Lecture 15

Two-Sample T-Tests

Today's Updates / Reminders

- Lab 6 is due Friday. You will use R and practice hypothesis testing.
- Homework 7 due end of the day, the first class day back from spring break.
- Homework 8 opens today. Lab 7 opens tomorrow.
- Exam 2 is 3 weeks from today. Exam review will post later this week.
- Wednesday, March 25 is the last day to drop. Please see me if you wish to discuss this.
- STAT Hub is open for business! From 5:30-8pm, M-Th in SAS1101. You can park right in front of SAS if driving!

One-Sample Proportion (β)

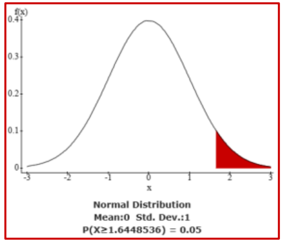
End of Lecture 14

Beta Example – Using Proportions

- A potato chip producer wonders whether the significance test of H_0: $p = 0.08$ versus H_a: $p > 0.08$ based on a random sample of 500 potatoes has enough power to detect a shipment with, say, 11% blemished potatoes. Use $\alpha = 0.05$.
  - Using null and alternative hypotheses, find critical z (z_0).
  - Find the proportion that corresponds with z^*.
  - Using this proportion, find its z with respect to p of 0.11.
  - Find area to the right.

Finding β

1. Use α = 0.05 to find z_α.



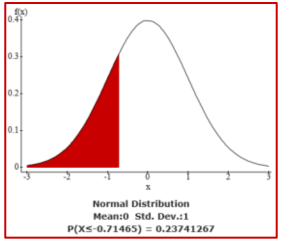
$$z_\alpha = 1.645$$

2. Using z_α, find the critical value. This is the limit of the proportion of potatoes we are willing to accept with blemishes.

$$p_{crit} = 1.645\sqrt{\frac{0.08 \cdot 0.92}{500}} + 0.08 \approx 0.1$$

$$z_\beta = \frac{0.1 - 0.11}{\sqrt{\frac{0.11 * 0.89}{500}}} = -0.71465$$

3. Now place yourself in the position of the alternative hypothesis. You believe the true proportion is 0.11. What is β (probability of failing to reject the null of 0.08 if you have a proportion of 0.11 potatoes with blemishes)?



β = P(z < -0.71465)

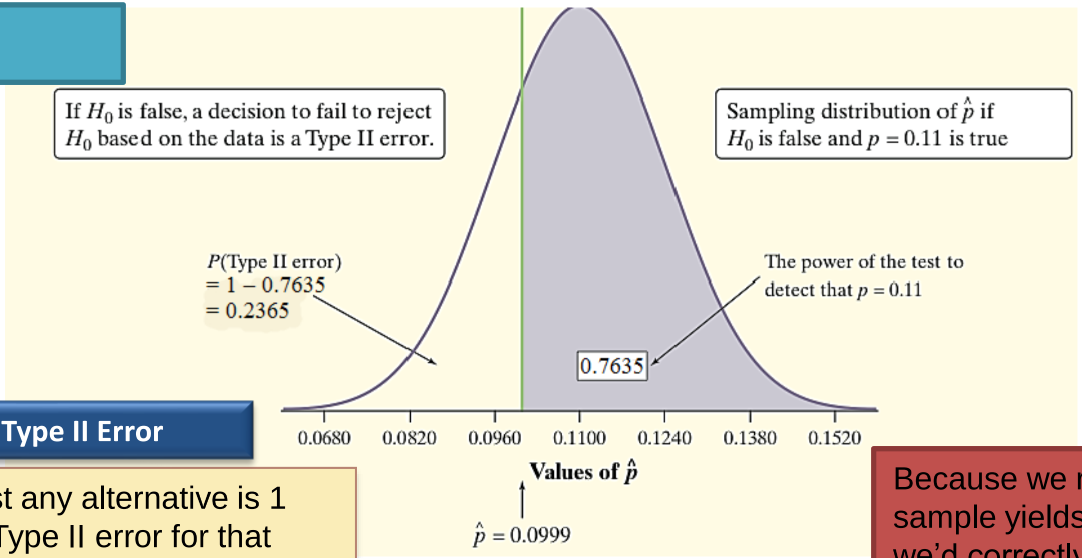
Power = P(z > -0.71465)

Power Example

- A potato chip producer wonders whether the significance test of H_0: p = 0.08 versus H_a: p > 0.08 based on a random sample of 500 potatoes has enough power to detect a shipment with, say, 11% blemished potatoes.

We would reject H_0 at α = 0.05 if our sample yielded a sample proportion to the right of the green line.

What if p = 0.11?



Power and Type II Error

The power of a test against any alternative is 1 minus the probability of a Type II error for that alternative; that is, power = 1 – β.

Because we reject H_0 at α = 0.05 if our sample yields a proportion > 0.0999, we'd correctly reject the shipment about 76% of the time.

Friday’s Lab

Discussion

A few points to note…

- You will need to remove any ages that are 98 or 99 from the list of data. 98 represents “don’t know” and 99 represents “refused to answer.”
- We will say, “Let’s keep all the ages that are 97 or less”
- Remember how to do this?
  - part1 <- data$age[(data$age<=97),]
- Then t.test accepts alternative=“two.sided”, “greater”, or “less”
- Otherwise, most of the lab is about understanding the basics of writing hypotheses and making conclusions based on R output.

Chapter 9.1

z-tests and CI for Difference Between Two Population Means

The Next 3 Lectures

- Section 9.1 – z-test for two means
  - When the population standard deviations, σ_1 and σ_2, are known – we will not talk much about this because it rarely exists
- Section 9.2 – t-test for two means
  - When the population standard deviations, σ_1 and σ_2, are NOT known, we will use 2-sample t-test
    - We must make an assumption about the variance of the two populations… equal or unequal?
- Section 9.3 – matched pairs
  - This is when we compare something to itself. Or we have data that come in natural pairs. This is usually done with “after minus before.”
- Section 9.4 – z-test for two proportions
  - When we compare two things that are expressed as a proportion

Types of Tests for Two Means

- The high-level summary of chapters 9.1 – 9.3
  - We can compare two different samples to see if there is a difference. “My roommate and I are on the NC State swim team. She is going to train in the mornings and I will train in the evenings and we will compare one of us against the other to see which time of day works better.” **Independent samples**
  - We can also compare two identical samples against each other. “I am on the NC State swim team. I will train in the morning this week and in the evenings next week and I will compare me to myself to see which time of day works better.” **Matched pairs**

Additional Types of Tests

	Two Samples are independent	Two Samples are dependent
$\bar{X}_1 - \bar{X}_2 \sim Normal$	Two-sample t-test	Paired t-test
$\bar{X}_1 - \bar{X}_2 \sim non - Normal$	Rank sum test	Signed rank test

- The first row are the ones we will address in this class.
- The second row is called non-parametric – this is not part of this class, but for those that are interested, some of the tests that do this work are the Mann-Whitney or Wilcoxon rank sum test and the Wilcoxon signed-rank test.

Set-Up

- We assume population 1 has a mean μ_1 and SD σ_1 and population 2 has a mean μ_2 and SD σ_2.
- Since neither μ_1 or μ_2 is known, we collect data.
- Note that we could say:
  - $\mu_1 > \mu_2$ but instead we will say, $\mu_1 - \mu_2 > 0$
  - $\mu_1 < \mu_2$ but instead we will say, $\mu_1 - \mu_2 < 0$
  - $\mu_1 \neq \mu_2$ but instead we will say, $\mu_1 - \mu_2 \neq 0$
- If we know that one is greater than the other from the outset, we may wish to state $\mu_1 - \mu_2$ is <, >, ≠ to D_0, where D_0 represents some presumed difference between the two populations.

Reminder of Rules for Mean and Variance

- If we wish to find $\mu_{\bar{X}_1-\bar{X}_2}$, we can use our rules that say:

$$\mu_{\bar{X}_1-\bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2}$$

- And the central limit theorem tells us that $\mu_{\bar{X}} = \mu$
- So we can say $\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2$
- Likewise, with variance, our rules say:

$$\sigma^2_{\bar{X}_1-\bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2}$$

- And the central limit theorem tells us that $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$
- So we can say $\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

See Lecture 5

Inferential Statistics

- Recall that if we don't know anything about the data being tested, we start with a confidence interval.
- If a claim has been made, we can do a hypothesis test.
- Confidence intervals are all the same:

$$point\ estimate\ \pm confidence\ factor * standard\ error\ of\ the\ mean$$

- In terms of this problem, we will say:

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \cdot \hat{\sigma}_{\bar{X}_1 - \bar{X}_2}$$

- Hypothesis tests are built as before $(sample - null)/standard\ error)$:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\hat{\sigma}_{\bar{X}_1 - \bar{X}_2}}$$

Usually assumed to be zero

- And we will assume that $\bar{X}_1 - \bar{X}_2 \sim N$ and the samples are independent

Example (9.1, #2 from text)

- The National Health Statistics Reports dated Oct. 22, 2008, included the following information on the heights (in.) for non-Hispanic white females:

Age	Sample Size	Sample Mean	Std. Error Mean
20–39	866	64.9	.09
60 and older	934	63.1	.11

- Calculate and interpret a 95% confidence interval for the difference between population mean height of younger women versus older women.

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \cdot \hat{\sigma}_{\bar{X}_1 - \bar{X}_2}$$

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Hypothesis Test

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\hat{\sigma}_{\bar{X}_1 - \bar{X}_2}}$$

Age	Sample Size	Sample Mean	Std. Error Mean
20–39	866	64.9	.09
60 and older	934	63.1	.11

- Set up a hypothesis test for this, assuming the difference in height is 0.

1) Write null and alternative hypotheses

2) Calculate the test statistic

3) Determine the p-value

4) Make a conclusion

Chapter 9.2

The Two-Sample t-Test and Confidence Interval

Going from z to t

- Recall, as we move from z to t, we move from σ to s.
- We shift from the z-table to the t-table.
  - As we make this shift to the t-table, we must begin to think about degrees of freedom. For a one-sample t-test, this was n – 1.
  - This becomes more complicated now.
- Since s is our estimate for population standard deviation, and we have s_1 and s_2 now, we use the standard error, $\hat{\sigma}_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
- To simplify the math, we can assume $s_1 = s_2$. But this assumption is often not valid. So sometimes, we have to go with $s_1 \neq s_2$.

Variance – Pooled vs. Unpooled

- If $\sigma_1^2 = \sigma_2^2$, then we use a common σ^2. This is called pooled variance because we pool the information from sample 1 and sample 2 to estimate σ.
- If $\sigma_1^2 \neq \sigma_2^2$, we use sample 1 to estimate σ_1 and we use sample 2 to estimate σ_2. This math is a bit more cumbersome. 🤯

Assumptions about Variance – Equal or Unequal?

- We can decide to pool the variance ($\sigma_1^2 = \sigma_2^2$), which means we make the assumption that the variance of the two populations are the same.
  - If we are comparing the tensile strength of two different batches of the same type of steel, this is likely a fair assumption
  - What we are saying when we pool the variance… "I think the bell curves for these two populations are roughly the same shape."
    - We will come back to this later in this section.
- We can decide to NOT pool the variance ($\sigma_1^2 \neq \sigma_2^2$). This is slightly more complex mathematically, but is usually the most conservative approach. We will start with this first.

Appendix: Test for Equal Variance

- If you want to know for sure, you can do a hypothesis test:

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_A: \sigma_1^2 \neq \sigma_2^2$$

- To test for homogeneity of variance (homoscedasticity), there are several tests (among others):
  - Levene
  - Bartlett
  - Brown-Forsythe
  - Cochran
  - Hartley F_{max}
- A p-value of less than 0.05 indicates a rejection of the null, which states that the variances are equal.
- In such cases, you would need to do a non-parametric analysis.

Unequal Variance

- If we go with $\sigma_1^2 \neq \sigma_2^2$, here is what happens:
  - We need two different estimators for σ_1^2 and σ_2^2. So we say s_1^2 and s_2^2 are our individual estimators.
  - Then, we get:

$$\hat{\sigma}_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

  - By itself, this looks simple, but unfortunately, this doesn't follow a t-distribution because we are estimating the standard deviations of two populations with the standard deviations of two samples of different size!! They would have two different degrees of freedom… 🤔

Satterthwaite's Approximation

$$\nu = df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

We will make (or not make) the assumption of equal variance many more times over the course of this class.

- By using this formula for degrees of freedom, we "correct" for the fact that the test statistic here does not have a t-distribution.
- This results in a decimal form of df. So we always round down (conservative). Most statistical software will calculate this for you.

Some textbooks use df equal to the smaller of $n_1 - 1$ and $n_2 - 1$. It's more conservative and avoids the nasty calculation above.

For purposes of tests, the questions will explicitly state whether variance is equal or unequal

Unequal Variance

- If we assume σ_1 and σ_2 are NOT equal… ($\sigma_1^2 \neq \sigma_2^2$):

$$\hat{\sigma}_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where $\upsilon = df = \dfrac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2+\frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2}$

Pooled Variance

- If we assume $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (populations have equal variance), we can get a pooled point estimator of σ^2 by combining the sum of squares from both samples.
- This leads to:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

- And since $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and we intend for s_p^2 to estimate σ^2, we get:

$$\hat{\sigma}_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

- And degrees of freedom will be: $(n_1 - 1) + (n_2 - 1)$

Equal Variance (Pooled Variance)

- If we assume σ$_1$ and σ$_2$ are equal… ($\sigma_1^2 = \sigma_2^2$):

$$\hat{\sigma}_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

where $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{(n_1-1)+(n_2-1)}$

and $\upsilon = df = (n_1 - 1) + (n_2 - 1)$

Example

- Are women more talkative than men? Here is the data from the paper… we'll use their nomenclature and state μ_1 = the daily number of words for men and μ_2 = the daily number of words for women
- So how do we use this data to make a conclusion about the population?

MEN	WOMEN
$n_1 = 186$	$n_2 = 210$
$\bar{x}_1 = 15668.5$	$\bar{x}_2 = 16215.0$
$s_1 = 8632.5$	$s_2 = 7301.2$

Assumptions about the Example

MEN	WOMEN
$n_1 = 186$	$n_2 = 210$
$\bar{x}_1 = 15668.5$	$\bar{x}_2 = 16215.0$
$s_1 = 8632.5$	$s_2 = 7301.2$

Other formula sheet

Are these normal????

Are these independent????

Is it reasonable to pool σ^2????

Confidence Interval

MEN	WOMEN
$n_1 = 186$	$n_2 = 210$
$\bar{x}_1 = 15668.5$	$\bar{x}_2 = 16215.0$
$s_1 = 8632.5$	$s_2 = 7301.2$

1. Pooled variance or not?
2. Calculate standard error.
3. Find df, or v.
4. Build the interval.

Hypothesis Test

MEN	WOMEN
n_1 = 186	n_2 = 210
$\bar{x}_1 = 15668.5$	$\bar{x}_2 = 16215.0$
s_1 = 8632.5	s_2 = 7301.2

1. Write the null and alternative hypotheses.

2. Pooled variance or not? Calculate standard error. Then calculate the test statistic.

3. Find df, or v. Then use the t-table.

4. Write conclusions.

When is it ok to use the t procedures?

Sample size	Use t procedures
n < 15	Ok if your **data look Normal distributed** (you can check this with a Normal Quantile Plot)
15 ≤ n < 40	Ok if data shows **some skewness.** Except in the presence of outliers or strong skewness.
n ≥ 40	Ok even if data are **strongly skewed.** Except in the presence of outliers.

Use different procedure or increase sample size

Investigate the cause: data wrongly recorded, equipment malfunction, Response bias

Important! No matter the sample size, t-test is not recommended if there are outliers in our data set!!