mean that the study was prospective?
8. Why do you think they used doctors? Why only males?
9. Why so many doctors?

After a number of years in the study, here were the results

	heart attack	no heart attack
aspirin	104	10933
placebo	189	10845

10. What should we conclude from this study?

Diabetes

An experiment evaluating three treatments for Type 2 Diabetes in patients aged 10–17 who were being treated with metformin is summarized in the table below. The three treatments considered were continued treatment with metformin (met), treatment with metformin combined with rosiglitazone (rosi), or a lifestyle intervention program. Each patient had a primary outcome, which was either lacked glycemic control (failure) or did not lack that control (success).³

²Professor Pruim’s father-in-law was a subject in this study.

³This example can be found in section 6.3.2 of *IMS*.

	failure	success	total
lifestyle	109	125	234
met	120	112	232
rosi	90	143	233
total	319	380	699

11. What are appropriate hypotheses for this test?
12. What are the degrees of freedom for this test if we use the theoretical method?
13. Compute the expected cell count and the contribution to the Chi-squared statistic for the lifestyle-failure cell of the table.
14. Now use `xchisq.test()` to test your hypothesis and to check your answers to the previous two questions. What conclusion can be drawn from this study?

Flycatchers

In northern Europe there are two species of flycatcher (a bird): collared and pied. Sometimes a male from one species will mate with a female from the other species. Researchers were interested in comparing the sex ratio for hybrid offspring vs “purebred” offspring.



collared flycatcher



pied flycatcher

Here is their data in table form:

	male	female
hybrid	16	10
purebred	72	73

15. Compute the proportion of offspring that are female in each mating type.
16. Do these data provide evidence that the sex ratios differ?
17. Construct a confidence interval for the difference in these proportions. How does this compare to your hypothesis test?

Aspirin and Cancer

A similar study (with women this time), investigated whether regular taking of aspirin has an affect on cancer rates. After the women had taken aspirin or placebo for ten years, researchers checked to see how many had been diagnosed with cancer. Here are the results.

	cancer	no cancer
aspirin	1438	18496
placebo	1427	18515

18. What should we conclude from this study? If the results are statistically significant, also compute a 95% confidence interval for the difference in the proportions of subjects who had cancer in each treatment group.