

Fisher's Exact Test

Fisher's Exact Test is useful tool for grouped count data and when the samples are particularly small. This test it is generally used on 2x2 tables. However, it can be extended to an $r \times c$ table. A lot of times Pearson's χ^2 is used for this type of analysis but when the assumptions for sample size and cell counts are not met then that approach is not acceptable. However, the Fisher's Exact Test makes the assumption that the margins are fixed fixed values and are not random. Take for example the classic example where the data was collected by Fisher himself and is seen in Table 1. The goal of Fisher's experiment was to determine if someone could determine whether milk or tea was poured first into a cup. To prove this theory Fisher presented eight randomized cups of tea. Four cups had milk poured first and four cups had tea poured first. The data is known as the Lady Tasting Tea.

Table 1: Lady Tasting Tea

		Guess Poured First		Total
		Milk	Tea	
Poured First	Milk	3	1	4
	Tea	1	3	4
	Total	4	4	8

Fisher's Exact Tests uses the hypergeometric distribution and does not rely on approximations. The probability density given the data can be calculated as seen in Equation 1. To determine the probability that the data is that extreme or greater we need the cumulative density. The only table that is more extreme is has $n_{11} = 4$. In Equation 2 we find the probability density of the only table more extreme than the observed table. Therefore, the p-value for this test is $0.22857 + 0.01429 = 0.24286$. This result does not establish any association between the guess on what was poured first and what actually was poured first. It is difficult to determine an association with a sample this small.

$$\frac{\binom{4}{3} \binom{4}{1}}{\binom{8}{4}} = 0.22857 \quad (1)$$

$$\frac{\binom{4}{3} \binom{4}{1}}{\binom{8}{4}} = 0.01429 \quad (2)$$

This example can also be done in the R environment using some simple commands. In this example both Fisher's Exact Test and the χ^2 test are provided so that side-by-side comparisons can be made.

```
tea <- matrix(c(3,1,1,3),
              nrow = 2,
              dimnames = list(Truth=c("Milk","Tea"),
                              Guess=c("Milk","Tea")));
chisq.test(tea);
fisher.test(tea, alternative = "greater");
```