One-Sample Hypothesis Testing

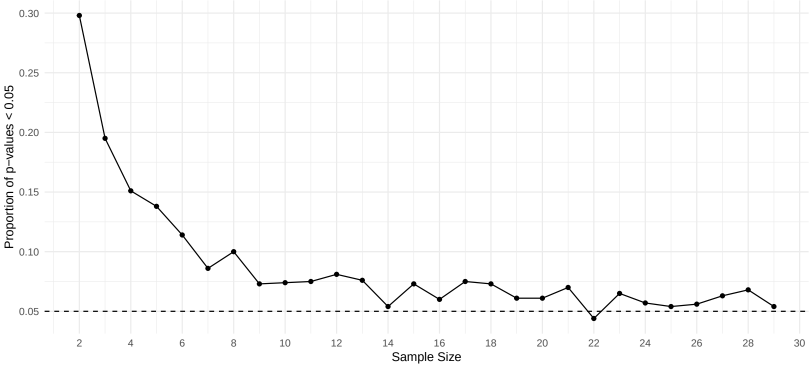
Part 3 - the T-test

Ryan Miller

Grinnell College
Statistics

Introduction

Suppose we repeatedly sample *quantitative data* from a population *where H_0 is true*, the graph below shows how often Z-test produces a p-value less than 0.05. Is this is a problem?

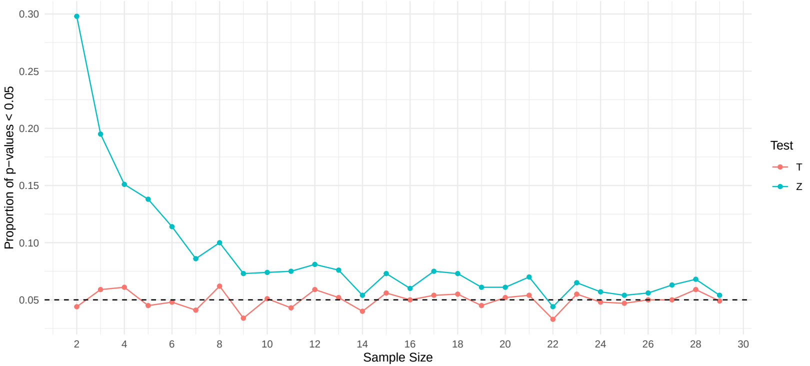


Limitations of the Z-test

- It is a problem... for one-sample quantitative data, the Z-test *systematically underestimates* the actual p-value for small sample sizes
  - When H_0 is true, we'd expect to see a p-value less than 0.05 only 5% of the time
- For small samples of quantitative data, the Z-test produces smaller p-values than it should, which makes it an unreliable decision making tool
  - In these scenarios we should use a modified procedure known as the T-test

The T-test

For one-sample quantitative data, here are how the Z-test and T-test compare:



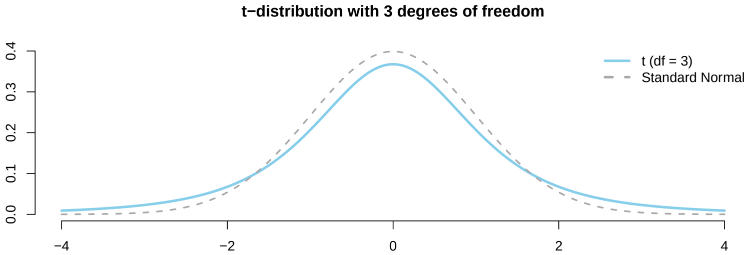
Grinnell College
Statistics

The T-test

- For one-sample *categorical data*, CLT gives us a standard error formula that only depends upon the hypothesized value, p
  - $Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$
- For one-sample *quantitative data*, the SE formula includes σ (the standard deviation describing all cases in the population), which is typically unknown
  - $Z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}}$
- To make the Z-test work, we need to estimate σ using the standard deviation of the sample, s.
  - However, because s differs from sample to sample, estimating σ introduces extra variability into our Z-score calculation which isn't properly accounted for by the Normal distribution

The T-test (cont.)

The t-distribution modifies the Normal curve to account for this extra uncertainty using a parameter known as *degrees of freedom*, or df:



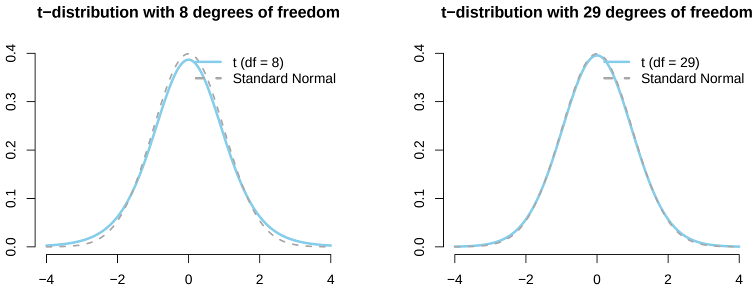
When working with a single mean, $df = n - 1$, which is because the calculation of the sample standard deviation involves the sample

dependent pieces of information are available

Grinnell College
Statistics

The T-test (cont.)

As the sample size increases, so do the degrees of freedom, and the *t*-distribution approaches the Normal distribution:



Some History on the T-test

- The T-test was first developed by William Gosset, a statistician working for Guinness Brewing
  - Gosset's work involved studying small samples to improve quality control, which exposed him to the unexpected behavior of the Z-test in certain circumstances
  - Gosset took a leave of absence from Guinness to study under the well-known statistician Karl Pearson
- In science it is common for the creator of a method to name it after themselves
  - Guinness forced Gosset to publish his work under a pseudonym, so Gosset named the distribution he developed "Student's t-distribution"
  - The T-test is now one of the most widely used statistical procedures

T-test Example

In a previous lab, you tested the hypothesis $H_0 : \mu = 120$ using the average systolic blood pressure of a sample of $n = 200$ ICU patients. The sample mean and standard deviation were $\overline{x} = 132.28$ and $s = 32.95$, respectively.

1) Identify the expected value and SE used in the test statistic
2) Calculate the test statistic (T-value)
3) Use the "t" menu of StatKey (under Theoretical Distributions) to calculate the one-sided p-value

Guidelines and Conclusion

- The Z-test generally works fine for one-sample categorical data so long as the sample is large enough
  - At least 10 instances of each outcome, or $n \cdot p \geq 10$ and $n \cdot (1-p) \geq 10$
- The Z-test *should not be used* for one-sample quantitative data, and we should use the T-test instead
  - The T-test is *designed* to work for small samples from a Normally distributed population
  - It is also appropriate for large samples ($n \geq 30$), regardless of how the data are distributed
- Going forward, we will prioritize these tests over StatKey simulations, and our next lab will cover how to perform them in `R`