Hypothesis Testing Revist

February 23, 2021

Brief Review

- Distribution $\rightarrow$ Population $\rightarrow$ Sample
  - Sampling, bias, and confounding
  - Parameter vs statistic
- Hypothesis Testing
  - Null hypothesis
  - Type I/Type II errors
- Sampling Distribution
  - Variability of statistic
  - Central limit theorem
  - Construction of confidence intervals
- Last week's homework

The Plan

Today, we will look more closely at *p*-values and power $(1 - \beta)$. Make ourselves aware of some important details when beginning a study

Next week is instructional break. I will compile a document covering in more detail everything we have covered up to the end of today. I will be on zoom during class time to review and answer questions, attendance not necessary

For the rest of the semester, we will be looking at specific tests and applications with genetic data

Take home midterm at the end of March, plan time for final presentations (May 10-14)

Standard Normal, $Z \sim N(0,1)$

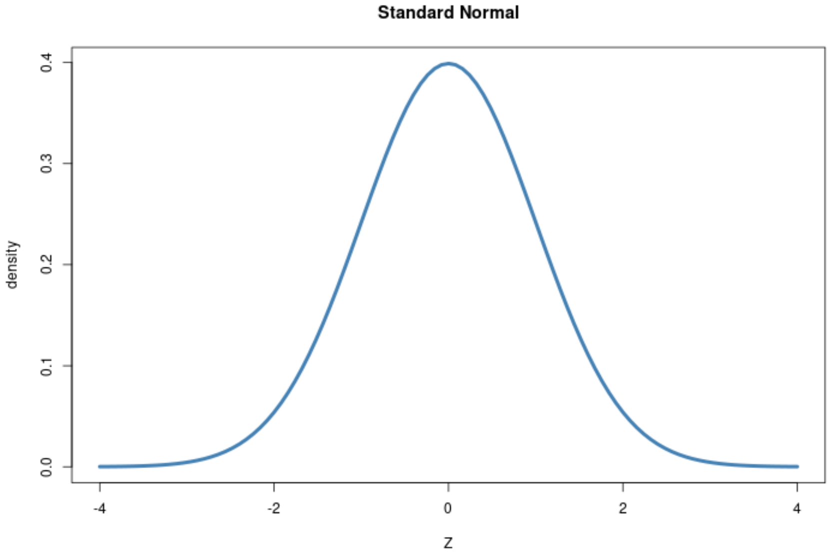
In the previous class, we were introduced to the standard normal distribution, Z, with the CLT, where

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

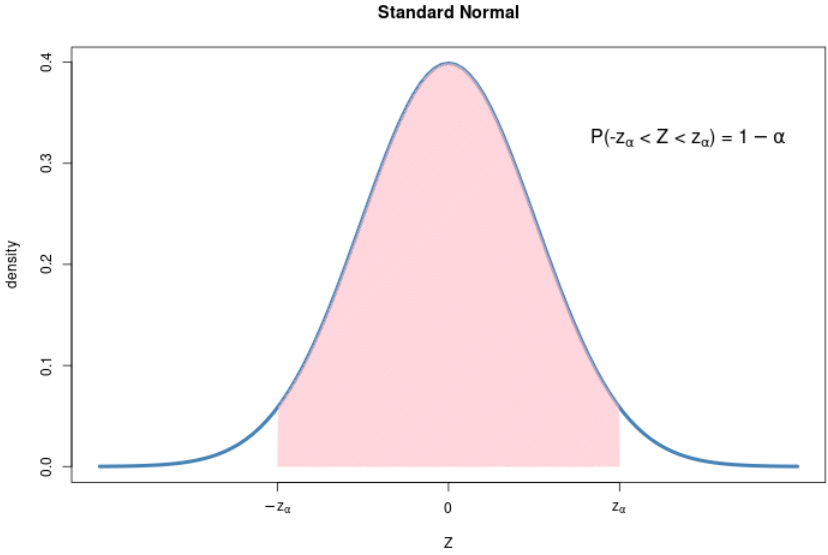
We may also recall the brief introduction of *critical values*, denoted z_α, which were chosen so that $P(-z_\alpha \leq Z \leq z_\alpha) = 1 - \alpha$.

Our interest now is how we might use this information in the context of hypothesis testing

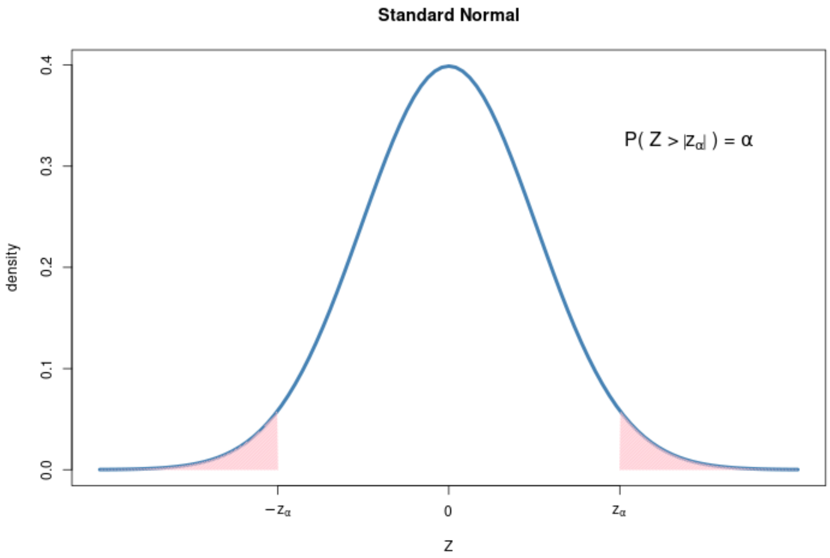
$N(0, 1)$



$N(0, 1)$



$N(0, 1)$



Generating test statistic

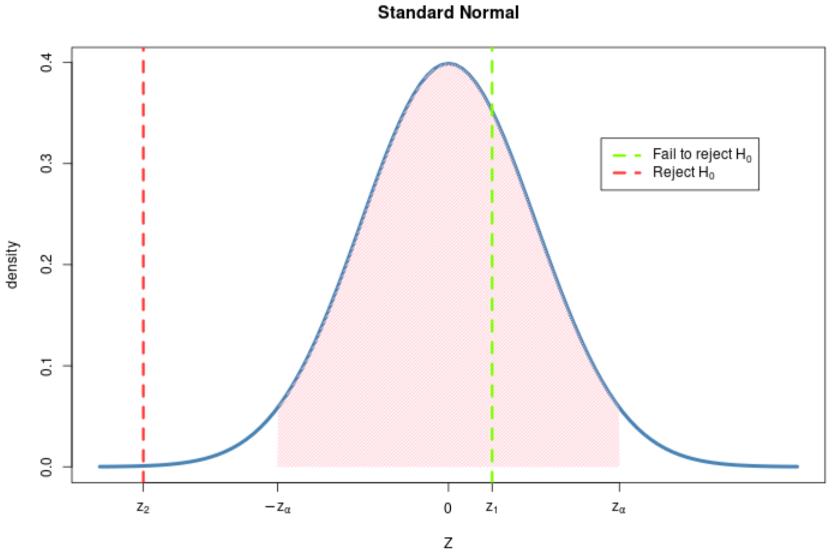
Under the null, with $H_0 : \mu = \mu_0$, we have that

$$\lim_{n \to \infty} \sqrt{n} \left(\frac{\overline{X} - \mu_0}{\sigma} \right) \to N(0, 1).$$

Assuming that σ is known, once we have collected our sample and computed our mean $\overline{x}$, we can punch in the known values to compute a *test statistic*

$$z = \sqrt{n} \left(\frac{\overline{x} - \mu_0}{\sigma} \right)$$

Test statistic



Weighing the Evidence

Given a level of α (which defines the size of the shaded region), we can come to the conclusion to either reject or fail to reject. However, it would also be nice to have some idea of the amount of evidence against our null hypothesis.

A natural candidate for this would be to consider the *magnitude* of deviation from the mean value of 0, which would be expressed $|z|$.

Treating this as a lower bound of deviation (that is, our deviation *is at least as great* as what we observed), we can now more succinctly define our p-value as

$$p = P(Z > |z|)$$

p-value

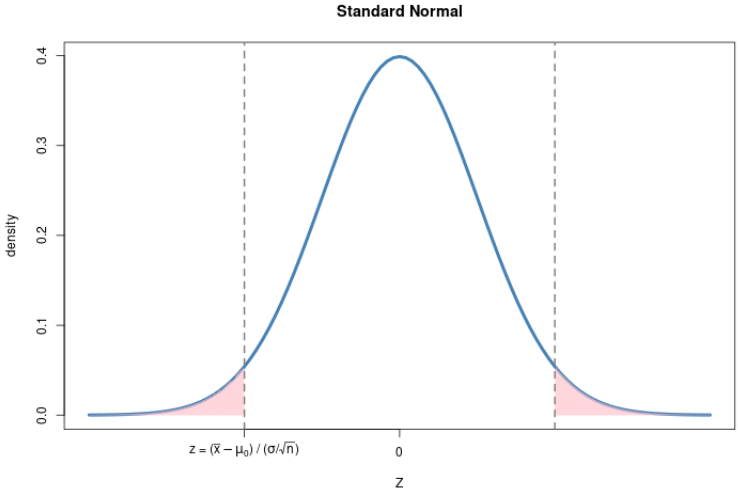
A few things to note about our value

$$p = P(Z > |z|)$$

- This is still under the assumption that H_0 is true; if it were not, $(\overline{x} - \mu_0)/(\sigma/\sqrt{n})$ would not be standard normal
- Z is the random variable. As such, this reads as "the probability of observing our data, *or something greater*"
- As this value is concerned with magnitude, we do not care if the observed z is positive or negative. This is known as a *two-sided hypothesis test*

p-value visualized

We can think of the p-value as the area under this curve



Power

Recall that our probability of committing a Type II error is given as

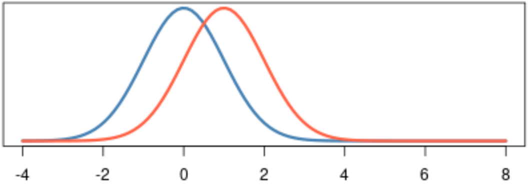
$$\beta = P(\text{Fail to reject } H_0 \mid H_0 \text{ is false})$$

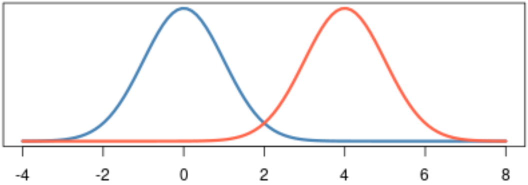
Our power, then, is the probability of correctly rejecting our null, written $1 - \beta$.

Given a statistical test, there are three things that impact our power

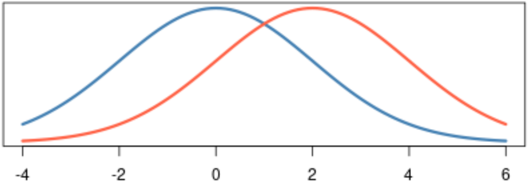
- Magnitude of departure of the observed data from the null. This is often referred to as the *effect size*, δ
- The variability of the population or response being studied
- Sample size

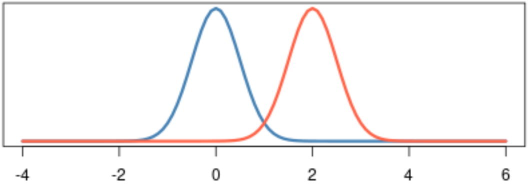
Magnitude of departure



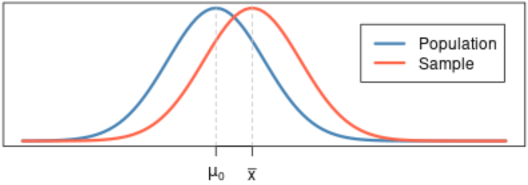


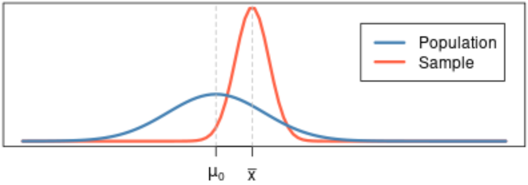
Variability of Population





Sample Size





Things to Consider

Prior to starting a study, it's important to determine a meaningful effect size, δ. A treatment wouldn't be worth pursuing, for example, if it was certain to have an effect on blood pressue that only lowered pressue 2 points

We should also try to anticipate the amount of variability we might see, usually an informed guess based on prior studies

Having determined a meaningful effect size and an estimate of variability, a *sensitivity analysis* can be performed to estimate power based on a number of sample sizes and different estimates of variability

Remarks on sample sizes

Some statistical test are more powerful than others, which can reduce the number of subjects needed. However, this is at the cost of greater assumptions

Attrition needs to be accounted for in all studies. Some subjects drop out, others lost for follow up, etc.,.

Depending on the cost of the study, one may have to concede the need for a larger effect size to account for a smaller number of participants

Sample sizes that are too large can be unnecessarily expensive, and samples that are too small may not be powerful enough, wasting resources. It's therefore often of critical importance to determine a reasonable sample size at the onset

References

- Dr. Deborah Dawson's GENE:6234 course notes
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
- Casella, G. and Berger, R. (2002) Statistical Inference. 2nd Edition, Duxbury Press, Florida.