Chapter 6
Compare Two Groups

Section 6.2:
Standard Error of the Mean Difference

School of Business Administration
Portland State University

- Standard Error of the Mean Difference
  - Definition and Estimation
  - Application

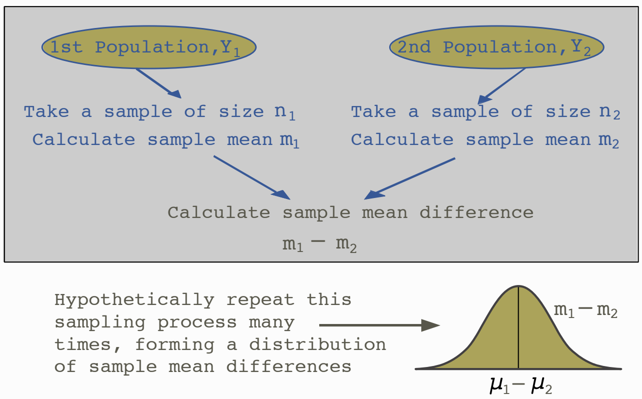
6.2a
Definition and Estimation

Statistical Inference Requires the Standard Error

Basic tools of inference

- Statistical inference, confidence interval or hypothesis test, is based on the concept of a standard error
- **Standard Error**: The (usually hypothetical) standard deviation of a statistic over multiple samples
- Instead of actually taking multiple samples, the standard error can be estimated from the information in a single sample, though it only has meaning defined over multiple samples
- In the case of the standard error of the mean difference ...
  - two independent samples are drawn
  - the corresponding means computed
  - the difference between the two sample means computed
- Conceptually, the meaning of standard error is the standard deviation of the entire distribution of the sample mean difference over many (usually hypothetical) repeated trials

Conceptual Meaning of Standard Error of Mean Difference



Standard error of the sample mean difference: Standard deviation of the distribution of sample mean differences

Standard Error: Homogeneity of Variance

A technical issue that requires consideration

- **Homogeneity of variance**: Equal population variances of Y for each group, that is, $\sigma_1^2 = \sigma_2^2$, which equal a common σ^2
- For technical, mathematical reasons, only when homogeneity of variance is true can the t-distribution be strictly applied to statistical inference of the mean difference
- In practice, homogeneity of variance is usually demonstrated, or at least assumed, for the analysis of the mean difference
- Fortunately, even if the assumption is not true, the t-distribution probabilities are still generally appropriate *unless* one of the sample sizes is particularly small and/or the two sample variances are dramatically discrepant
- If the population variances do not appear to be equal, as shown by some analyses described shortly, use a second version of the t-test which is free of this assumption

Standard Error: Within-group Variability

The average of the variability of Y within each group

- **Key Concept**: To estimate a single common population variance, σ^2, combine the two separate sample variances for each group, s_1^2 and s_2^2, into a single estimate of σ^2
- Combine the sample variances by calculating the average of the two sample variances
- To calculate this average, weight each variance by its degree of freedom, $df_1 = n_1 - 1$ and $df_2 = n_2 - 1$, to give the variance from the larger sample more weight
- Average the variance of Y within the first group with the variance within the second group, so denote as s_w^2
- Overall average variability within the groups:

$$s_w^2 = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2}$$

Standard Error of the Mean Difference: Formula

Generalize from the mean to the mean difference

- Estimated standard error of the mean: $s_m = \frac{s}{\sqrt{n}} = s\sqrt{\frac{1}{n}}$
- To generalize to two groups
  - Replace s, the standard deviation of Y for one group, with its corresponding averaged value for two groups, $s_w = \sqrt{s_w^2}$
  - Add a second term for the second sample size
- The straightforward generalization to the estimated standard error of the mean difference: $s_{m_1-m_2} = s_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
- For equal sample sizes, $n_1 = n_2$, or n, $s_{m_1-m_2} = s_w\sqrt{\frac{2}{n}}$

Evaluate Homogeneity of Variance

Formal hypothesis tests

- Evaluate the assumption of homogeneity of variance, such as with two different hypothesis tests described next
- The null hypothesis for each test is equal population variances

  $H_0 : \sigma_1^2 = \sigma_2^2$
- When evaluating the p-value from a specific hypothesis test of equal variances, the goal is usually a p-value larger than $\alpha = 0.05$, consistent with the assumption of equal variances
- Unfortunately, hypothesis tests of equal variances have low power at low samples sizes, exactly when the test becomes more important
- Instead of a literal interpretation of the corresponding p-values from these hypothesis tests, probably best to interpret the results as an informal "rule of thumb", or heuristic

Evaluate Homogeneity of Variance: Test #1

Variance Ratio Test

- Calculate the ratio of the two variances, called an F-statistic, often for convenience with the larger variance on top
- To calculate this ratio in terms of hypothetical repeated sampling, draw a sample from each population, so that
  - Each numerator variance is of a sample from 1^{st} population
  - Each corresponding denominator variance is of a sample from 2^{nd} population
- The distribution of the ratio of variances over repeated pairs of samples follows an F-distribution, one distribution for each combination of numerator *df* and denominator *df*
- Null hypothesis: Equal population variances, so their ratio is 1
- Is the obtained variance so much larger than 1 that the null hypothesis of variance equality becomes untenable?

Evaluate Homogeneity of Variance: Test #2

Levene's Test

- Levene proposed the following test for equal group variances given the null hypothesis of equal population group variances
  - Deviate each data value Y_i from its own group mean
  - Evaluate the mean difference of these deviation scores for the groups, such as with the t-test
- Brown and Forsythe modified this procedure: Calculate each deviation from the respective group *median*
- The result of this modification is a more robust hypothesis test, able to provide more accurate results across a wider range of different scenarios, and generally accepted as the preferred method

The Standard Error for Unequal Group Variances

What to do if the group variances are *not* equal?

- If the population variances of the two groups cannot be assumed equal, particularly at small sample sizes, use another version of the standard error to define a different t-test
- This version, $s'_{m_1-m_2}$, does *not* assume equality of the variances as it separately includes the variance of each group

$$s'_{m_1-m_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- A t-test based on this version of the standard error is called the Welch Two-Sample t-test
- The problem from using two separate variances instead of a single average is that the corresponding t-statistic, the ratio of the mean difference divided by this standard error, only *approximately* follows the mathematical t-distribution

Degrees of Freedom Depending on the Standard Error

Assume homogeneity of variance

- For the usual case of equal variances, the degrees of freedom for the analysis is just the sum of the degrees of freedom for each group

$$df = df_1 + df_2 = (n_1 - 1) + (n_2 - 1)$$

Do not assume homogeneity of variance

- For unequal variances, a rather extensive expression for this *df* approximates the proper use of *t*-distribution probabilities

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

- Generally this *df* has decimal digits

6.2b Application

Homogeneity of Variance

- Evaluate Homogeneity of Variance with the lessR ttest function, here for the Supplier data: `> ttest(Time ~ Supplier)`

```
--- Assumptions ---

These hypothesis tests can perform poorly, and the
t-test is typically robust to violations of assumptions.
Use as heuristic guides instead of interpreting literally.

Null hypothesis is equal variances of Time, i.e., homogeneous.
Variance Ratio test:  F = 2.52/2.09 = 1.20,  df = 14;18,
                      p-value = 0.700
Levene's test, Brown-Forsythe:  t = -0.621,  df = 32,
                      p-value = 0.539
```

- Both *p*-values > 0.05, so evidence that the populations from which the data values are sampled have at least approximately equal variances

Within-Group Standard Deviation: Traditional Notation

Descriptive statistics

- **Data Summary**

  $n_1 = 19,\ m_1 = 9.574$ days, $s_1 = 1.447$ days
  $n_2 = 15,\ m_2 = 8.087$ days, $s_2 = 1.587$ days

- **Within-group variance**

  $$s_w^2 = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2} = \frac{(18)1.447^2 + (14)1.587^2}{18 + 14} = 2.28$$

- **Within-group standard deviation**

  $$s_w = \sqrt{s_w^2} = \sqrt{2.28} = 1.51 \text{ days}$$

Standard Error for Classic *t*-test

The standard error of the sample mean difference

- Get the needed standard error and *df* for the hypothesis test

Description	Name	Value	Formula
total deg of freedom	df	32	(nn1-1)+(nn2-1)
estimated std error	sterr	0.522	sw*SQRT((1/_nn1)+(1/_nn2))

- **df** $df = df_1 + df_2 = (19 - 1) + (15 - 1) = 32$

- **Standard Error** $s_{m_1 - m_2} = s_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

  $$= 1.51\sqrt{\frac{1}{19} + \frac{1}{15}} = 0.5215$$

Standard Error for Welch *t*-test

The standard error of the sample mean difference

- **n, s** $n_1 = 19, n_2 = 15, s_1 = 1.447, s_2 = 1.587$
- **df**

  $$df' = \frac{\left[(s_1^2/n_1) + (s_2^2/n_2)\right]^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)} = 28.762$$

- **Standard Error** $s'_{m_1 - m_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

  $$= \sqrt{\frac{1.447^2}{19} + \frac{1.587^2}{15}} = 0.5274$$

R Analysis of the Mean Difference, Welch *t*-test

R input and output

lessR t-test function provides for both the *t*-test with (Section 6.1) and without the assumption of equal group variances

```
> tt(Time ~ Supplier)
```

```
-- Do not assume equal population variances of Time
     for each Supplier

t-cutoff: tcut =  2.046
Standard Error of Mean Difference: SE =  0.53

Hypothesis Test of 0 Mean Diff:
     t = 2.82,  df = 28.76,  p-value = 0.009

Margin of Error for 95% Confidence Level:  1.08
95% Confidence Interval for Mean Difference:
     0.41 to 2.57
```


Equal vs Unequal Variances Assumption

Compare results of classic and Welch *t*-tests

- Assuming equal population variances yields the classic *t*-test
- Allowing unequal population variances yields the Welch *t*-test

Estimate	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$
df	32	28.762
lower bound	0.4247	0.4080
upper bound	2.5493	2.5661
t-value	2.8513	2.8195
p-value	0.0076	0.0086

- As noted, homogeneity of variance is supported in these data, and here there is no practical distinction between between classic and Welch analyses
- If population variances demonstrated not equal, then results of the two tests may diverge, particularly in small samples

Index

Subtract 2 from each listed value to get the Slide #

- The End