Two Group Differences

Grinnell College

November 18, 2024

Review

Suppose that we have a diagnostic test for an infectious disease which has a Type I error rate of 5% and a Type II error rate of 1%. Answer the following:

1. What is the null hypothesis for a diagnostic test? What constitutes a Type I and Type II error?
2. Suppose we use it to test for the disease on a population of 1,000 people where 40% have the disease. Construct a table showing the number of correct and incorrect conclusions based on the truth of H_0
3. Of individuals with a positive test, what percentage actually had the disease?
4. For testing for an infectious disease, is Type I or Type II error more important to control?

The process is the same

What we did before, we will do today:

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

$$t = \frac{\overline{X} - \mu_0}{\hat{\sigma}/\sqrt{n}}$$

4. Reject or fail to reject hypothesis
   - Type I errors
   - Type II errors

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, μ_A and μ_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

Just as in the univariate case for testing the mean, we can use a t-test to evaluate the difference in means between two groups

There are a number of various assumptions about our data, 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

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

Two-sampled t-test

Just as in the univariate case for testing the mean, we can use a t-test to evaluate the difference in means between two groups

There are a number of various assumptions about our data, all resulting in slightly different tests (degrees of freedom and standard error):

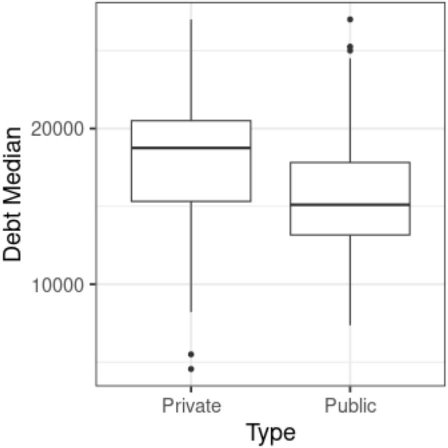
1. ~~Independent, groups same size and have same variance~~
2. ~~Independent, groups have unequal sizes and similar variance~~
3. Independent, groups have different sizes and different variances
4. Paired testing

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

Example

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
  - $\overline{x}_2 = 15627$
  - $\hat{\sigma}_2 = 3111$
  - $n_2 = 559$



t-test, Independent samples, heterogeneous groups

Our t-statistic takes the form

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}}$$

This t-statistic only approximately follows a t-distribution, making the calculation of its degrees of freedom non-trival, usually approximated using $n_1 + n_2 - 2$ (or with software)

Otherwise, the process for constructing confidence intervals or testing hypotheses is exactly the same

Example

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
```

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 it ostensibly shares many characteristics with 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 *difference*, respectively

Paired t-test

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

$$H_0 : \mu_0 = 0$$

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

Example – French Institute

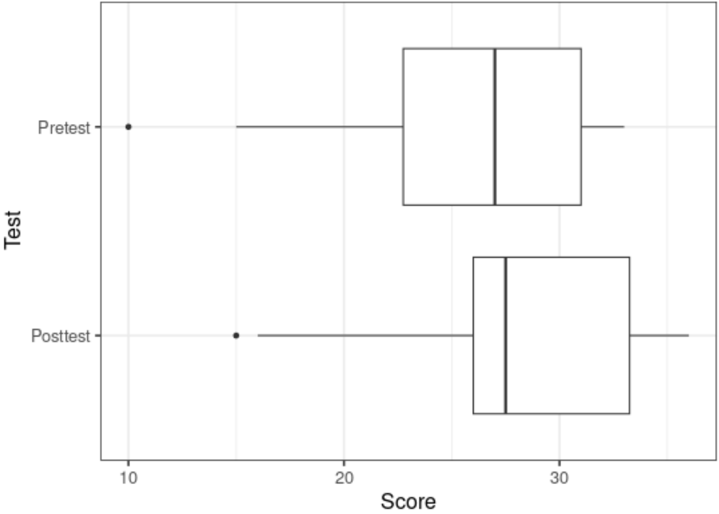
1. What is the null hypothesis for this study?
   - What would be a Type I error?
   - A Type II error?
2. How many total subjects do we have?
3. How many recorded observations do we have?

Example – French Institute

The results of the tests are as follows:

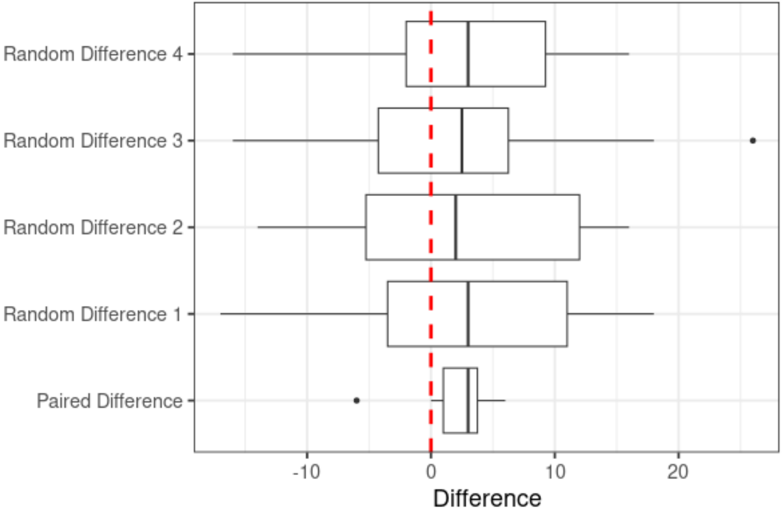
ID	Pretest	Posttest	Difference	ID	Pretest	Posttest	Difference
1	32	34	2	11	30	36	6
2	31	31	0	12	20	26	6
3	29	35	6	13	24	27	3
4	10	16	6	14	24	24	0
5	30	33	3	15	31	32	1
6	33	36	3	16	30	31	1
7	22	24	2	17	15	15	0
8	25	28	3	18	32	34	2
9	32	26	-6	19	23	26	3
10	20	26	6	20	23	26	3

Example – French Institute



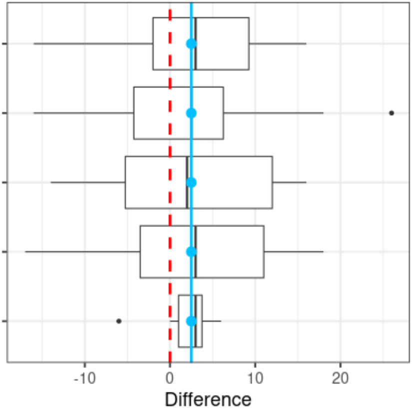
Example – French Institute

Plotted below is a boxplot of observed differences if people were randomly shuffled and repaired in each group



There are a few things to notice here:

- The mean value for each arrangement is *identical*
- The groups that were randomly assigned show far greater variability
- Less variability = more power



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
```

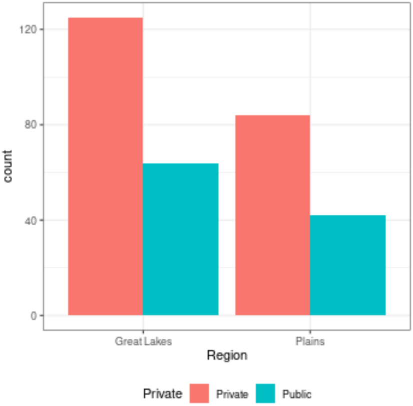
Review

- Hypothesis testing works nearly identically for two groups as it did with one group
- CLT applies for both difference in proportions as well as difference in group means
- Two-sample t-tests have a paired version
    1. Reduces variability
    2. Also reduces degrees of freedom
- We can use R to do most of these for us

Difference in Proportions

Suppose we are interested in determining if the composition of public and private schools is the same between the Plains region and the Great Lakes

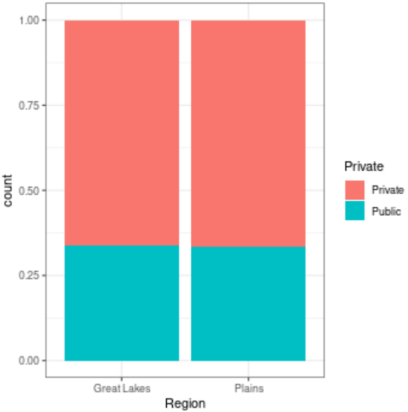
	Private	Public
Great Lakes	125	64
Plains	84	42



Difference in Proportions

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

- $H_0 : p_1 - p_2 = 0$
- $\hat{p}_1 = 0.661, n_1 = 189$
- $\hat{p}_2 = 0.666, n_2 = 126$



Differences in Proportion

The central limit theorem gives

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2,\ \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$$

The procedure for hypothesis testing is exactly the same:

1. State null hypothesis and construct distribution of values under the null
2. Create t-statistic using point estimate and standard error
3. Determine probability of observing t-statistic under the null, get p-value (`prop.test()` in R)
4. Reject or fail to reject H_0

	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
```