**Section 1.1.2 – Inference for the probability of success (continued)**

Hypothesis tests for π

When only one simple parameter is of interest, such as π, confidence intervals are generally preferred over hypothesis tests. 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:π = π0 vs. Ha:π ≠ π0, one way is to use the test statistic of



as mentioned earlier when discussing the Wilson confidence interval. Reject H0 if |ZS| > Z1-α/2. This type of test is usually called a “score” test. Score tests are general likelihood-based inference procedures – please see Appendix B. Because the Wilson interval results from a score test, the interval is often referred to as a “score interval” as well.

Another way to perform a hypothesis test of H0:π = π0 vs. Ha:π ≠ π0 is a likelihood ratio test (LRT). Because this test is VERY frequently used in the analysis of categorical data (beyond testing for π), we will spend some time discussing it here.

The LRT statistic, Λ, is the ratio of two likelihood functions. The numerator is the likelihood function maximized over parameters restricted under the null hypothesis. The denominator is the likelihood function maximized over the parameters restricted by null and alternative hypotheses. The test statistic is written as:



Wilks (1935, 1938) shows that –2log(Λ) can be approximated by a  for a large sample and under H0 where u is the difference in dimension between the alternative and null hypothesis parameter spaces. See Appendix B for more background on the LRT.

Questions:

* Suppose Λ is close to 0, what does this say about H0? Explain.
* Suppose Λ is close to 1, what does this say about H0? Explain.
* When using –2log(Λ), will large or small value values indicate H0 is false?

Example: Field goal kicking (no program)

Continuing the field goal example, suppose the hypothesis test H0:π = 0.5 vs. Ha:π ≠ 0.5 is of interest. Remember that w = 4 and n = 10.

The numerator of Λ is the maximum possible value of the likelihood function under the null hypothesis. Because π = 0.5 is the null hypothesis, the maximum can be found by just substituting π = 0.5 in the likelihood function:



Then



The denominator of Λ is the maximum possible value of the likelihood function under the null OR alternative hypotheses. Because this includes all possible values of π here, the maximum is achieved when the MLE is substituted for π in the likelihood function! As shown previously, the maximum value is 0.001194.

Therefore,



Then -2log(Λ) = -2log(0.8179) = 0.4020 is the test statistic value. The critical value is  = 3.84 using α = 0.05:

> qchisq(p = 0.95, df = 1)

[1] 3.841459

There is not sufficient evidence to reject the hypothesis that π = 0.5.

After recording the video: Research papers, textbooks, and data analysts in general are not consistent with what “LRT statistic” refers to. Some will have it refer to Λ alone while others will have it refer to the transformed statistic -2log(Λ). My co-author and I decided to use the following meanings in the book:

Λ: This is the LR statistic

-2log(Λ): This is the LRT statistic because it is compared to the chi-square distribution quantile during the test

This standardizing of the terminology was one of our last changes to the book before submission to the publisher. Because my notes and videos were created prior to this decision on standardization, they do not reflect it.