**Section 1.2.3 – Test for the difference of two probabilities**

Confidence intervals are preferred over hypothesis tests when a simple set of parameters, like π1 – π2, are of interest. This is because a confidence interval gives a range of possible parameter values, which a hypothesis test cannot.

Still, if you want to perform a hypothesis test of H0:π1 – π2 = 0 vs. Ha:π1 – π2 ≠ 0 using a test statistic and p-value, one way is to use the score test statistic of



where . This test statistic has an approximate standard normal distribution for a large sample. Therefore, you can reject H0 if |ZS| > Z1-α/2.

Notes:

* + -  is included because this is the MLE for π1 and π2 when π1 – π2 = 0 (i.e., when the null hypothesis is true).
    - A general version of the test statistic for H0:π1 – π2 = d vs. Ha: π1 – π2 ≠ d is

,

where  and  denote the MLEs of 1 and 2 under the constraint that π1 – π2 = d. This version is usually not of interest except for when one wants to invert it to find a confidence interval.

Another way to perform the hypothesis test is through using a Pearson chi-square test. This is a general testing procedure that can be used for a large number of problems including the test of interest here.

The general form of a test statistic of this type is



which is formed across all cells of the contingency table. The estimated expected count is what we would expect a cell count to be if the null hypothesis was true. For our test in a 2×2 table,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Response | |  |
|  |  | Success | Failure |  |
| Group | 1 | w1 | n1 – w1 | n1 |
| 2 | w2 | n2 – w2 | n2 |
|  |  | w+ | n+ – w+ | n+ |

the estimated expected count is  for the success column and  for the failure column. The test statistic is



Through using some algebra, we can simplify the test statistic to be



For large samples, this test statistic has an approximate  probability distribution. Large values indicate evidence against the null hypothesis so that we reject H0 if 

Question: Why do large values indicate evidence against the null hypothesis?

One can show that ! Also, because the square of a standard normal random variable is the same as chi-square random variable (e.g., ), the Pearson chi-square test is the same as the score test here!

Example: Larry Bird (Bird.R)

Below is some code and output that was shown earlier.

> c.table <- array(data = c(251, 48, 34, 5), dim = c(2,2), dimnames = list(First = c("made", "missed"), Second =

c("made", "missed")))

> c.table

Second

First made missed

made 251 34

missed 48 5

The purpose of this problem is to test H0: π1 – π2 = 0 vs. Ha: π1 – π2 ≠ 0 (1 = First FT made, 2 = First FT missed)

Note that . Then



Because -1.96 < -0.5222 < 1.96, do not reject H0 when α = 0.05. There is not sufficient evidence to conclude that the probability of success on the second free throw differs based on what happened for the first free throw.

New code and output:

> prop.test(x = c.table, conf.level = 0.95, correct = FALSE)

2-sample test for equality of proportions without continuity correction

data: c.table

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

Note that R gives . Also, R gives the Wald confidence interval in its output.

There are other ways to perform this test in R. For example,

> chisq.test(x = c.table, correct = FALSE)

Pearson's Chi-squared test

data: c.table

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

Another way to perform a test for the difference of two probabilities is through a likelihood ratio test (LRT). The test statistic is



Notice the similar format of the statistic compared to what we saw for a test involving a single probability of success parameter π. The null hypothesis is rejected if .

Larntz (1978) showed that the score (Pearson) test is better to use than the LRT in these situations.