Review session 2: notes on power and sample size

29 April, 2020

Recall the simple conditions for inference

Recap - confidence intervals and testing

How confidence intervals behave

Recall the form of a CI:

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

Where $z^* \frac{\sigma}{\sqrt{n}}$ is the margin of error.

How confidence intervals behave

The margin of error gets smaller when:

- z^* is smaller (i.e., you change to a smaller confidence level). Thus, there is a trade-off between the confidence level and the margin of error.
- σ is smaller. You might be able to reduce σ if there is measurement error. Often times, the σ can't be reduced, it is just a characteristic of the population
- n is larger.

How hypothesis tests behave

- Statistical significance depends on sample size (since sample size determines the standard error of the sampling mean)
- Recall the form of the z-test:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\text{magnitude of observed effect}}{\text{size of chance variation}} = \frac{\textit{signal}}{\textit{noise}}$$

- The numerator quantifies the distance between what you observe in the sample and the null hypothesized parameter.
- The denominator represents the size of chance variations from sample to sample

How hypothesis tests behave

- Statistical significance depends on:
  - The size of the observed effect $(\bar{x} - \mu)$
  - The variability of individuals in the population (σ)
  - The sample size (n)
  - Your criteria for rejection the null (α)

If you obtain a small p-value it is not necessarily because the effect size is large.

Type I error, and Type II error in hypothesis tests

	H_a is true	H_0 is true
Reject H_0	Correct decision	Type I error (α)
Fail to reject H_0	Type II error (β)	Correct decision

This table should remind you of something we have seen before. . . .

Power

- The power is the chance of making the correct decision when the alternative hypothesis is true.
- Thus, it is the complement of β
- Power $= 1 - \beta$

	H_a is true	H_0 is true
Reject H_0	Correct decision	Type I error (α)
Fail to reject H_0	Type II error (β)	Correct decision

However, there are an infinite number of possible values that μ could assume that are not $= \mu_0$

Thus we must choose a value at which to evaluate the β and power for an alternative hypothesis...

When we evaluate β we do so at a single such value μ_1

Calculating power in R

Example of calculating power

Suppose you have a known standard deviation $\sigma = 1$. $H_0 : \mu = 0$ vs. $H_a : \mu > 0.8$ and choose $\alpha = 0.05$. Calculate the power when $n = 10$.

You can calculate the minimum z-value required to reject H_0:

```
qnorm(p = 0.05, mean = 0, sd = 1/sqrt(10), lower.tail = F)
```

```
## [1] 0.5201484
```

So for any z-test with this value or higher, you will reject H_0 in favor of H_a.

This is often called Z_α

Example of calculating power

Now suppose that H_a is true. The test will reject H_0 about what percent of the time when H_a is true? To calculate this probability, we take the value from the previous calculation and calculate the *probability* above its value under H_a:

```
pnorm(q = 0.5201484, mean = 0.8, 1/sqrt(10), lower.tail = F)
```

```
## [1] 0.8119132
```

Example of calculating power, illustrated

Size of the difference

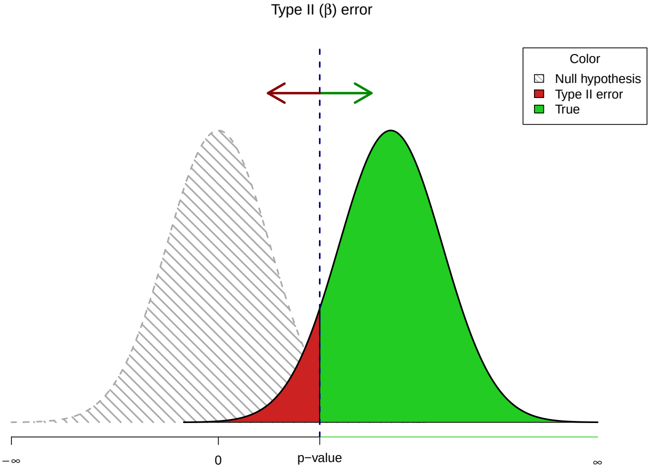
Effect of changing the μ_1

Imagine we our H_0 is a standard normal (mean=0, SD=1) and we set our α at 0.05.

If the true mean of our sampled population is 1.7 standard deviations above the μ_0,

what does our β look like?

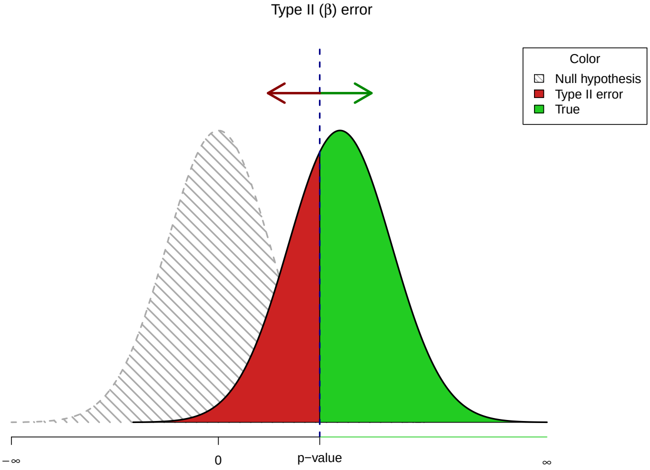
Effect of changing the μ_1



Effect of changing the μ_1

What happens if the “true” mean is closer to the Null?

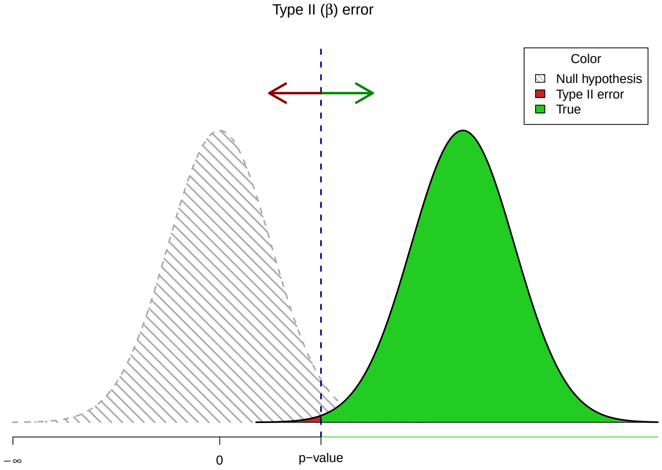
Effect of changing the μ_1



Effect of changing the μ_1

What happens if the "true" mean is further from the null?

Effect of changing the μ_1



Example

Lecture 21 has a worked example from the Pagano text that is worth reviewing Here we will go through another example from the Baldi and Moore textbook. This example assumes you are planning a quality control study to look at whether storage impacts the percieved sweetness of a beverage. Ten professional tasters will rate the sweetness on a 10 point scale before and after storage. We know that the standard deviation of sweetness ratings is $= 1$. We also know that a mean sweetness change of 0.8 on this scale is noticed by consumers. We want 90% power and an alpha of 0.05 for our study. We have a set of 10 values representing the difference in sweetness caused by storage.

What is the null hypothesis here?

What is our alternative?

Is our hypothesis one or two sided?

Example

We will start by finding the Z alpha:

$$Z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

```r
qnorm(.05)
```

```
## [1] -1.644854
```

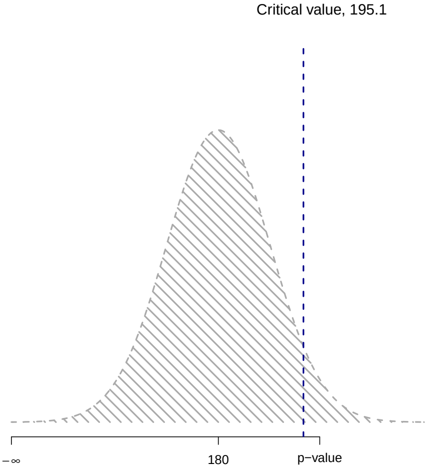
$$-1.645 = \frac{\overline{x} - 0}{\frac{1}{\sqrt{10}}}$$

Solve this for $\overline{x}$

$$\bar{x} = -1.645 \times \frac{1}{\sqrt{10}} = -0.522$$

null distribution

So here we have our null distribution with the value at which we reject the null



Example

We must choose a value at which to evaluate β. Here we will choose an alternate hypothesis that the mean sweetness difference is -0.8. Since we know a sample mean greater than -0.522 causes us to fail to reject H_0 we need to calculate the proportion of a distribution centered at 0.8 that would be below this value.

$$Z = \frac{-0.8}{\frac{1}{\sqrt{10}}}$$

$$Z = -0.253$$

Example:

Using R to calculate the probability,

```r
pnorm(-0.253, mean=0.8)
```

```
## [1] 0.1461705
```

Thus β P(do not reject null(0)|Null is false (true sweetness change is -0.8)) is $\sim$ 0.146

Remember that Power is 1-β = P(reject null | null is false)

In this example, Power is 1-0.146 or $\sim$ 0.854

Calculating sample size

sample size

When we are calculating sample size the steps we follow are: - Find the Z alpha and use this to calculate the value of our variable at which we would reject the null. - Find Z beta and use that to calculate what value this would be on the curve of the alternative hypothesis - Set these values equal to each other and solve for n

Example

For example, using our previous study of mean serum cholesterol levels, if we remember that we assumed the following:

$H_0 : \mu \leq 180 mg/100 ml$

α: 0.01

σ: 46

If the true population mean is as large as 211 and we want to risk only a 5% chance of failing to reject the null, so $\beta = 0.05$ and power would be $= 1 - \beta = 0.95$

Calculating Sample Size

We start by finding the Z value at which we would reject H_0 at $\alpha = 0.01$

We call this value Z_α

```r
qnorm(0.01, lower=FALSE)
```

```
## [1] 2.326348
```

Solve for $\overline{x}$

$$2.32 = \frac{\overline{x} - 180}{\frac{46}{\sqrt{n}}}$$

$$\overline{x} = 180 + 2.32(\frac{46}{\sqrt{n}})$$

Calculating Sample Size

Next we find the Z value at which we would reject H_A at $\beta = 0.05$

We call this value Z_β

```
qnorm(0.05)
```

```
## [1] -1.644854
```

Solve for $\overline{x}$

$$-1.645 = \frac{\overline{x} - 211}{\frac{46}{\sqrt{n}}}$$

$$\overline{x} = 211 - 1.645(\frac{46}{\sqrt{n}})$$

Calculating Sample Size

Because the value of $\overline{x}$ represents the same cutpoint for each of these, we can set the two equations equal to each other and solve for n.

$$180 - 2.32\left(\frac{46}{\sqrt{n}}\right) = 211 - 1.645\left(\frac{46}{\sqrt{n}}\right)$$

$$\sqrt{n}(211 - 180) = (2.32 - (-1.645)) * 46$$

$$n = \left(\frac{(2.32 + 1.645) * (46)}{(211 - 180)}\right)^2$$

$$n = 34.6$$

As we cannot include 0.6 of a person, the convention is to round up. So we would need 35 people in our sample.

Determining sample size for a margin of error

Example of calculation sample size

Body temperature has a known $\sigma = 0.6$ degrees F. We want to estimate the mean body temperature μ for healthy adults within ± 0.05 F with 95% confidence. How many healthy adults must we measure?

$$n = \left(\frac{z^*\sigma}{m}\right)^2$$

$$n = \left(\frac{1.96 \times 0.6}{0.05}\right)^2 = 553.2$$

We must recruit 554 (round up!) healthy adults for this study.

Sample size for a proportion

How big should the sample be to estimate a proportion?

Suppose that you want to estimate a sample size for a proportion within a given margin of error. That is, you want to put a maximum bound on the width of the corresponding confidence interval.

Let m denote the desired margin of error. Then $m = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

We can solve this equation for n, but we also need to plug in a value for p. To do that we make a guess for p denoted by p^*.

p^* is your best estimate for the underlying proportion. You might gather this estimate from a completed pilot study or based on previous studies published by someone else. If you have no best guess, you can use $p^* = 0.5$. This will produce the most conservative estimate of n. However if the true p is less than 0.3 or greater than 0.7, the sample size estimated may be much larger than you need.

How big should the sample be to estimate a proportion?

Rearranging the formula on the last slide for n, we get:

$$m = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\sqrt{n} m = z^* \sqrt{p(1-p)}$$

$$\sqrt{n} = \frac{z^*}{m} \sqrt{p(1-p)}$$

$$n = (\frac{z^*}{m})^2 p^*(1-p^*)$$

This last formula is the one we will use to estimate the required sample size.

Example of estimating sample size

Suppose after the midterm vote, you were interested in estimating the number of STEM undergraduate students who voted. First you need to decide what margin of error you desire. Suppose it is 4 percentage points or $m = 0.04$ for a 95% CI.

If you had no idea what proportion of STEM students voted then you let $p^* = 0.5$ and solve for n:

$$n = (\frac{z^*}{m})^2 p^*(1 - p^*) = (\frac{1.96}{0.04})^2 \times 0.5 \times 0.5 = 600.25 = 601$$

However, suppose you found a previous study that estimated the number of STEM students who voted to be 25%. Then what sample size would you need to detect this proportion?

$$n = (\frac{z^*}{m})^2 p^*(1 - p^*) = (\frac{1.96}{0.04})^2 \times 0.25 \times 0.75 = 450.19 = 451$$

Example of estimating sample size

What if you want the width of the 95% confidence interval to be 6 percentage points. What would m be in this case?