**Inferences for the difference of two proportions**



Notation:

* Let y1 be the observed number of successes out of n1 trials (or sample size) from population #1.
* Let y2 be the observed number of successes out of n2 trials (or sample size) from population #2.
* Suppose the probability of success is π1 for population 1 and π2 for population 2. Then

 estimates π1

and

 estimates π2.

Wald CI for π1 – π2:



Where does this interval come from?

A very similar derivation can be performed as in the last section!

Similar to the Wald CI in this last section, this interval is not very good. Below is the interval that we will use instead.

Agresti and Caffo CI for π1 – π2:

Let

 and 

be adjusted estimators for π1 and π2. The (1-α)100% confidence interval for π1 – π2 is:



Notes:

* Agresti and Caffo recommend using the “add two successes and two failures” methods for a CI of ANY level of confidence. Again, Agresti and Caffo do not change the adjustment for different confidence levels! In their paper, they examined adding other number of successes and failures and found this to be the best.
* Notice that 4 new values are added to the data, but only 2 are used for the sample from population #1 and the other 2 are used for the sample from population #2.
* Why adjust the estimates for π1 and π2? It helps to move the estimate away from 0 or 1. In the end, the confidence interval has better “coverage”. This means the true percentage of times the Agresti and Caffo CI will contain π1 – π2 will be closer to the (1-α)100% confidence level than the true percentage of times the Wald CI will contain π1 – π2.
* There is an interval for π1 – π2 similar to the Wilson interval. This “score” interval finds all possible values of π1 – π2 that satisfy the test statistic and critical values in a hypothesis test. Unfortunately, there are no closed form expressions that can be written out like what we had for the Wilson interval. The diffscoreci() function in the PropCIs package can calculate this interval using iterative numerical methods.

Example: Larry Bird (Bird.R)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Second |   |
|   |   | Made | Missed | Total |
| First | Made | 251 | 34 | 285 |
| Missed | 48 | 5 | 53 |
|   | Total | 299 | 39 | 338 |

Previously, we examined this data and considered it the “entire population”. This is not the best thing to do. Instead, it is better to consider this as a sample from the population of all Bird’s free throw attempts ever taken or will be taken (if he was still playing).

Question: Is this a representative sample?

Consider the first free throw made as one population and first free throw missed as the second population. We are interested in estimating the second free throw probability of success for each first free throw outcome population. In other words,

π1 = probability of success (made) on the second attempt for the first attempt = made population

and

π2 = probability of success (made) on the second attempt for the first attempt = missed population

Then

 and 

Question: Suppose π1 – π2 = 0. What does this mean in terms of the problem?

Find the 95% confidence interval to estimate the difference between proportions (probabilities). Note that



and

.

The 95% confidence interval for π1 – π2 is:



-0.1035 < π1 – π2 < 0.0778

I am 95% confident that the difference between the probability of success for the 2nd free throw attempts is between -0.1035 and 0.0778 for successful vs. unsuccessful first attempts.

Below are the calculations from R:

> y1 <- 251

> n1 <- 285

> y2 <- 48

> n2 <- 53

> alpha <- 0.05

> pi.tilde1 <- (y1+1)/(n1+2)

> pi.tilde2 <- (y2+1)/(n2+2)

> var.AC <- pi.tilde1\*(1-pi.tilde1) / (n1+2) +

 pi.tilde2\*(1-pi.tilde2) / (n2+2)

> lower.AC <- pi.tilde1 - pi.tilde2 - qnorm(p = 1-alpha/2)

 \* sqrt(var.AC)

> upper.AC <- pi.tilde1 - pi.tilde2 + qnorm(p = 1-alpha/2)

 \* sqrt(var.AC)

> data.frame(lower.AC, upper.AC)

 lower.AC upper.AC

1 -0.1035325 0.07781192

Below is an easier way to do the calculations, but you need to be careful with it:

> #Wald CI

> prop.test(x = c(y1, y2), n = c(n1, n2), conf.level =

 0.95, correct = FALSE)

2-sample test for equality of proportions without continuity correction

data: c(y1, y2) out of c(n1, n2)

X-squared = 0.2727, df = 1, p-value = 0.6015

alternative hypothesis: two.sided

95 percent confidence interval:

 -0.11218742 0.06227017

sample estimates:

 prop 1 prop 2

0.8807018 0.9056604

> #Agresti-Caffo CI

> prop.test(x = c(y1, y2) + 1, n = c(n1, n2) + 2,

 conf.level = 0.95, correct = FALSE)

2-sample test for equality of proportions without continuity correction

data: c(y1, y2) + 1 out of c(n1, n2) + 2

X-squared = 0.0723, df = 1, p-value = 0.7879

alternative hypothesis: two.sided

95 percent confidence interval:

 -0.10353254 0.07781192

sample estimates:

 prop 1 prop 2

0.8780488 0.8909091

Note that ONLY the confidence intervals should be used from the above output. The hypothesis testing differences of proportions should not be used due to the adjusted data values.

Questions:

* Is π1 – π2 = 0?
* Are first and second free throw attempt results independent?

What about hypothesis testing?

Tests can be performed using confidence intervals, test statistics, and p-values in a similar manner as we have seen earlier. We just now use different formulas for confidence intervals and test statistics than we did in the last chapter.

Example: Larry Bird (Bird.R)

Are the first and second attempt outcomes independent? In other words, is π1 – π2 = 0?

Hypothesis test using the CI method:

1. Ho: π1 – π2 = 0
 Ha: π1 – π2 ≠ 0
2. -0.1035 < π1 – π2 < 0.0778
3. Because 0 is in the CI, do not reject Ho.
4. There is not sufficient evidence to conclude that the outcome on the first throw has an effect on the outcome of the second free throw.

Of course, you could also do the hypothesis test using the test statistic and p-value methods.

We can write the hypotheses for a two-tail test as

Ho: π1 – π2 = 0
Ha: π1 – π2 ≠ 0

and left-tail or right-tail tests would have the usual adjustments. Notice that I use 0 as the hypothesized value because this value will be of interest in almost all applications.

Test statistic: Suppose  and . One can show that



because



has an approximate standard normal probability distribution. What is the test statistic?



where  was substituted for π1 and π2 in the denominator. Now, we need to estimate πp. Because π1 = π2 under Ho, we could say that this means both populations are the same (because this is the only parameter for both). Thus, we could then “pool” the results from both samples to form an estimator for πp:



The final test statistic becomes,



What are the critical values for two-tail, left-tail, and right-tail tests?

P-value for two-tail test: 2×P(Z > |z|)

What is the p-value for a left-tail or right-tail test?

When is the normal probability distribution APPROXIMATION for Z good enough to be used?

My recommendation is to look at the cell counts in the contingency table representation of the data. If there are any very small cell counts (say, < 5), you may want to investigate further if the normal approximation works. This investigation can be performed using exact inference procedures discussed in a categorical data analysis course.

Example: Larry Bird (Bird.R)

Test statistic method:

1. Ho: π1 – π2 = 0
 Ha: π1 – π2 ≠ 0
2. Note that . Then



1. ±z0.025 = ±1.96
2.



 Because -1.96 < -0.5222 < 1.96, do not reject Ho.

1. There is not sufficient evidence to conclude that the outcome on the first throw has an effect on the outcome of the second free throw.

P-value method:

1. Ho: π1 – π2 = 0
Ha: π1 – π2 ≠ 0
2. 2×P(Z > |-0.5222|) = 0.6015
3. α = 0.05
4. Since 0.6015 > 0.05, do not reject Ho.
5. There is not sufficient evidence to conclude that the outcome on the first throw has an effect on the outcome of the second free throw.

Below are the results from bird.R.

> pi.hat.p <- (y1 + y2)/(n1 + n2)

> z.stat <- (pi.hat1 - pi.hat2)/sqrt(pi.hat.p\*(1-pi.hat.p)

 \* (1/n1 + 1/n2))

> z.crit <- qnorm(p = 1-alpha/2, mean = 0, sd = 1)

> pvalue <- 2\*(1 - pnorm(q = abs(z.stat), mean = 0, sd =

 1))

> data.frame(pi.hat1, pi.hat2, z.stat, z.stat^2, z.crit,

 pvalue)

 pi.hat1 pi.hat2 z.stat z.stat.2 z.crit pvalue

1 0.8807018 0.9056604 -0.5222416 0.2727363 1.959964 0.6015021

Another way to perform the calculations in R:

> prop.test(x = c(y1, y2), n = c(n1, n2), conf.level =

 0.95, correct = FALSE, alternative = "two.sided")

2-sample test for equality of proportions without continuity correction

data: c(y1, y2) out of c(n1, n2)

X-squared = 0.2727, df = 1, p-value = 0.6015

alternative hypothesis: two.sided

95 percent confidence interval:

 -0.11218742 0.06227017

sample estimates:

 prop 1 prop 2

0.8807018 0.9056604

The test statistic given is called X-squared and this can be compared to a chi-square distribution critical value (to be discussed later in the course). The square root of the statistic is our z, where the + or - sign of z is the sign of . This equivalence result would be shown in a mathematical statistics class where you would learn that a squared standard normal random variable is the same as a chi-square random variable with 1 degree of freedom.

Questions:

* Are first and second free throw attempts independent?
* Suppose the purpose of the problem was changed to:

Perform a hypothesis test to determine if the probability of success for the second free throw attempt decreases if the first throw is missed rather than made.

This is what most basketball fans probably think. How would you perform the test?

1. Ho: π1 – π2 ≤ 0
 Ha: π1 – π2 > 0

Notice that in Ha, π1 > π2. Thus, the probability of success for the second free throw in the first free throw is made population **is greater than** the probability of success for the second free throw in the first free throw is missed population. If we reject Ho to conclude Ha is true, the probability of an incorrect decision (type I error) is α.

1. z = -0.5222

1. +z0.05 = +1.645

1. Because -0.5222 < 1.645, do not reject Ho

1. There is not sufficient evidence to conclude that the probability of success on the second throw decreases when the first free throw is missed rather than made.

The p-value is P(Z > -0.5222) = 0.6992. Below are the calculations from R:

> z.crit <- qnorm(p = 1-alpha, mean = 0, sd = 1)

> pvalue <- 1 - pnorm(q = z.stat, mean = 0, sd = 1)

> data.frame(pi.hat1, pi.hat2, z.stat, z.crit, pvalue)

 pi.hat1 pi.hat2 z.stat z.crit pvalue

1 0.8807018 0.9056604 -0.5222416 1.644854 0.6992489

> prop.test(x = c(y1, y2), n = c(n1, n2), conf.level =

 0.95, correct = FALSE, alternative = "greater")

2-sample test for equality of proportions without continuity correction

data: c(y1, y2) out of c(n1, n2)

X-squared = 0.2727, df = 1, p-value = 0.6992

alternative hypothesis: greater

95 percent confidence interval:

 -0.09816334 1.00000000

sample estimates:

 prop 1 prop 2

0.8807018 0.9056604

* What if the number of successes were not already summarized? Suppose the “raw” data is in a data frame called raw.data2:

> head(raw.data2)

 first second

1 made made

2 made made

3 made made

4 made missed

5 missed made

6 made made

> tail(raw.data2)

 first second

333 made made

334 made made

335 made made

336 made made

337 made made

338 made made

> cont.table <- table(raw.data2)

> cont.table

 second

first made missed

 made 251 34

 missed 48 5

> y1 <- cont.table[1,1]

> y2 <- cont.table[2,1]

> n1 <- cont.table[1,1] + cont.table[1,2]

> n2 <- cont.table[2,1] + cont.table[2,2]

> data.frame(y1, y2, n1, n2)

 y1 y2 n1 n2

1 251 48 285 53

* Should we be concerned about the normal distribution approximation for this data set? Notice that there is a cell count of size 5 in the contingency table. Using a permutation test method that would be taught in a categorical data analysis course, we can visualize the sampling distribution of the test statistic:





There is a lot of discreteness in the distribution. However, it appears the standard normal distribution works relatively well (even though there is one high bar in the histogram).