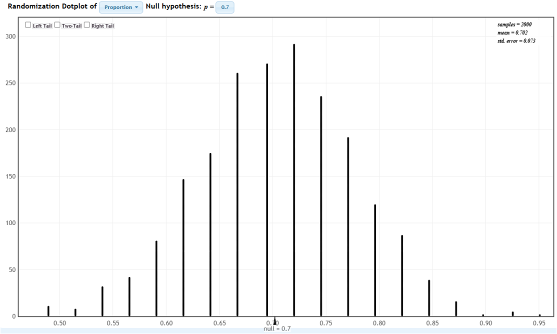
One-Sample Hypothesis Testing

Part 2 - Z and T Tests

Ryan Miller

Introduction

So far, we've relied upon *simulations* to determine the *null distribution* and *p*-value for our hypothesis tests.



You might have noticed that these simulations often produce a *bell-shaped* distribution.

Normal Distributions

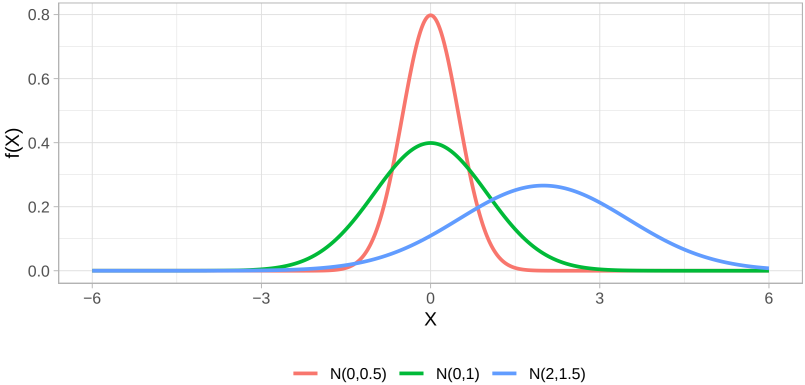
This shape is no coincidence, but rather a consequence of **Central Limit theorem** (CLT). But before we discuss CLT, we'll need to cover a few details about the **Normal distribution**:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- μ is the center (mean) of the distribution
- σ is the standard deviation of the distribution
- We use the shorthand $N(\mu,\sigma)$ to express a Normal distribution
  - For example, N(3,1) is a curve centered at 3 with a standard deviation of 1
- You don't need to know the formula for the Normal curve, but you should know that it depends on μ and σ

Normal Distributions

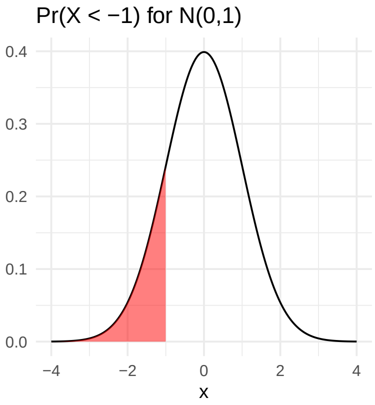
Below are three different Normal Distributions displayed on the same x-axis:

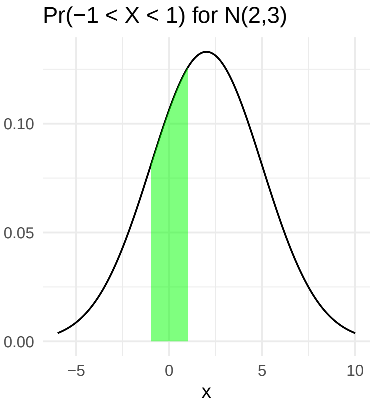


Grinnell College
Statistics

Probability and Normal Distribution

The Normal curve is a **probability distribution**, meaning the area under the curve can be used to model/calculate the probabilities of certain events:





Grinnell College
Statistics

Z-Scores

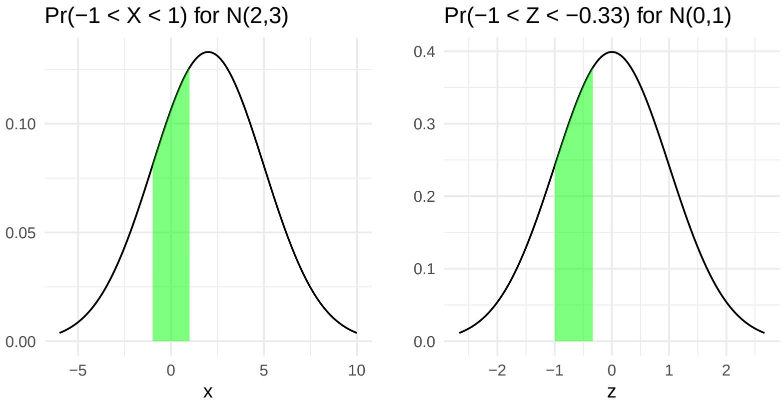
Because it is inconvenient to work with a distribution that uses a different scale in each new analysis statisticians frequently use a standardization approach known as the **Z-score transformation**:

$$Z_i = \frac{X_i - \text{Expected Value}}{\text{Standard Deviation}}$$

- For example, an ICU patient in the data from our previous lab had a systolic blood pressure of 162.
  - The mean systolic blood pressure of the entire sample was 132.28, and the sample standard deviation was 32.95
- So, this individual's Z-score is $Z = \frac{162-132.28}{32.95} = 0.90$
  - Meaning they are almost 1 standard deviation above the sample average

Z-Scores and Probability

Z-score transformations allow us to use the $N(0, 1)$ curve as a probability model for any scenario, regardless of the measurement units:



Central Limit Theorem

Mathematically, **Central Limit theorem** (CLT) states:

$$\lim_{n\to\infty} \sqrt{n}\left(\frac{\overline{x}-\mu}{\sigma}\right) \to N(0,1)$$

In more practical terms (using slightly informal notation), CLT suggests any random variable that is an average of sufficiently many independent observations will follow a Normal distribution with a predictable mean and standard deviation:

$$\overline{X} \sim N(\mu, \tfrac{\sigma}{\sqrt{n}})$$

Central Limit Theorem

In hypothesis testing, we can make this even more general:

$$\text{Sample Estimate} \sim N\big(\text{Expected value under } H_0, SE\big)$$

$$\implies \frac{\text{Sample Estimate} - \text{Expected value under } H_0}{SE} \sim N(0,1)$$

Here SE is the **standard error** of the sample estimate. We won't get into how these are derived, but CLT gives us the following SE formulas:

- $SE = \sqrt{\frac{p\cdot(1-p)}{n}}$ for a single proportion
- $SE = \frac{\sigma}{\sqrt{n}}$ for a single mean (note that σ is the standard deviation of cases in the population)

Grinnell College
Statistics

The Z-test

Central Limit theorem allows for a standardized hypothesis testing approach whenever we are studying a sample average or a sample proportion (which is just an average of 0's and 1's)

1. Apply the Z-score transformation to the observed sample mean or proportion using the null hypothesis and SE derived from CLT
2. Compare the resulting Z-score to a $N(0,1)$ distribution to find the p-value by looking at the area corresponding to Z-scores at least as extreme as the one from the sample data

Z-test Example

In our infant toy choice example, $H_0 : p = 0.5$ and we observed $\hat{p} = 14/16 = 0.875$, or 14 of 16 infants choosing the "helper". Carry out a one-sample Z-test by performing the following steps:

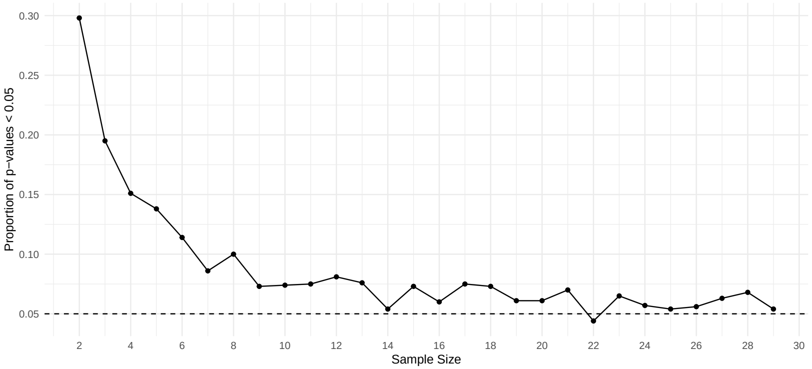
1) Identify the expected value and SE for the Z-transformation
2) Calculate the test statistic (Z-value)
3) Use the "Normal" menu of StatKey (under Theoretical Distributions) to calculate the one-sided p-value

Z-test Example (solution)

1) Under H_0, CLT implies $SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.5(1-0.5)}{16}} = 0.125$
2) Then, $Z = \frac{\hat{p}-p}{SE} = \frac{0.875-0.5}{0.125} = 3$
3) Using a $N(0,1)$ distribution, $Pr(Z \geq 3) = 0.0013$, which is the one-sided p-value. The two-sided p-value would be 0.0026 due to the symmetry of the Normal distribution

Problems with the Z-test

Consider $H_0 : \mu = 0$, if we sample data from a $N(0,1)$ (reflecting H_0 being true) here is the percentage of time the Z-test produces a p-value less than 0.05. Is this is a problem?



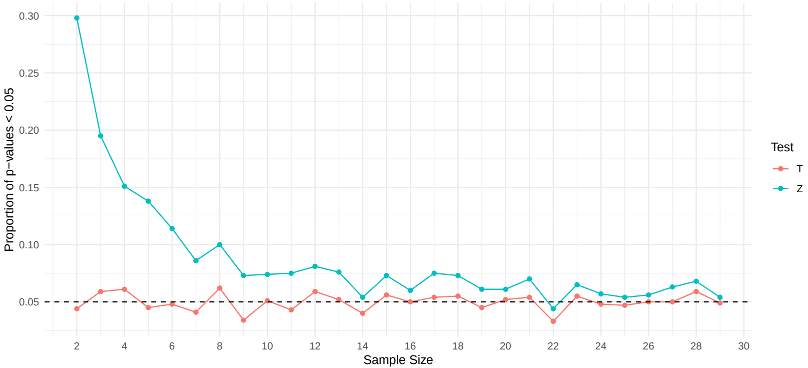
Grinnell College
Statistics

Problems with the Z-test (cont.)

- Yes! 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
  - The Z-test produces such p-values far more often than it should

The T-test

For one-sample quantitative data, we should use a similar procedure known as the T-test:

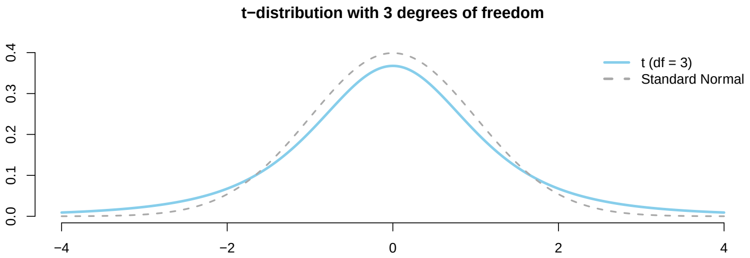


Grinnell College Statistics

The T-test

- For one-sample *categorical data*, CLT gives us a standard error formula that only depends upon the hypothesized value, p
- For one-sample *quantitative data*, the SE formula includes σ (the standard deviation describing all cases in the population), which is typically unknown
  - So, to make the Z-test work, we need to estimate σ using the standard deviation of the sample, s.
  - However, this step introduces extra variability into our Z-score calculation which isn't properly accounted for by the Normal distribution

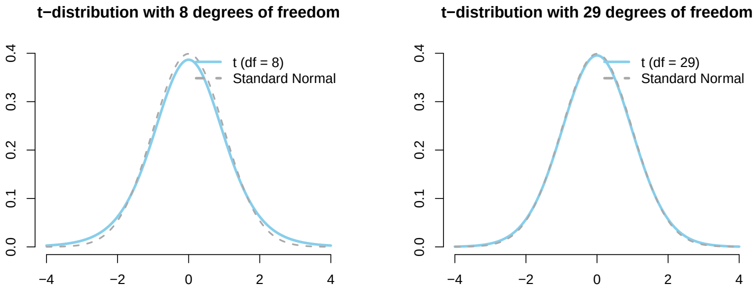
The *t*-distribution modifies the Normal curve to account for this extra uncertainty using a parameter known as *degrees of freedom*, or *df*:



When working with a single mean, $df = n - 1$

The T-test (cont.)

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



Grinnell College
Statistics

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 Question #5 of Lab 3, 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 132.28 and 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`