Contents

R> Fisher's Exact Test

The command `fisher.test()` can be used to perform the chi-squared test in R. The function `table()` is useful for preparing the contingency table input.

```
> d <- read.csv("titanic.csv", header=T)
> d$Two.Class <- rep(T, length(d$PClass))
> d$Two.Class <- d$PClass=="3rd"
> fisher.test(table(d[,c("Two.Class","Survived")]))

          Fisher's Exact Test for Count Data

data:  table(d[, c("Two.Class", "Survived")])
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.1737612 0.2880626
sample estimates:
odds ratio
 0.2241543
```

3.6.5 Chi-Square Test of Homogeneity

Chi-Square Test for Homogeneity

Purpose. The chi-square test for homogeneity is used to compare whether *count data* collected from $J > 1$ populations come from the same sampling distribution. It is related to, but distinct from the following tests we have already learned:

- *rank sum test.* Given two samples $X_1, \ldots, X_n \sim F$ and $Y_1, \ldots, Y_m \sim G$ of continuous data, the rank sum test tests $H_0 : F = G$ for two samples.
- *goodness-of-fit test.* Given a sample of continuous or discrete data $X_1, \ldots, X_n \sim F$ and a named distribution G, the goodness-of-fit test tests $H_0 : F = G$.

Theory. Suppose you collect J independent samples of counts

$$\begin{aligned} (n_{11}, n_{21}, \ldots, n_{I1}) &\sim \text{Multinomial}\,(n_{\cdot 1}, (\pi_{11}, \ldots, \pi_{I1})) \\ (n_{12}, n_{22}, \ldots, n_{I2}) &\sim \text{Multinomial}\,(n_{\cdot 2}, (\pi_{12}, \ldots, \pi_{I2})) \\ &\vdots \\ (n_{1J}, n_{2J}, \ldots, n_{IJ}) &\sim \text{Multinomial}\,(n_{\cdot J}, (\pi_{1J}, \ldots, \pi_{IJ})) \end{aligned} \tag{4}$$

where $n_{\cdot j} = \sum_{i=1}^{I} n_{ij}$ are the column sums of the data matrix $\{n_{ij}\}$.

The null hypothesis is

$$H_0 : \pi_{i1} = \pi_{i2} = \cdots = \pi_{iJ} = \pi_i, \ \forall\, i \in \{1, \ldots, I\}$$

In words, H_0 claims that all J samples are drawn by repeated sampling from the same discrete distribution with probability mass function $\pi = (\pi_1, \ldots, \pi_I)$. In other words, the row sum vector $(n_{1\cdot}, \ldots, n_{I\cdot})$, with $n_{i\cdot} = \sum_{j=1}^{J} n_{ij}$ has a multinomial distribution with total $n_{\cdot\cdot} = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}$ trials and probability vector π.

We will construct a statistics not unlike the one used for goodness-of-fit tests, and for this, we need to compute the expected count in each category i under H_0. The MLEs of the null distribution π are

$$\hat{\pi}_i = \frac{n_{i\cdot}}{n_{\cdot\cdot}}$$

so we can estimate the expected number of type i observations in sample j as

$$E_{ij} = n_{\cdot j}\hat{\pi}_i = \frac{n_{i\cdot}n_{\cdot j}}{n_{\cdot\cdot}}$$

The statistic is

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

For much the same reasons it worked for the goodness-of-fit test, it can be shown that $X^2 \sim \chi^2_{(I-1)(J-1)}$, a chi-square distribution with $(I-1)(J-1)$ degrees of freedom. To determine the number of degrees of freedom, notice there are IJ data values n_{ij}. Each constraint on the model or parameter estimated from the data, removes one degree of freedom. Immediately, we lose J degrees of freedom because the sample sizes $n_{\cdot j}$ are fixed by assumption of the independent multinomial model, Eq. (4). There are $i-1$ equations needed to estimate $\hat{\pi}_i$ for all $i < I$, then $\hat{\pi}_I = 1 - \hat{\pi}_1 - \cdots - \hat{\pi}_{I-1}$. The result is

$$IJ - J - (I-1) = (I-1)(J-1)$$

degrees of freedom left for the statistic X^2.

Example: TV Preferences

Example:

	TV Show			Row
	Lone Range	Sesame Street	The Simpsons	Total
Boy	50	30	20	100
Girl	50	80	70	200
Col. Total	100	110	90	300

The number of degrees of freedom is $df = (3-1)(2-1) = 2$. The expected counts are

	TV Show		
	Lone Range	Sesame Street	The Simpsons
Boy	$\frac{100}{3}$	$\frac{110}{3}$	30
Girl	$\frac{200}{3}$	$\frac{220}{3}$	60

yielding

$$X^2 = 19.91$$

A quick call in `R` to `1-pchisq(19.91, df=2)` produces a p-value of 4.75×10^{-5} and rejection of H_0 homogeneity. This is almost always a one-sided test because you are rarely interested in outcomes in the left tail, where the data show unusually high homogeneity.

Application and Interpretation.

Merging samples. When using this test with $J > 2$ samples, it can arise that a subset of samples are expected to be homogeneous. If a subset of samples are known to be homogeneous, it makes sense to combine all samples in the subset. However, before assuming homogeneity in a the subset, it is good practice to test the assumption of homogeneity. If H_0 cannot be rejected for the subset, then the samples are merged, and additional tests of homogeneity can be performed. Keep in mind that if this procedure requires multiple tests, the significance level α should probably be adjusted, for example with Bonferroni correction. For example, it may be known that girls and boys react similarly to the relatively gender-neutral shows *Sesame Street* and *The Simpsons*. A test of homogeneity (or Fisher's exact would work here) for just the last two columns yields $X^2 = 0.6734$ with p-value 0.8237, permitting us to merge the last two columns to produce data

	TV Show	
	Lone Range	not *Lone Range*
Boy	50	50
Girl	50	150

Interpretting Rejected H_0. If the null hypothesis of homogeneity is rejected, then it becomes interesting to evaluate which cells in the table most deviate from homogeneity. One can look for specific deviations qualitatively by producing a table of values $\frac{(n_{ij}-E_{ij})^2}{E_{ij}}$. In some cases, the inhomogeneity is distributed evenly throughout the table and no cell contributes a large amount, but in other cases, one or a few cells will have substantially larger contributions. In our example, the table of contributions is

	TV Show		
	Lone Range	Sesame Street	The Simpsons
Boy	8.38	1.22	3.33
Girl	4.70	0.61	1.67

and it seems that girls and boys most disagree in their preference for the Lone Ranger.

3.6.6 Chi-Square Test of Independence

Chi-Square Test of Independence

Purpose. The chi-square test of independence uses the same test statistic and sampling distribution as the chi-square test of homogeneity. It differs in its null hypothesis and interpretation. It is thus related to the following tests

- *Fisher's Exact Test.* It generalizes the Fisher's exact test to handle more than two categories in each dimension.
- *Chi-Square Test of Homogeneity.* Shares the same statistic and sampling distribution but yields different conclusion.

Theory. The data n_{ij} are presumed to come from a multinomial distribution with total number of trials $n_{..}$ and probabilities π_{ij}. Notice, this is a *single* multinomial distribution for all the data; there is no concept of independent samples as for the last test. On the other hand, one of the dimensions, say i, may index samples. Under the assumption of independence, we have

$$H_0 : \pi_{ij} = \pi_{i\cdot}\pi_{\cdot j}$$

where $\pi_{i\cdot} = \sum_{j=1}^{J} \pi_{ij}$ is the probability of outcome i in the first dimension, and $\pi_{\cdot j} = \sum_{i=1}^{I} \pi_{ij}$ is the probability of outcome j in the second.

Under H_0, the MLE estimates of the probabilities are

$$\hat{\pi}_{i\cdot} = \frac{n_{i\cdot}}{n_{..}} \qquad\qquad \hat{\pi}_{\cdot j} = \frac{n_{\cdot j}}{n_{..}}$$

which implies expectation estimate

$$E_{ij} = n_{..}\frac{n_{i.}}{n_{..}}\frac{n_{.j}}{n_{..}} = \frac{n_{i.}n_{.j}}{n_{..}}$$

Since this is the same expectation obtained for the chi-square test of homogeneity, we know the statistic

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

is also the same.

The number of degrees of freedom is IJ minus 1 for the constrained number of samples $n_{..}$ and the $I - 1$ estimated $\hat{\pi}_{i.}$ and the $J - 1$ estimated $\hat{\pi}_{.j}$, so

$$IJ - 1 - (I - 1) - (J - 1) = (I - 1)(J - 1)$$

degrees of freedom, also the same.

Example: Politics

	Voting Preference			Row
	Republican	Democrat	Independent	Sum
Male	200	150	50	400
Female	250	300	50	600
	450	450	100	1000

In this case $X^2 = 16.2$, which yields p-value 0.0003, for a one-sided test (usually you aren't interested in the *unusually independent* outcome in the left tail).

Interpretation. Notice that many datasets like the one above are based on observational data. In such studies, we have no ability to randomly assign sex to voters. Thus, when an association (non-independence) is found, we can only conclude there is an association, nothing about what caused the association. In this case, for example, there could be an association because unemployed people had more time that allowed them to participate in the study, and these tended to be women and democrats.

3.6.7 Matched-Pair Design

Matched-Pair Design

Purpose. We have seen how matched pair studies tend to have more power to detect difference in population means for continuous data. The same strategy can be used to reduce variance in count data and detect more subtle differences as well. The test we are about to describe, McNemar's test, is thus related to the following other tests.

- *Matched-Pair t-Test.* Works on matched pair data when the observed data are continuous.
- *Fisher's Exact Test.* Tests for independence in a 2-by-2 table, but is inappropriate if the data are matched into pairs.

Theory. A binary response is recorded on pairs of individuals. If 1 is the positive response and 2 is negative, then a pair may contribute a count to n_{11} if both show positive response, n_{12} if the first in pair is positive, second is negative, n_{21} vice versa, and n_{22} if both are negative. These data can be arranged in a table as below with $n = n_{11} + n_{12} + n_{21} + n_{22}$ the total number of pairs.

		Individual Type A	
		Positive	Negative
Individual Type B	Positive	n_{11}	n_{12}
	Negative	n_{21}	n_{22}

Notice that typically the paired individuals differ in some key quantity which is the real focus of interest. For example, the type A individual in the pair may be a patient afflicted with the disease, while the other, type B, is not. One individual may be treated with drug A, while the other with drug B. Thus, n_{12} and n_{21} are distinguishable categories.

Let π_{ij} be the probability that a pair lands in category ij. Then, the null hypothesis of independence is encoded as

$$H_0 : \pi_{21} = \pi_{12}.$$

To see this, define $\pi_{i\cdot} = \pi_{i1} + \pi_{i2}$ and $\pi_{\cdot j}$ similarly. If there is independence of condition, A or B, and outcome, positive or negative, then both types of individuals, A and B, should have equal probabilities of outcome 'positive," i.e. $\pi_{1\cdot} = \pi_{\cdot 1}$. Similarly, $\pi_{2\cdot} = \pi_{\cdot 2}$. However, these two equations both simplify to the null hypothesis as stated above. To see it for the first equation, note that $\pi_{1\cdot} = \pi_{11} + \pi_{12} = \pi_{\cdot 1} = \pi_{11} + \pi_{21}$ or $\pi_{12} = \pi_{21}$.

Under H_0, we can estimate

$$\hat{\pi}_{11} = \frac{n_{11}}{n} \qquad \hat{\pi}_{22} = \frac{n_{22}}{n}$$

$$\hat{\pi}_{12} = \hat{\pi}_{21} = \frac{n_{12} + n_{21}}{2n}$$

which yields

$$E_{11} = n_{11} \qquad E_{22} = n_{22} \qquad E_{12} = E_{21} = \frac{n_{12} + n_{21}}{2}$$

The same old statistic is

$$X^2 = \frac{(n_{12} - E_{12})^2 + (n_{21} - E_{21})^2}{E_{12}} = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}$$

The degrees of freedom is

$$4 - 1 - 2 = 1$$

because of the four data values, all must sum to n, and two parameters $\hat{\pi}_{11}$ and $\hat{\pi}_{12}$ (then $\hat{\pi}_{22} = 1 - \hat{\pi}_{11} - 2\hat{\pi}_{12}$) were estimated.

Example: Honesty

Students are asked to agree/disagree with a statement: "If a test is unfair, it is OK to cheat." They are then subject to three sessions where they watch videos, do activities, and discuss academic honesty. After the third session, they are asked the question again. The data are naturally paired because the same subject is asked two questions. A table of the data is shown below.

		Post-Test	
		Disagree	Agree
Pre-Test	Disagree	134	36
	Agree	86	22

The statistic is

$$X^2 = \frac{(36 - 86)^2}{36 + 86} = 20.49$$

which is significant against χ^2_1, with p-value 5.99×10^{-06}.

Reference. Ciechalski, J.C., Pinkney, J.W., Weaver, F.S. A Method for Assessing Change in Attitude: The McNemar Test. Annual Meeting of the American Educational Research Association (New Orleans, LA, April 1-5, 2002).

3.6.8 Odds Ratio

Odds Ratio

Purpose. We understand that it is desirable to take random samples from populations, but sometimes random samples are a foolish waste of money. Consider the case where random individuals are sampled from a population in order to study a rare disease and its association with some rare risk factor. Let D be the event that a random individual has the disease. Otherwise, they do not $\bar{D}$. Let X be the event that a random invididual is exposed to the risk factor. Otherwise, they are not $\bar{X}$. If both the disease and exposure to the risk factor are rare, then even a very large sample may include no individuals with the disease, no individuals exposed, or, especially, no individuals both exposed and diseased.

In these situtations, it is far more logical to do targetted sampling, either in a

- *Prospective study,* where a fixed number of exposed and unexposed individuals are sampled and followed forward in time until they develop, or don't, the disease, or a
- *Retrospective study,* where a fixed number of diseased and healthy individuals are sampled, and queried whether they were exposed to the risk or not.

The trouble with the prospective or retrospective study is that it is not possible to estimate $P(D \cap X)$ or $P(D)$ or similar unconditional probabilities. For example, estimating $\hat{P}(D) = \frac{n_d}{n}$, where n_d is the number of diseased in a retrospective study is laughable (since n_d and n are fixed by the experimenter).

Theory We take some time to define some new terms.

Definition: *odds*

The odds of event A is defined as

$$\text{odds}(A) = \frac{P(A)}{1 - P(A)} = \frac{P(A)}{P(\bar{A})}$$

We can also define conditional odds in terms of conditional probabilities

$$\text{odds}(A|B) = \frac{P(A|B)}{1 - P(A|B)} = \frac{P(A|B)}{P(\bar{A}|B)}$$

Definition: *odds ratio*

The odds ratio

$$\Delta = \frac{\text{odds}(A|B)}{\text{odds}(A|C)}$$

quantifies how the odds change under different circumstances B and C. Typically $C = \bar{B}$, so the two conditions are complementary.

We will show that the odds ratio is equivalent under the prospective and retrospective studies. For the disease/exposure models, we can summarize the relevant model parameters in a table not unlike others we have seen.

	$\bar{D}$	D	
$\bar{X}$	π_{00}	π_{01}	$\pi_{0\cdot}$
X	π_{10}	π_{11}	$\pi_{1\cdot}$
	$\pi_{\cdot 0}$	$\pi_{\cdot 1}$	1

The various conditional probabilities are

$$P(D \mid X) = \frac{P(D \cap X)}{P(X)} = \frac{\pi_{11}}{\pi_{1\cdot}} \qquad P(\bar{D} \mid X) = \frac{P(\bar{D} \cap X)}{P(X)} = \frac{\pi_{10}}{\pi_{1\cdot}}$$

$$P(D \mid \bar{X}) = \frac{P(D \cap \bar{X})}{P(\bar{X})} = \frac{\pi_{01}}{\pi_{0\cdot}} \qquad P(\bar{D} \mid \bar{X}) = \frac{P(\bar{D} \cap \bar{X})}{P(\bar{X})} = \frac{\pi_{00}}{\pi_{0\cdot}}$$

which yields

$$\text{odds}(D \mid X) = \frac{\pi_{11}}{\pi_{10}} \qquad \text{odds}(D \mid \bar{X}) = \frac{\pi_{01}}{\pi_{00}}$$

and odds ratio

$$\Delta(D \mid X) = \frac{\pi_{11}\pi_{00}}{\pi_{10}\pi_{01}}$$

One can similarly show that $\Delta(X \mid D) = \Delta(D \mid X)$, so the prospective and retrospective studies result in the same odds ratio. We also simply write Δ.

Given that we observe count data n_{00}, n_{01}, n_{10}, and n_{11}, it is possible to estimate the conditional probabilities, for example

$$\hat{P}(X \mid D) = \frac{n_{11}}{n_{\cdot 1}}.$$

Since the odds ratio is

$$\Delta = \frac{P(X \mid D)P(\bar{X} \mid \bar{D}}{P(X \mid \bar{D})P(\bar{X} \mid D)},$$

it can be estimated as

$$\hat{\Delta} = \frac{n_{11}n_{00}}{n_{10}n_{01}}$$

Example: Feeding

A study (Westergren, A., Karlsson, S., Andersson, P., Ohlssom, O. and Hallberg, H.R. (2001) Eating difficulties, need for assisted eating, nutritional status and pressure ulcers in patients admitted for stroke rehabilitation. *J. Clinical Nursing*. **10**, 257.) examined whether patients who ate independently or dependently through the help of someone else, ate 3/4 or more of their food.

		Dependent	Independent	
Eats $\leq 3/4$	Yes	59	33	92
of food	No	17	44	61
	Total	76	77	153

The odds of dependent feeding in those who eat $\leq 3/4$ vs. those who don't is

$$\Delta = \frac{59/33}{17/44} = 4.63$$

so there is an increased risk of dependent feeding among those who eat $\leq 3/4$ food. Since we can flip the odds ratio around, we could say that the risk of eating $\leq 3/4$ food is increased in dependent eaters over independent eaters.

Confidence

We have not worked out the sampling distribution for Δ, so we have no way to assess whether the increased risk is significant or not. A natural hypothesis of interest is

$$H_0 : \Delta = 1$$

implying no independence of one condition (disease) from the other (exposure).

Challenge yourself. How might you use a computer to test the above null hypothesis, or provide confidence intervals for $\hat{\Delta}$?