# Hypothesis Testing pt. 6

## More on Difference Tests for Means

Grinnell College

# Review – Testing

Hypothesis Testing Procedure

1. Construct null and alternate hypotheses, $H_0$ and $H_A$
2. Collect data and compute our sample statistic (i.e., $\overline{x}$)
3. Evaluate that statistic in the context of a null distribution, i.e.,

$$T = \frac{\overline{x} - \mu_0}{\hat{\sigma}/\sqrt{n}} \sim t_{df=n-1}$$

4. Reject or fail to reject hypothesis
   - Type I errors (false positive)
   - Type II errors (false negative)

# Review – Group Differences

Often in statistical inference, we are interested in investigating the *difference* between two or more groups

For example, we may have two groups, $A$ and $B$, with a mean value for each group, $\mu_A$ and $\mu_B$

Expressed in our null hypothesis, this equates to

$$H_0 : \mu_A = \mu_B \quad \text{or} \quad H_0 : \mu_A - \mu_B = 0$$

# Two-sampled t-test

There are a number of various assumptions we can make about the data for testing diff. in means, all resulting in slightly different tests (degrees of freedom and standard error):

1. Independent, groups same size and have same variance
2. Independent, groups have unequal sizes and similar variance
3. Independent, groups have different sizes and different variances
4. Paired testing

Of (1), (2), and (3)... (3) is the most versatile.

In general, we will concern ourselves with (3) and (4)

# Two-sampled t-test

There are a number of various assumptions we can make about the data for testing diff. in means, all resulting in slightly different tests (degrees of freedom and standard error):

1. ~~Independent, groups same size and have same variance~~
2. ~~Independent, groups have unequal sizes and similar variance~~
3. Independent, groups have different sizes and different variances
4. Paired testing

Of (1), (2), and (3)... (3) is the most versatile.

In general, we will concern ourselves with (3) and (4)

# Review – Hypothesis Test for Difference of Means

$H_0$: $\mu_1 - \mu_2 = \mu_0 = 0$

**Conditions:**

- Random Sample
- Normal population **OR** $n_1 \geq 30$ and $n_2 \geq 30$

**When $\sigma$ is not known:**

$$T := \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim \mathbf{t(df = min(}n_1,\ n_2\mathbf{)\ \text{-}\ 1)}$$

- use pt() function with value of T and df

# Degrees of Freedom

With our test-statistic from above:

$$T := \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{df},$$

there is actually controversy over how to calculate the df for this

**Option 1:** df $= min(n_1, n_2) - 1$ is the most *conservative* option

- gives us largest p-value, and thus weakest evidence of the 3 options
- simple: use this if you don't have access to computer for Option 3

**Option 2:** df $= n_1 + n_2 - 2$ is the least conservative

- gives us smallest p-value, and thus strongest evidence of the 3 options
- relies on assumption of $\sigma_1 = \sigma_2$ (how do we know this?)

# Degrees of Freedom

With our test-statistic from above:

$$T := \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{df},$$

there is actually controversy over how to calculate the df for this

**Option 3:** Satterthwaite Approximation

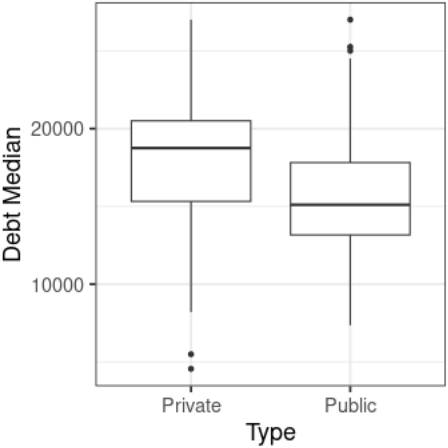
$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

- closest to the truth, but fractional value and hard to interpret!
- df will be inbetween values for the other options
- when we use R to do a t-test later on, look for fractional values of df

# Example – College data

Consider our college data, where we might investigate the differences in median debt upon graduate for public and private schools

- Private Schools
  - $\overline{x}_1 = 18028$
  - $\hat{\sigma}_1 = 3995$
  - $n_1 = 647$
- Public Schools
  - $\overline{x}_2 = 15627$
  - $\hat{\sigma}_2 = 3111$
  - $n_2 = 559$



# Example – College Data

Again, we will use R to compute this, utilizing a special “formula” syntax when using data.frames (will cover in lab)

```
> t.test(Debt_median ~ Private, college)

  Welch Two Sample t-test

data:  Debt_median by Private
t = 11.2, df = 1075, p-value <0.0000000000000002
alternative hypothesis: true difference in means between group
    Private and group Public is not equal to 0
95 percent confidence interval:
 1981.0 2820.6
sample estimates:
mean in group Private
                18028
 mean in group Public
                15627
```

- min(n1,n2)-1 = 559 - 1 = 558
- n1 + n2 - 2 = 647 + 559 = 1206

# Paired t-test

The **paired t-test** or **paired difference test** is a test for assessing differences in group means where the groups consist of the same subjects with multiple observations

While you may want to treat this as a two-sample t-test, in practice it more closely resembles that of a one-sample test:

$$T_{\text{paired}} = \frac{\overline{x}_d - \mu_0}{\hat{\sigma}_d / \sqrt{n}}$$

where $n$ represents the number of *unique* subjects and $\overline{x}_d$ and $\hat{\sigma}_d$ represent the mean and standard deviation of the *differences* between observations for each subject

# Paired t-test

Just as with the unpaired case, our null hypothesis is typically that

$$H_0 : \mu_d := \mu_1 - \mu_2 = \mu_0 = 0$$

Paired testing between groups allows us to control for within-subject variation, effectively reducing variation and making it easier to detect a true difference (power)

This comes at a cost, however – for $n$ subjects we are required to make $2n$ unique observations

# Example – French Institute

Consider the results of a summer institute program sponsored by the National Endowment for the Humanities to improve language abilities in foreign language high school teachers

Twenty teachers were given a listening test of spoken French before and after the program, with a maximum score of 36. We are interested in determining the efficacy of the summer institute

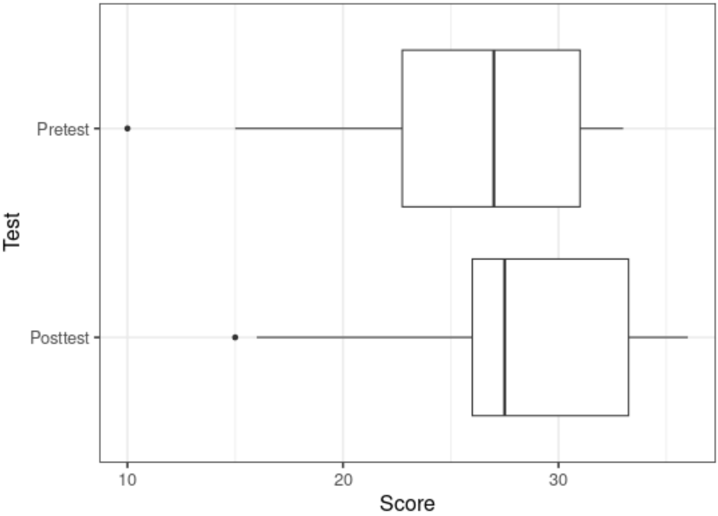
1. What is the null hypothesis for this study?
   - What would be a Type I error?
   - A Type II error?
2. How many total subjects do we have?
3. How many recorded observations do we have?

# Example – French Institute

The results of the tests are as follows:

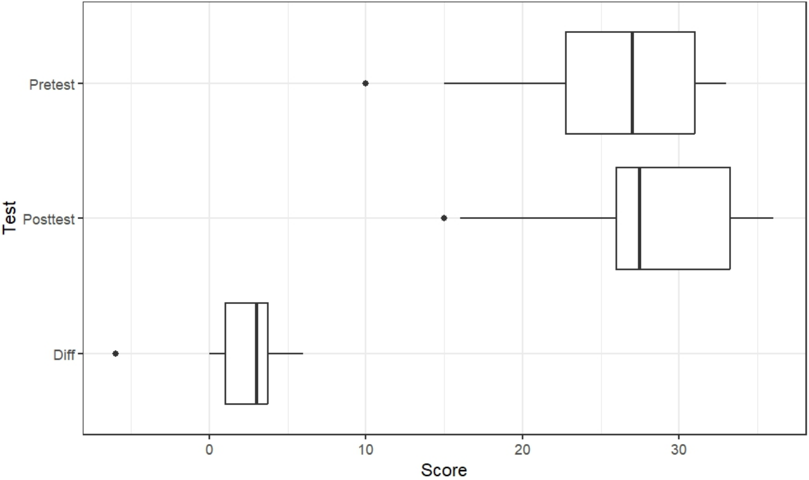
| ID | Pretest | Posttest | Difference | ID | Pretest | Posttest | Difference |
|---|---|---|---|---|---|---|---|
| 1 | 32 | 34 | 2 | 11 | 30 | 36 | 6 |
| 2 | 31 | 31 | 0 | 12 | 20 | 26 | 6 |
| 3 | 29 | 35 | 6 | 13 | 24 | 27 | 3 |
| 4 | 10 | 16 | 6 | 14 | 24 | 24 | 0 |
| 5 | 30 | 33 | 3 | 15 | 31 | 32 | 1 |
| 6 | 33 | 36 | 3 | 16 | 30 | 31 | 1 |
| 7 | 22 | 24 | 2 | 17 | 15 | 15 | 0 |
| 8 | 25 | 28 | 3 | 18 | 32 | 34 | 2 |
| 9 | 32 | 26 | -6 | 19 | 23 | 26 | 3 |
| 10 | 20 | 26 | 6 | 20 | 23 | 26 | 3 |

# Example – French Institute




- if I do a two-sample t-test for a diff. in means, will I find a difference?

# Example – French Institute




- If we do a paired t-test to see if differences are zero, what will we find?

# Example – French Institute

Results of the *paired t-test*

```
> t.test(post, pre, paired = TRUE)

  Paired t-test

data:  post and pre
t = 3.86, df = 19, p-value = 0.001
alternative hypothesis: true mean difference is
    not equal to 0
95 percent confidence interval:
 1.1461 3.8539
sample estimates:
mean difference
            2.5
```

# Example – French Institute

Results of the unpaired t-test, no power to find difference

```
> t.test(post, pre, paired = FALSE)

  Welch Two Sample t-test

data:  post and pre
t = 1.29, df = 37.9, p-value = 0.2
alternative hypothesis: true difference in
   means is not equal to 0
95 percent confidence interval:
 -1.424  6.424
sample estimates:
mean of x mean of y
     28.3      25.8
```

# Review

- There are different ways of calculating the df for diff. in means test
- Two-sample t-tests have a paired version
    1. Reduces variability
    2. Also reduces degrees of freedom
    3. see if you can use paired test by checking if there are multiple observations per subject
- We can use R to do most of these for us
    - examples in today's lab

# Testing Proportions in R

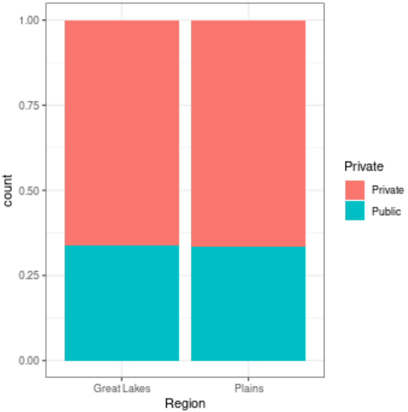
We can also use R to do the following:

- test if a proportion is equal to some value
  - $H_0 : p = p_0$
- test if two (or more!) proportions equal zero
  - $H_0 : p_1 = p_2$
  - $H_0 : p_1 - p_2 = 0$
  - $H_0 : p_1 = p_2 = ... = p_N$ (N is # of groups, not sample size)

# Difference in Proportions

| | Private | Public | Total |
|---|---|---|---|
| Great Lakes | 125 | 64 | 189 |
| Plains | 84 | 42 | 126 |

- $H_0 : p_1 - p_2 = 0$
- $\hat{p}_1 = 0.661, n_1 = 189$
- $\hat{p}_2 = 0.666, n_2 = 126$
- will a test find a difference?



# Review – Hypothesis Test for Difference of Proportions

$\mathrm{H}_0$: $p_1 - p_2 = 0$

Under $H_0$, $p_1 = p_2$, both are estimating the same thing.

$$\text{Let} \quad \widehat{p}_{pool} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

If we are simulating what the null hypothesis looks like, then

$$Z := \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\widehat{p}_{pool}(1-\widehat{p}_{pool})}{n_1} + \frac{\widehat{p}_{pool}(1-\widehat{p}_{pool})}{n_2}}} = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\widehat{p}_{pool}(1-\widehat{p}_{pool})(\frac{1}{n_1} + \frac{1}{n_2})}} \sim \mathbf{N(0,1)}$$

- use pnorm() with value of Z

| | Private | Public | Total |
|---|---|---|---|
| Great Lakes | 125 | 64 | 189 |
| Plains | 84 | 42 | 126 |

```
> prop.test(x = c(125, 84), n = c(189, 126))

  2-sample test for equality of
  proportions with continuity
  correction

data:  c(125, 84) out of c(189, 126)
X-squared < 3.74E-30
df = 1, p-value = 1
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.11701  0.10643
sample estimates:
 prop 1  prop 2
0.66138 0.66667
```

- The test-statistic R uses here is actually the square of the test-statistic we previously found ($Z^2$)
- the distribution is not standard Normal, it is something else that we will talk about on Monday. Regardless, the p-value from ‘prop.test()‘ is still the same