

Inference for Proportions

Nathaniel E. Helwig

Associate Professor of Psychology and Statistics
University of Minnesota



October 17, 2020

Copyright © 2020 by Nathaniel E. Helwig

Table of Contents

1. Testing a Single Proportion

- Problem of Interest

- Exact Test

- Asymptotic Test

- Confidence Intervals

2. Testing Proportion Difference

- Problem of Interest

- Fisher's Exact Test

- Asymptotic Test

- Confidence Interval

- Odds, Relative Risk, Odds Ratios, and Phi

Table of Contents

1. Testing a Single Proportion

Problem of Interest

Exact Test

Asymptotic Test

Confidence Intervals

2. Testing Proportion Difference

Problem of Interest

Fisher's Exact Test

Asymptotic Test

Confidence Interval

Odds, Relative Risk, Odds Ratios, and Phi

Data Generating Assumptions

Suppose that we have collected an independent and identically distributed (iid) sample of observations x_1, \dots, x_n .

Assume each observation follows a Bernoulli distribution with probability of success π , i.e., $x_i \stackrel{\text{iid}}{\sim} \text{Bern}(\pi)$.

We want to test a null hypothesis about the probability of success.

Hypothesis Testing Options

Could test the following hypotheses:

- $H_0 : \pi = \pi_0$ versus $H_1 : \pi \neq \pi_0$ (exact H_0 with two-sided H_1)
- $H_0 : \pi \geq \pi_0$ versus $H_1 : \pi < \pi_0$ (inexact H_0 with less than H_1)
- $H_0 : \pi \leq \pi_0$ versus $H_1 : \pi > \pi_0$ (inexact H_0 with greater than H_1)

$\pi_0 \in (0, 1)$ is the null hypothesized value of the probability of success

The inexact null hypothesis (with one-sided H_1) is often preferred, because researchers often have an idea about the effect direction.

The Exact Null Distribution

If the null hypothesis H_0 is true, then $X = \sum_{i=1}^n x_i \sim B(n, \pi_0)$.

- Binomial PMF: $f(x; \pi) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}$ for $x \in \{0, 1, \dots, n\}$

For the directional (i.e., one-sided) tests, the p-values are

- $H_1 : \pi < \pi_0$, the p-value is computed as $p = \sum_{k=0}^X f(k; \pi_0)$
- $H_1 : \pi > \pi_0$, the p-value is computed as $p = \sum_{k=X}^n f(k; \pi_0)$

Calculate the probability of being “as or more extreme” than X in the direction specified by the alternative hypothesis.

Two-Sided P-Value Calculations

For two-sided tests, the p-value can be computed as

$$p = \begin{cases} 1 & \text{if } X = \mu_0 \\ \sum_{k=0}^X f(k; \pi_0) + \sum_{k=n-Y+1}^n f(k; \pi_0) & \text{if } X < \mu_0 \\ \sum_{k=0}^{Y-1} f(k; \pi_0) + \sum_{k=X}^n f(k; \pi_0) & \text{if } X > \mu_0 \end{cases}$$

where $\mu_0 = n\pi_0$ is the expected number of successes under H_0 and

$$Y = \begin{cases} \sum_{k=\lceil \mu_0 \rceil}^n I(f(k; \pi_0) \leq f(X; \pi_0)) & \text{if } X < \mu_0 \\ \sum_{k=0}^{\lfloor \mu_0 \rfloor} I(f(k; \pi_0) \leq f(X; \pi_0)) & \text{if } X > \mu_0 \end{cases}$$

with $I(\cdot)$ denoting an indicator function.

These are previous (one-sided) formulas plus an additional term.

Example 1

Assume that a researcher suspects that 1% of people in a particular country have contracted some disease (e.g., COVID-19).

- Goal: test $H_0 : \pi = 0.01$ versus $H_1 : \pi \neq 0.01$

Researcher samples $n = 1000$ individuals and records whether the individual has antibodies for the disease ($X = 1$) or not ($X = 0$).

Researcher finds that $X = 19$ individuals in the sample test positive for antibodies.

Example 1 (continued)

We could compute this p-value directly, but the `binom.test` function in R (R Core Team, 2020) does the hard work for us:

```
> binom.test(x = 19, n = 1000, p = 0.01)
```

Exact binomial test

data: 19 and 1000

number of successes = 19, number of trials = 1000, p-value = 0.009584

alternative hypothesis: true probability of success is not equal to 0.01

95 percent confidence interval:

0.01147704 0.02951240

sample estimates:

probability of success

0.019

The Asymptotic Null Distribution

If the null hypothesis $H_0 : \pi = \pi_0$ is true, then

$$Z = \frac{X - n\pi_0}{\sqrt{n\pi_0(1 - \pi_0)}} \dot{\sim} N(0, 1)$$

if the sample size n is large enough.

We can test null hypotheses about π by comparing the observed Z to the quantiles of a standard normal distribution.

Yate's correction for continuity subtracts $1/2$ from the numerator

$$Z = \frac{X - n\pi_0 - 1/2}{\sqrt{n\pi_0(1 - \pi_0)}}$$

Example 1 (revisited)

Can implement an asymptotic version of this test using the R code:

```
> prop.test(x = 19, n = 1000, p = 0.01)
```

1-sample proportions test with continuity correction

```
data: 19 out of 1000, null probability 0.01
X-squared = 7.298, df = 1, p-value = 0.006903
alternative hypothesis: true p is not equal to 0.01
95 percent confidence interval:
 0.01180558 0.03008791
sample estimates:
      p
0.019
```

Note: default use of `prop.test` implements continuity correction.

Different Methods for Proportion Confidence Intervals

Letting $z = z_{1-\alpha/2}$ denote the $1 - \alpha/2$ quantile of $N(0, 1)$ distribution

- Normal Approximation: $\hat{\pi} \pm z\sqrt{\hat{\pi}(1 - \hat{\pi})/n}$
- Wilson Score Method: $\left(1 + \frac{z^2}{n}\right)^{-1} \left(\hat{\pi} + \frac{z^2}{2n} \pm z\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n} + \frac{z^2}{4n^2}}\right)$
- Agresti-Coull Method: $\tilde{\pi} \pm z\sqrt{\frac{\tilde{\pi}(1-\tilde{\pi})}{n+z^2}}$ where

$$\tilde{\pi} = \left(1 + \frac{z^2}{n}\right)^{-1} \left(\hat{\pi} + \frac{z^2}{2n}\right)$$
- Arcsin Method: $\sin^2 \left(\arcsin(\sqrt{p}) \pm \frac{z}{2\sqrt{n}} \right)$
- Clopper-Pearson: $[B(\alpha/2; X, n - X + 1), B(1 - \alpha/2; X + 1, n - X)]$
 where $B(\cdot; \alpha, \beta)$ is the quantile function for the beta distribution
 with shape parameters α and β

R Function for Proportion CIs

```
prop.ci <- function(x, n, level = 0.95){

  p <- x / n
  se <- sqrt(p * (1 - p) / n)
  z <- qnorm(1 - alpha/2)
  normal <- c(p - z * se, p + z * se)

  denom <- 1 + z^2 / n
  p.adj <- p + z^2 / (2 * n)
  se.adj <- sqrt(p * (1 - p) / n + z^2 / (4 * n^2))
  wilson <- c(p.adj - z * se.adj, p.adj + z * se.adj) / denom

  p.adj <- p.adj / denom
  se.adj <- sqrt(p.adj * (1 - p.adj) / (n + z^2))
  agresti.coull <- c(p.adj - z * se.adj, p.adj + z * se.adj)

  asrp <- asin(sqrt(p))
  asinz <- z / (2 * sqrt(n))
  arcsin <- c(sin(asrp - asinz)^2, sin(asrp + asinz)^2)

  ci.lower <- qbeta(alpha/2, x, n - x + 1)
  ci.upper <- qbeta(1 - alpha/2, x + 1, n - x)
  clopper.pearson <- c(ci.lower, ci.upper)

  ci <- rbind(normal, wilson, agresti.coull, arcsin, clopper.pearson)
  rownames(ci) <- c("normal", "wilson", "agresti.coull", "arcsin", "clopper.pearson")
  colnames(ci) <- c("lower", "upper")
  res <- list(x = x, n = n, level = level, conf.int = ci)
  return(res)

} # end prop.ci
```

Example 1 Confidence Intervals

```
> prop.ci(x = 19, n = 1000)$conf.int
              lower      upper
normal      0.01053827 0.02746173
wilson      0.01219689 0.02948446
agresti.coull 0.01200381 0.02967753
arcsin      0.01146726 0.02837989
clopper.pearson 0.01147704 0.02951240
```

Clopper-Pearson from `binom.test`, and Wilson from `prop.test`:

```
> binom.test(x = 19, n = 1000, p = 0.01)$conf.int
[1] 0.01147704 0.02951240
attr("conf.level")
[1] 0.95
> prop.test(x = 19, n = 1000, p = 0.01, correct = FALSE)$conf.int
[1] 0.01219689 0.02948446
attr("conf.level")
[1] 0.95
```

Table of Contents

1. Testing a Single Proportion

Problem of Interest

Exact Test

Asymptotic Test

Confidence Intervals

2. Testing Proportion Difference

Problem of Interest

Fisher's Exact Test

Asymptotic Test

Confidence Interval

Odds, Relative Risk, Odds Ratios, and Phi

Assumptions and Hypotheses

$x_i \stackrel{\text{iid}}{\sim} \text{Bern}(\pi_x)$ for $i = 1, \dots, n_x$ and $y_i \stackrel{\text{iid}}{\sim} \text{Bern}(\pi_y)$ for $i = 1, \dots, n_y$,
and all observations are independent of one another.

We could test three different hypotheses:

- $H_0 : \pi_x = \pi_y$ vs $H_1 : \pi_x \neq \pi_y$ (exact H_0 with two-sided H_1)
- $H_0 : \pi_x \geq \pi_y$ vs $H_1 : \pi_x < \pi_y$ (inexact H_0 with less than H_1)
- $H_0 : \pi_x \leq \pi_y$ vs $H_1 : \pi_x > \pi_y$ (inexact H_0 with greater than H_1)

In many applications, the directional alternatives are preferable.

The Exact Null Distribution

Fisher's exact test involves forming the 2×2 contingency table

	Success	Failure	Total
Population 1	a	b	$a + b = n_x$
Population 2	c	d	$c + d = n_y$
Total	$a + c = n_1$	$b + d = n_0$	$a + b + c + d = n$

- $a = \sum_{i=1}^{n_x} x_i$ is the number of successes for population 1
- $b = n_x - a$ is the number of observed failures for population 1
- $c = \sum_{i=1}^{n_y} y_i$ is the number of successes for population 2
- $d = n_y - c$ is the number of observed failures for population 2

The Exact Null Distribution (continued)

Conditioned on the marginals of the table, Fisher showed that the probability of observing a particular combination of cell values was given by the hypergeometric distribution.

Assuming that the row marginals $n_x = a + b$ and $n_y = c + d$ are fixed and the column marginals $n_1 = a + c$ and $n_0 = b + d$ are fixed, the probability of observing a successes in population 1 is given by

$$f(a; n_x, n_y, n_1) = \frac{\binom{n_x}{a} \binom{n_y}{n_1 - a}}{\binom{n_x + n_y}{n_1}}$$

which is the PMF of the hypergeometric distribution.

P-Values for Fisher's Exact Test

For the one-sided tests, the p-values are simple to compute, given that they just involve summing the hypergeometric PMF for a number of successes as or more extreme than a (in the direction of H_1)

- $H_1 : \pi_x < \pi_y$, the p-value is $p = \sum_{k=0}^a f(k; n_x, n_y, n_1)$
- $H_1 : \pi_x > \pi_y$, the p-value is $p = \sum_{k=a}^{n_1} f(k; n_x, n_y, n_1)$

For the two-sided tests, the p-value can be computed as

$$p = \sum_{k=a_L}^{a_U} f(k; n_x, n_y, n_1) I(f(k; n_x, n_y, n_1) \leq f(a; n_x, n_y, n_1) \delta)$$

where $a_L = \max(0, n_x - n_0)$ and $a_U = \min(n_x, n_1)$. The `fisher.test` function in R includes a tolerance factor of $\delta = 1.0000001$

Example 2: Problem Setup

Contingency table from Table 4 of Radelet and Pierce (1991):

Defendant	Death Penalty		Total
	Yes	No	
White	53	430	483
Black	15	176	191
Total	68	606	674

Want to test $H_0 : \pi_x = \pi_y$ versus $H_1 : \pi_x \neq \pi_y$

- π_x is probability of receiving death penalty for white defendants
- π_y is probability of receiving death penalty for black defendants

Example 2: Estimates and P-Value

The sample estimates of the probability are

$$\hat{\pi}_x = 53/483 = 0.10973085 \quad \text{and} \quad \hat{\pi}_y = 15/191 = 0.07853403$$

but is this difference significant?

The probability of the observed table is given by

$$P(a = 53 \mid n_x = 483, n_y = 191, n_1 = 68, n_0 = 606) = \frac{\binom{483}{53} \binom{191}{15}}{\binom{674}{68}} = 0.056329$$

and the p-value for the two-sided alternative hypothesis is given by

$$p = \sum_{k=0}^{68} \frac{\binom{483}{k} \binom{191}{15}}{\binom{674}{68}} I \left(\frac{\binom{483}{k} \binom{191}{15}}{\binom{674}{68}} \leq 0.05632907 \right) = 0.2577816$$

Example 2: R Implementation

```
> xtab <- matrix(c(53, 15, 430, 176), 2, 2)
> colnames(xtab) <- c("Yes", "No")
> rownames(xtab) <- c("White", "Black")
> xtab
      Yes  No
White  53 430
Black  15 176
> fisher.test(xtab)
```

Fisher's Exact Test for Count Data

```
data:  xtab
p-value = 0.2578
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.777655 2.837046
sample estimates:
odds ratio
 1.445462
```

The Asymptotic Null Distribution

For large n , we have $\hat{\pi}_x \dot{\sim} N\left(\pi_x, \frac{\pi_x(1-\pi_x)}{n_x}\right)$ and $\hat{\pi}_y \dot{\sim} N\left(\pi_y, \frac{\pi_y(1-\pi_y)}{n_y}\right)$.

If the null hypothesis $H_0 : \pi_x = \pi_y$ is true, then

$$Z = \frac{\hat{\pi}_x - \hat{\pi}_y}{\sqrt{\hat{\pi}(1-\hat{\pi})(1/n_x + 1/n_y)}} \dot{\sim} N(0, 1)$$

for large enough n , where $\hat{\pi} = (a + c)/(n_x + n_y)$ is the common proportion estimate.

Can test null hypotheses about $\pi_x - \pi_y$ by comparing Z to a $N(0, 1)$, or by comparing Z^2 to a χ_1^2 .

Example 2: Asymptotic Inference

```
> nx <- sum(xtab[1,])
> ny <- sum(xtab[2,])
> px <- xtab[1,1] / nx
> py <- xtab[2,1] / ny
> p0 <- sum(xtab[,1]) / (nx + ny)
> pdif <- px - py
> pdif.se <- sqrt(p0 * (1 - p0) * (1/nx + 1/ny))
> z <- pdif / pdif.se
> z^2
[1] 1.468519
> 1 - pchisq(z^2, df = 1)
[1] 0.2255796
```


Example 2: Asymptotic Inference (continued)

```
> prop.test(xtab, correct = FALSE)
```

2-sample test for equality of proportions without continuity correction

```
data:  xtab
```

```
X-squared = 1.4685, df = 1, p-value = 0.2256
```

```
alternative hypothesis: two.sided
```

```
95 percent confidence interval:
```

```
-0.01605167  0.07844531
```

```
sample estimates:
```

```
prop 1      prop 2
```

```
0.10973085 0.07853403
```

Asymptotic CI for Proportion Difference

The $100(1 - \alpha)\%$ confidence interval for $\pi_x - \pi_y$ has the form

$$\hat{\pi}_x - \hat{\pi}_y \pm z_{1-\alpha/2} \sqrt{\frac{\hat{\pi}_x(1 - \hat{\pi}_x)}{n_x} + \frac{\hat{\pi}_y(1 - \hat{\pi}_y)}{n_y}}$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the $N(0, 1)$ distribution.

Note that, unlike the denominator of the Z test statistic that was used for the hypothesis test, the standard error formula for the confidence interval does *not* use the pooled proportion estimate.

The confidence interval for $\pi_x - \pi_y$ is not formed under the assumption that $H_0 : \pi_x = \pi_y$ is true.

Example 2: Confidence Interval

Continuing with the same example, we could form the confidence interval “by hand”, but the `prop.test` function outputs this result for us (see previous example).

```
> pdif <- px - py
> pdif.se <- sqrt(px * (1 - px) / nx + py * (1 - py) / ny)
> c(pdif - qnorm(0.975) * pdif.se, pdif + qnorm(0.975) * pdif.se)
[1] -0.01605167  0.07844531
```

Odds of Success

The odds of success is the ratio of the probability of success over the probability of failure, i.e., $\text{odds} = \frac{\pi}{1-\pi}$ where $\pi = P(X = 1)$ and $1 - \pi = P(X = 0)$.

If someone says that “the odds are A to B” this means that the probability of success is $\frac{A}{A+B}$ and the probability of failure is $\frac{B}{A+B}$.

For example, if the odds are 3 to 1, this means that the probability of success is $3/4$ and the probability of failure is $1/4$.

Risk Ratio and Odds Ratio

The relative risk, also known as the risk ratio, is the ratio of the probability of success for two populations, i.e., $RR = \pi_x / \pi_y$, where $\pi_x = P(X = 1)$ and $\pi_y = P(Y = 1)$.

The odds ratio is the ratio of the odds of success for two populations

$$OR = \frac{\pi_x / (1 - \pi_x)}{\pi_y / (1 - \pi_y)} = \frac{\pi_x(1 - \pi_y)}{\pi_y(1 - \pi_x)} = \frac{ad}{bc}$$

where (a, b, c, d) denote the entries of the 2×2 contingency table.

Phi Coefficient

The ϕ coefficient is another statistic that can be used to measure the strength of association in a 2×2 contingency table.

The ϕ coefficient is defined as

$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

and note that $\phi = 0$ indicates no association between the variables.

The ϕ coefficient is related to Pearson's chi-square test statistic, i.e., $\phi = \sqrt{X^2/n}$, which will be discussed in the next chapter.

Example 2: Odds and Relative Risk

The odds of receiving the death penalty are

$$\text{White : } \frac{\hat{\pi}_x}{1 - \hat{\pi}_x} = 0.1232558 \quad \text{and} \quad \text{Black : } \frac{\hat{\pi}_y}{1 - \hat{\pi}_y} = 0.08522727$$

so the odds of the death penalty are larger for white defendants.

The relative risk of receiving a death penalty verdict is given by

$$\text{RR} = \hat{\pi}_x / \hat{\pi}_y = 1.397239$$

so the probability of receiving the death penalty is about 1.4 times larger for white defendants compared to black defendants.

Example 2: Odds Ratio

The odds ratio for receiving the death penalty verdict is

$$\hat{\theta} = \frac{\hat{\pi}_x / (1 - \hat{\pi}_x)}{\hat{\pi}_y / (1 - \hat{\pi}_y)} = \frac{(53)(176)}{(430)(15)} = 1.446202$$

so the odds of receiving the death penalty is about 1.45 times larger for white defendants compared to black defendants.

The output of the `fisher.test` function provides this odds ratio estimate, as well as a 95% confidence interval for the odds ratio.

Example 2: Odds Ratio Asymptotic CI

The `fisher.test` function uses an exact method to obtain the 95% confidence interval, which is `[0.777655, 2.837046]`.

We could calculate the confidence interval using an asymptotic argument as well (forms the interval on log-scale).

- $\log(\hat{\theta}) \sim N(0, V)$ where $V = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$

```
> OR <- xtab[1,1] * xtab[2,2] / (xtab[1,2] * xtab[2,1])
> logOR <- log(OR)
> logOR.se <- sqrt(1/xtab[1,1] + 1/xtab[2,2] + 1/xtab[1,2] + 1/xtab[2,1])
> logci <- c(logOR - qnorm(.975) * logOR.se, logOR + qnorm(.975) * logOR.se)
> exp(logci)
[1] 0.7941306 2.6336964
```

References

- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. R version 4.0.2.
- Radelet, M. L. and G. L. Pierce (1991). Choosing those who will die: Race and the death penalty in Florida. *Florida Law Review* 43(1), 1–34.