

Inference for Proportions

Nathaniel E. Helwig

University of Minnesota

1 Testing a Single Proportion

1.1 Problem of Interest

Suppose that we have collected an independent and identically distributed (iid) sample of observations x_1, \dots, x_n , where each observation is assumed to follow a Bernoulli distribution with probability of success π , i.e., $x_i \stackrel{\text{iid}}{\sim} \text{Bern}(\pi)$. Furthermore, suppose that we want to test a null hypothesis about the probability of success. As a reminder from the previous chapter, we 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)

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

1.2 Exact Test

If the null hypothesis H_0 is true, then the random variable $X = \sum_{i=1}^n x_i$ follows a binomial distribution with size n and probability of success π_0 , i.e., $X \sim B(n, \pi_0)$. Thus, we can use X as a test statistic to compute an exact p-value for testing the null hypothesis of interest. As a reminder, the p-value for testing H_0 is the probability of observing a test statistic as or more extreme than the observed test statistic under the assumption that H_0 is true. What

it means to be “as or more extreme than the observed test statistic” will depend on the type of null hypothesis is being tested.

As a reminder, the binomial probability mass function has the form

$$f(x; \pi) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}$$

for any input $x \in \{0, 1, \dots, n\}$. For the directional (i.e., one-sided) tests, the p-values are

- $H_0 : \pi \geq \pi_0$ versus $H_1 : \pi < \pi_0$, the p-value is computed as $p = \sum_{k=0}^X f(k; \pi_0)$
- $H_0 : \pi \leq \pi_0$ versus $H_1 : \pi > \pi_0$, the p-value is computed as $p = \sum_{k=X}^n f(k; \pi_0)$

given that we just need to calculate the probability of being “as or more extreme” than X in the direction specified by the alternative hypothesis.

For the two-sided test $H_0 : \pi = \pi_0$ versus $H_1 : \pi \neq \pi_0$, the p-value is a bit more complicated given that we need to calculate the probability of being “as or more extreme” than X in both directions. 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 the null hypothesis 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.

To understand the formulas for the two-sided p-value calculations, note that when $X < \mu_0$ the p-value formula contains the p-value formula used for the “less than” alternative plus a term that sums up the probabilities of observing an X that would be equivalently extreme in the upper tail. Similarly, when $X > \mu_0$ the p-value formula contains the p-value formula used for the “greater than” alternative plus a term that sums up the probabilities of observing an X that would be equivalently extreme in the lower tail. When the null hypothesized value is $\pi_0 = 0.5$, the two-tailed p-value will simply be two times the one-tailed p-value.

Example 1. Assume that a researcher suspects that 1% of people in a particular country have contracted some disease (e.g., COVID-19), and wants to test the null hypothesis $H_0 : \pi = 0.01$ versus the alternative hypothesis $H_1 : \pi \neq 0.01$. To test this hypothesis, the researcher collects a random sample of $n = 1000$ individuals and records whether the individual has antibodies for the disease ($X = 1$) or not ($X = 0$). Suppose that the researcher finds that $X = 19$ individuals in the sample test positive for antibodies. We could use the above formula for calculating the two-sided p-value; however, the `binom.test` function in R 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 observed p-value is $p = 0.009584$, so using a standard significance level of $\alpha = 0.05$ or $\alpha = 0.01$, we would reject the null hypothesis that 1% of people have contracted the disease.

Example 2. Consider the previous example, but now suppose that the researcher believes that the percentage of people who have contracted the disease is less than or equal to 1% of the population. This would involve testing the null hypothesis $H_0 : \pi \leq 0.01$ versus the alternative hypothesis $H_1 : \pi > 0.01$. To test this hypothesis, we would just need to change the default of the “alternative” argument in the `binom.test` function from “two-sided” to “greater”. Assuming that we observe the same number of individuals with antibodies, i.e., $X = 19$, the p-value for testing the one-sided hypothesis would be $p = 0.006905$. Thus, using a standard significance level, we would reject the researcher’s hypothesis that $H_0 : \pi \leq 0.01$ in favor of the alternative hypothesis $H_1 : \pi > 0.01$.

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

```
Exact binomial test
```

```
data: 19 and 1000
number of successes = 19, number of trials = 1000, p-value = 0.006905
alternative hypothesis: true probability of success is greater than 0.01
95 percent confidence interval:
 0.01247677 1.00000000
sample estimates:
probability of success
      0.019
```

1.3 Asymptotic Test

Given that an exact test is possible, we would not really want to use this asymptotic test, which only works well for large n . However, it is pedagogically useful to introduce this asymptotic test for a single proportion, given that we will leverage these ideas for testing the multiple proportions. As a reminder, for large enough n , the number of success $X = \sum_{i=1}^n x_i$ is approximately normally distributed with mean $E(X) = n\pi$ and variance $\text{Var}(X) = n\pi(1 - \pi)$. So, if the null hypothesis is true, then

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

for large enough n . Thus, we can test null hypotheses about π by comparing the observed Z to the quantiles of a standard normal distribution. This sort of test is implemented in the `prop.test` function in R. Note that the default use of the `prop.test` function uses Yate's correction for continuity, which involves subtracting $1/2$ from the numerator in the calculation of Z , i.e.,

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

when using the continuity correction. The continuity correction is typically a good thing to do if you have small samples, because it avoids overly-optimistic proportion estimates.

Example 3. Revisiting the previous example of testing the null hypothesis $H_0 : \pi = 0.01$ versus the alternative hypothesis $H_1 : \pi \neq 0.01$, we could implement an asymptotic version of this test in R using the 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
```

which produces a p-value that is similar to the p-value obtained from the exact test.

1.4 Confidence Intervals

Letting $z = z_{1-\alpha/2}$ denote the $1 - \alpha/2$ quantile of the standard normal distribution, we could form a confidence interval for π using the various methods:

- 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

Note that the first method (i.e., the normal approximation) is what is typically taught in introductory statistics courses; however, this method tends to perform the worst of all of the methods. This is because the normal approximation method is a large sample (asymptotic) method that performs well for large n but does not necessarily perform well for finite n . The last method (i.e., the Clopper-Pearson method) is an exact method, which uses the relationship between the Binomial and Beta distributions to form the confidence interval. However, this method tends to be conservative, i.e., it may not produce the shortest possible interval that has the desired confidence level. The other three methods are various approximations that are designed to be (i) more accurate than the asymptotic normal approximation, and (ii) less conservative than the exact Clopper-Pearson method.

Example 4. The `prop.ci` function given in the Appendix forms the various confidence intervals. Continuing with the same example as before:

```
> 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
```

As you can see, all of the methods aside from the normal approximation give a similar interval, with the Clopper-Pearson interval being the widest. Also, note that the Clopper-Pearson interval is the same as the interval produced by the `binom.test` function, and the Wilson interval is the same as that produced by the `prop.test` function.

```
> 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
```

2 Testing Proportion Difference

2.1 Problem of Interest

Suppose that $x_i \stackrel{\text{iid}}{\sim} \text{Bern}(\pi_x)$ for $i = 1, \dots, n_x$ and $y_i \stackrel{\text{iid}}{\sim} \text{Bern}(\pi_y)$ for $i = 1, \dots, n_y$, and assume that all x_i and y_i are independent of one another. Furthermore, suppose that we want to test a null hypothesis about the difference in the probability of successes for the two populations. As in the one sample case, we could test three different hypotheses:

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

2.2 Exact Test

Fisher derived a method for exactly testing null hypotheses about proportion differences. 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	$a + c = n_1$	$b + d = n_0$	$a + b + c + d = n$

where $a = \sum_{i=1}^{n_x} x_i$ is the number of observed successes for population 1 (the “x” group), $b = n_x - a$ is the number of observed failures for population 1, $c = \sum_{i=1}^{n_y} y_i$ is the number of observed successes for population 2 (the “y” group), $d = n_y - c$ is the number of observed failures for population 2, and $n = n_x + n_y$ is the total sample size. Fixing the marginals of the table, Fisher showed that the probability of observing a particular combination of cell values was given by the hypergeometric distribution.

Specifically, 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 probability mass function of the hypergeometric distribution. This implies that the (conditional) p-values for testing null hypotheses about differences in proportions can be computed directly from the hypergeometric probability mass function. Note that these p-values are “conditional” in the sense that they are conditioned on the fixed row marginals (n_x and n_y) and column marginals (n_1 and n_0).

For the directional (one-sided) tests, the p-values are simple to compute, given that they just involve summing the hypergeometric probability mass function for a number of successes as or more extreme than the observed a (in the direction of H_1), i.e.,

- $H_0 : \pi_x \geq \pi_y$ versus $H_1 : \pi_x < \pi_y$, the p-value is $p = \sum_{k=0}^a f(k; n_x, n_y, n_1)$
- $H_0 : \pi_x \leq \pi_y$ versus $H_1 : \pi_x > \pi_y$, the p-value is $p = \sum_{k=a}^{n_1} f(k; n_x, n_y, n_1)$

For the two-sided hypothesis $H_0 : \pi_x = \pi_y$ versus $H_1 : \pi_x \neq \pi_y$, the p-value is a bit more complicated given that we need to calculate the probability of being “as or more extreme” than a in both directions. 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 $a_U = \min(n_x, n_1)$. Note that this p-value calculation sums up the probabilities for all tables (with the same marginals) that have probabilities smaller than or equal to the observed table’s probability of $f(a; n_x, n_y, n_1)$.

The `fisher.test` function in R includes a tolerance factor of $\delta = 1.0000001$, which ensures that outcomes that are essentially as extreme as the observed table are included in the p-value calculation. This is often useful because evaluating the hypergeometric probability mass function can be slightly inexact when the row and column marginals are moderate to large—because it is necessary to calculate factorials of large numbers. However, this could inflate the p-value if there are tables with probabilities that are just slightly larger than the observed table’s probability. Note that for any $\delta > 1$, the Fisher test may be a bit conservative (i.e., return p-values that are slightly too large), especially if the sample size is large, because a large n means that there exists many possible tables. So, when conducting two-sided Fisher exact tests in R, it may be useful to compare the results of the Fisher test obtained from R’s `fisher.test` function with a manual calculation of the p-value using the hypergeometric distribution.

Example 5. The following contingency table is from Table 4 of Radelet and Pierce (1991), which cross-classifies individuals who received a death penalty sentence by race:

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

Suppose that we want to test the null hypothesis that the probability of receiving the death penalty is the same for both white and black defendants, i.e., $H_0 : \pi_x = \pi_y$, versus the two-sided alternative that these two probabilities are different, i.e., $H_1 : \pi_x \neq \pi_y$. Note that 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

$$P(a = 53 \mid n_x = 483, n_y = 191, n_1 = 68, n_0 = 606) = \frac{\binom{483}{53} \binom{191}{15}}{\binom{674}{68}} = 0.05632907$$

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

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

so we fail to reject the null hypothesis $H_0 : \pi_x = \pi_y$ using any standard significance level. We can use the `fisher.test` function in R to confirm our results:

```
> 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
```

2.3 Asymptotic Test

Given that an exact test is possible, we would not really want to use this asymptotic test, which only works well for large $n = n_x + n_y$. However, it is pedagogically useful to introduce this asymptotic test for the equality of two proportions, given that we will leverage these ideas for testing the multiple proportions (in our discussion of chi-square tests next week). As a reminder, for large enough n_x , the number of success $a = \sum_{i=1}^{n_x} x_i$ is approximately normally distributed with mean $E(a) = n_x \pi_x$ and variance $\text{Var}(a) = n_x \pi_x (1 - \pi_x)$. Similarly, for large enough n_y , the number of success $c = \sum_{i=1}^{n_y} y_i$ is approximately normally distributed with mean $E(c) = n_y \pi_y$ and variance $\text{Var}(c) = n_y \pi_y (1 - \pi_y)$. So, for large enough samples, the proportion estimates are approximately normal with $\hat{\pi}_x \dot{\sim} N(\pi_x, \pi_x(1 - \pi_x)/n_x)$ and $\hat{\pi}_y \dot{\sim} N(\pi_y, \pi_y(1 - \pi_y)/n_y)$. 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—which is the correct thing to use under the assumption $H_0 : \pi_x = \pi_y$ is true. Thus, we can test null hypotheses about $\pi_x - \pi_y$ by comparing the observed Z to the quantiles of a standard normal distribution (or by comparing Z^2 to the quantiles of a χ_1^2 distribution).

Example 6. Revisiting the example that we used for Fisher’s exact test, we could use the asymptotic test. The R code to do this “by hand” is pretty simple, but the `prop.test` function in R does the work for us:

```
> 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

> 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
```

Using the asymptotic test, we would fail to reject the null hypothesis $H_0 : \pi_x = \pi_y$, and the p-value is quite similar to what we obtained using Fisher’s exact test.

2.4 Confidence Intervals

It is typical to use the large sample (normal) approximation to form confidence intervals for differences in proportions. 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 standard normal 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. This is because the standard error used in the denominator of Z is the appropriate standard error under the assumption that $H_0 : \pi_x = \pi_y$ is true. In contrast, the confidence interval for $\pi_x - \pi_y$ is not formed under the assumption that $H_0 : \pi_x = \pi_y$ is true, so the standard error formula uses the individual estimates of the proportions. Thus, in the case of inference with proportions, the confidence interval will not use the same standard error as is used for the test of the null hypothesis—because the parameters of interest (π_x and π_y) appear within the standard error formula.

Example 7. 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
```

2.5 Odds, Relative Risk, Odds Ratios, and Phi

Definition. The odds of success is the ratio of the probability of success over the probability of failure, i.e., $\text{odds} = \pi / (1 - \pi)$ 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 $A / (A + B)$ and the probability of failure is $B / (A + B)$. For example, if the odd are 3 to 1, this means that the probability of success is $3/4$ and the probability of failure is $1/4$.

Definition. 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 $\pi_y = P(Y = 1)$.

Definition. The odds ratio is the ratio of the odds of success for two populations, i.e.,

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

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

Definition. 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 $\phi = 0$ this indicates no association between the variables. Note that 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 8. For the death penalty example, the odds of receiving the death penalty are

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

so the odds of receiving the death penalty are slightly larger for white defendants. The relative risk of receiving a death penalty verdict is given by

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

so the probability of receiving the death penalty is about 1.4 times larger for white defendants compared to black defendants. 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 (in Example 5) provides this odds ratio estimate, as well as a 95% confidence interval for the odds ratio, which is given by [0.777655, 2.837046]. Since this confidence interval contains 1, we would conclude that the odds of receiving the death penalty verdict are not significantly different for white and black defendants. The `fisher.test` function uses an exact method to obtain the 95% confidence interval, but we could calculate the confidence interval using an asymptotic argument as well. The large sample confidence interval for the (log of the) odds ratio has the form

$$\log(\hat{\theta}) \pm z_{1-\alpha/2} \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

Using the asymptotic confidence interval gives a similar result as the exact interval that is produced by the `fisher.test` function:

```
> 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
```

The estimate of the ϕ coefficient is given by

$$\hat{\phi} = \frac{(53)(176) - (430)(15)}{\sqrt{(483)(191)(68)(606)}} = 0.04667773$$

and note that the “X-squared” that is output by the `prop.test` function (in Example 6) is related to this ϕ coefficient estimate, such that

$$X^2 = n\hat{\phi}^2 = 1.468519$$

References

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.

Appendix

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

  # checks
  x <- as.integer(x[1])
  if(x < 0) stop("Input 'x' must be a positive integer")
  n <- as.integer(n[1])
  if(n < 0) stop("Input 'n' must be a positive integer")
  if(x > n) stop("Inputs must satisfy: x <= n")
  level <- as.numeric(level[1])
  if(level <= 0 | level >= 1) stop("Input 'level' must be between 0 and 1")
  alpha <- 1 - level

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

  # wilson
  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

  # agresti-coull
  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)

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

  # clopper-pearson
  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)

  # return results
  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
```