

12.7: The Fisher Exact Test

What should you do if your cell counts are too small, but you'd still like to test the null hypothesis that the two variables are independent? One answer would be "collect more data", but that's far too glib: there are a lot of situations in which it would be either infeasible or unethical to do that. If so, statisticians have a kind of moral obligation to provide scientists with better tests. In this instance, Fisher (1922) kindly provided the right answer to the question. To illustrate the basic idea, let's suppose that we're analysing data from a field experiment, looking at the emotional status of people who have been accused of witchcraft; some of whom are currently being burned at the stake.¹⁸¹ Unfortunately for the scientist (but rather fortunately for the general populace), it's actually quite hard to find people in the process of being set on fire, so the cell counts are awfully small in some cases. The `saalem.Rdata` file illustrates the point:

```
load("./rbook-master/data/saalem.Rdata")
```

```
saalem.tabs <- table( trial )  
print( saalem.tabs )
```

```
##           on.fire  
## happy  FALSE TRUE  
## FALSE      3    3  
## TRUE     10    0
```

Looking at this data, you'd be hard pressed not to suspect that people not on fire are more likely to be happy than people on fire. However, the chi-square test makes this very hard to test because of the small sample size. If I try to do so, R gives me a warning message:

```
chisq.test( saalem.tabs )
```

```
## Warning in chisq.test(saalem.tabs): Chi-squared approximation may be  
## incorrect
```

```
##  
## Pearson's Chi-squared test with Yates' continuity correction  
##  
## data:  saalem.tabs  
## X-squared = 3.3094, df = 1, p-value = 0.06888
```

Speaking as someone who doesn't want to be set on fire, I'd *really* like to be able to get a better answer than this. This is where **Fisher's exact test** comes in very handy.

The Fisher exact test works somewhat differently to the chi-square test (or in fact any of the other hypothesis tests that I talk about in this book) insofar as it doesn't have a test statistic; it calculates the p-value "directly". I'll explain the basics of how the test works for a 2×2 contingency table, though the test works fine for larger tables. As before, let's have some notation:

	Happy	Sad	Total
Set on fire	O_{11}	O_{12}	R_1
Not set on fire	O_{21}	O_{22}	R_2
Total	C_1	C_2	N

In order to construct the test Fisher treats both the row and column totals (R_1 , R_2 , C_1 and C_2) are known, fixed quantities; and then calculates the probability that we would have obtained the observed frequencies that we did (O_{11} , O_{12} , O_{21} and O_{22}) given those

totals. In the notation that we developed in Chapter 9 this is written:

$$P(O_{11}, O_{12}, O_{21}, O_{22} \mid R_1, R_2, C_1, C_2)$$

and as you might imagine, it's a slightly tricky exercise to figure out what this probability is, but it turns out that this probability is described by a distribution known as the *hypergeometric distribution*.¹⁸² Now that we know this, what we have to do to calculate our p-value is calculate the probability of observing this particular table *or a table that is "more extreme"*.¹⁸³ Back in the 1920s, computing this sum was daunting even in the simplest of situations, but these days it's pretty easy as long as the tables aren't too big and the sample size isn't too large. The conceptually tricky issue is to figure out what it means to say that one contingency table is more "extreme" than another. The easiest solution is to say that the table with the lowest probability is the most extreme. This then gives us the p-value.

The implementation of the test in R is via the `fisher.test()` function. Here's how it is used:

```
fisher.test( salem.tabs )
```

```
##
## Fisher's Exact Test for Count Data
##
## data: salem.tabs
## p-value = 0.03571
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.000000 1.202913
## sample estimates:
## odds ratio
##           0
```

This is a bit more output than we got from some of our earlier tests. The main thing we're interested in here is the p-value, which in this case is small enough ($p=.036$) to justify rejecting the null hypothesis that people on fire are just as happy as people not on fire.

This page titled [12.7: The Fisher Exact Test](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Danielle Navarro](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.