BIOS:4120 – Introduction to Biostatistics
Unit 10: Comparison of Two Means

Knute D. Carter

Department of Biostatistics
The University of Iowa

October 16, 2025

Overview

- Paired Samples
- Independent Samples
  - Variances Known
  - Common Unknown Variances
  - Unknown Differing Variances

Introduction

- In the last chapter, we learned how to test whether the mean of the population from which we drew our sample was different from some pre-specified value, μ_0.
- In many inferential settings, what we really want to do is compare two means from two populations, say μ_1 and μ_2.
- A natural comparison would be between two sample groups, such as a study (intervention) group and a control group, i.e. two independent samples.

Introduction

A comparison of μ_1 and μ_2 can be made in two ways:

1. We can conduct a hypothesis test where the null hypothesis H_0 corresponds to $\mu_1 = \mu_2$ and the alternative hypothesis H_A corresponds to $\mu_1 \neq \mu_2$, $\mu_1 < \mu_2$, or $\mu_1 > \mu_2$.
2. We can construct a confidence interval for the difference $\mu_1 - \mu_2$.

Paired Samples

- A *Paired Sample* arises when two samples of measurements are collected in such a manner that every observation in the first sample can be matched to a corresponding observation in the second sample.

- Examples of paired data:
  - Measurements taken on a patient before and after an intervention or treatment.
  - Observations collected on each of two twins.
  - Observations collected on a husband and a wife.
  - Observations collected on each of two animals from the same litter.
  - Measurements taken on each of a subject's two eyes (two arms, two legs, etc.).

Paired Samples

- In the paired data situation, it is natural to compute the difference (D) in the response between the two members of the pair (Y_1 and Y_2).
- Assess whether the two differ by testing to see if the mean difference is zero.

Paired Samples

- Consider a paired sample represented as follows:

Measurement 1	Measurement 2
y_{11}	y_{12}
y_{21}	y_{22}
y_{31}	y_{32}
$\vdots$	$\vdots$
y_{n1}	y_{n2}

- So y_{ij} is measurement j for person i.

Paired Samples

- Let μ_1 denote the mean for the population associated with the first sample, and let μ_2 denote the mean for the population associated with the second sample.
- Let $\mu_d = \mu_1 - \mu_2$ denote the difference between μ_1 and μ_2. (This difference is denoted by δ in the text.)

Inferential procedures on μ_d are conducted as follows:-

1. Reduce the two samples of observations to a single sample by taking differences:

Paired Samples

Measurement 1	Measurement 2	Difference
y_{11}	y_{12}	$d_1 = y_{11} - y_{12}$
y_{21}	y_{22}	$d_2 = y_{21} - y_{22}$
y_{31}	y_{32}	$d_3 = y_{31} - y_{32}$
$\vdots$	$\vdots$	$\vdots$
y_{n1}	y_{n2}	$d_n = y_{n1} - y_{n2}$

2. Compute the mean $\bar{d}$ and the standard deviation s_d of these differences.

3. Apply the inferential procedures in Units 8 and 9 to the parameter μ_d, which represents the mean for the population of differences.

Paired Samples

- If a confidence interval is constructed, the interval estimates $\mu_d = \mu_1 - \mu_2$.
- If a hypothesis test is conducted, the hypotheses are formulated as follows:

- For a <u>two-sided test</u>, we have

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

$$H_A : \mu_d \neq 0 \iff \mu_1 - \mu_2 \neq 0 \iff \mu_1 \neq \mu_2$$

Paired Samples

- For a one-sided test, we have

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

$$H_A: \mu_d > 0 \iff \mu_1 - \mu_2 > 0 \iff \mu_1 > \mu_2$$

or

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

$$H_A: \mu_d < 0 \iff \mu_1 - \mu_2 < 0 \iff \mu_1 < \mu_2$$

- Note that the baseline value of μ_d under H_0 is $\mu_0 = 0$.

Paired Samples

- **Note**: The number of differences n is the relevant sample size, and thereby dictates whether the z procedures or the t procedures should be used for conducting hypotheses tests and constructing confidence intervals.
- If $n \geq 30$, we may use the z procedures.
  - Normality (of the population of differences) is not required.
  - The population standard deviation of differences σ_d can be replaced by the sample standard deviation of differences s_d in the formulas for the confidence intervals and the test statistic.
- If $n < 30$ we should use the t procedures.
  - Normality (of the population of differences) is required.

Paired t-test

- A $100(1-\alpha)\%$ confidence interval for μ_d would be:

$$\bar{d} \pm t_{n-1,\alpha/2} \frac{s_d}{\sqrt{n}}$$

where s_d is the sample standard deviation of the n differences.

- Tests of $H_0 : \mu_d = 0$ are based on the test statistic:

$$\frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{\bar{d}\sqrt{n}}{s_d}$$

which has a t_{n-1} distribution when H_0 is true.

Paired t-test

- The test on the preceding slide is known as the *paired t-test*, and is an application of the general one-sample t-test with unknown variance.
- Hence the same assumptions hold, only now on the differences d_i.

Example

- Measurements of cell fluidity were made in 20 dishes of pulmonary artery cells from 10 dogs.
- Cell samples from each dog were randomly assigned to an oxygen (O_2) treatment or to a control (no O_2) treatment.
- The following table below gives the results, where
  Y_1 = cell fluidity in the no O_2 group, and
  Y_2 = cell fluidity in the O_2 group.

Example

Dog	Y_1 (No O_2)	Y_2 (O_2)	$D = Y_1 - Y_2$
Jake	0.308	0.308	0.000
Nathan	0.304	0.309	−0.005
Monty	0.305	0.305	0.000
Kai	0.304	0.311	−0.007
Bella	0.301	0.303	−0.002
Rocky	0.278	0.293	−0.015
Casala	0.296	0.302	−0.006
Louie	0.301	0.300	0.001
Tiger	0.302	0.308	−0.006
BB	0.237	0.250	−0.013
Mean	0.2936	0.2989	−0.0053
Std Dev	0.0216	0.0180	0.0054

Example

- $\bar{d} = -0.0053$ is the sample mean of the differences.
- $s_d = 0.0054$ is the standard deviation of the differences.
- $s_d/\sqrt{10} = 0.0017$ is the standard error of the mean of the differences.
- Note that even though $\bar{d} = \bar{y}_1 - \bar{y}_2$, we cannot determine s_d from s_1 and s_2, because it depends on the correlation between the paired measurements

Example

- Test the hypothesis that $\mu_d = 0$ versus the alternative that it is not at $\alpha = 0.05$ significance level.

Example

Independent Samples

- A very common research goal is to compare two (or more) independent groups to see if they have the same mean.
- We will focus on comparing two groups.
- *Independent Samples* arise when two sets of observations are collected based on two independently drawn random samples from two populations. The two samples may or may not have the same size.

Independent Samples

We can think of this two ways:

1) there is one population of subjects, but we have subdivided them into two (or more) groups, producing sub-populations.

2) there are two (or more) separate populations which are comparable except in their responses to their respective treatments.

- If the treatments are assigned randomly, the two situations above are equivalent.
- Non-random separation: e.g., separate males and females as two groups.

Independent Samples

- Consider a situation where we have two random samples of size n_1 and size n_2 from two (independent) groups.

$$Y_{11}, Y_{21}, \ldots, Y_{n_1,1} : \text{ sample mean, } \bar{Y}_1; \text{ and sample SD, } s_1$$

$$Y_{12}, Y_{22}, \ldots, Y_{n_2,2} : \text{ sample mean, } \bar{Y}_2; \text{ and sample SD, } s_2$$

- Further assume that both samples are from normal distributions:

$$Y_{i1} \sim N(\mu_1, \sigma_1^2) \text{ and } Y_{i2} \sim N(\mu_2, \sigma_2^2).$$

Independent Samples

- Often we are interested in testing one of the following sets of hypotheses:

$$H_0 : \mu_1 = \mu_2 \text{ vs. } H_A : \mu_1 \neq \mu_2$$

$$H_0 : \mu_1 \leq \mu_2 \text{ vs. } H_A : \mu_1 > \mu_2$$

$$H_0 : \mu_1 \geq \mu_2 \text{ vs. } H_A : \mu_1 < \mu_2$$

- Similarly, we might be interested in confidence intervals (one- or two-sided) on the difference $\mu_1 - \mu_2$.

- How do we do this?
  We can estimate $\mu_1 - \mu_2$ with $\bar{Y}_1 - \bar{Y}_2$.
  But what is the standard error of the estimate $\bar{Y}_1 - \bar{Y}_2$?

Independent Samples: Case I

Case I: σ_1^2 and σ_2^2 are known, (unrealistic).

- It can be shown that

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

  which means you can standardize to a $N(0,1)$ distribution as we did before.

- A $100(1-\alpha)\%$ confidence interval on $\mu_1 - \mu_2$ is

$$(\bar{Y}_1 - \bar{Y}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Independent Samples: Case I

- Tests of $H_0 : \mu_1 - \mu_2 = \mu_0$ are based on the test statistic

$$\frac{(\bar{Y}_1 - \bar{Y}_2) - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

which has a $N(0, 1)$ distribution if H_0 is true.

Independent Samples: Case II

Case II: $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (Equal, but unknown. More realistic.)

- It is still true that

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

which can now be written as

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1 - \mu_2,\ \sigma^2 \left\{\frac{1}{n_1} + \frac{1}{n_2}\right\}\right)$$

Independent Samples: Case II

- It can be shown that when we replace σ^2 by an estimate (s_p^2), then

$$\frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

where

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

is a pooled estimate of the common variance using a weighted average of the two individual sample variances (s_1^2 and s_2^2).

Independent Samples: Case II

- A $100(1-\alpha)\%$ confidence interval on $\mu_1 - \mu_2$ is

$$(\bar{Y}_1 - \bar{Y}_2) \pm t_{n_1+n_2-2,\,\alpha/2} \times s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\,.$$

- Tests of $H_0 : \mu_1 - \mu_2 = \mu_0$ are based on the test statistic

$$t = \frac{(\bar{Y}_1 - \bar{Y}_2) - \mu_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

which has a $t(n_1 + n_2 - 2)$ distribution if H_0 is true.

Independent Samples: Case II, Example

- In an investigation of pregnancy-induced hypertension, one group of women with this disorder was treated with low-dose aspirin, and a second group was given a placebo.
- A sample consisting of 23 women who received aspirin has a mean arterial blood pressure of 111 mm Hg with a standard deviation of 8 mm Hg.
- A sample of 24 women who were given the placebo has a mean blood pressure of 109 mm Hg with a standard deviation of 7 mm Hg.

Independent Samples: Case II, Example

(a) At the 0.01 level of significance, test the null hypothesis that the two populations of women have the same mean arterial blood pressure.

Independent Samples: Case II, Example

Independent Samples: Case II, Example

(b) Construct a 99% confidence interval for the true difference in population means. Does this interval contain the value 0?

Notes on Pooling Variance Estimates

- We have two samples:

$$n_1, \;\; \bar{y}_1, \;\; s_1^2, \;\; \sqrt{s_1^2} = s_1, \;\; E(Y_{i1}) = \mu_1, \;\; Var(Y_{i1}) = \sigma_1^2$$

$$n_2, \;\; \bar{y}_2, \;\; s_2^2, \;\; \sqrt{s_2^2} = s_2, \;\; E(Y_{i2}) = \mu_2, \;\; Var(Y_{i2}) = \sigma_2^2$$

- If we assume the underlying populations have the same variance, then $\sigma_1^2 = \sigma_2^2$. Call this common variance σ^2.

- Under this common variance assumption, it makes sense to *pool* our estimates of that common value.
  That is, s_1^2 and s_2^2 both estimate the common σ^2.

Notes on Pooling Variance Estimates

- We do this by creating a weighted average, where the weights depend on the sample sizes in the two samples (n_1 and n_2).

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

- More weight is given to the sample estimate from the sample with the larger sample size.

- If $n_1 = n_2$, then the pooled estimate is a simple average of the two.

Notes on Pooling Variance Estimates

- Computationally ...

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (y_{i1} - \bar{y}_1)^2$$

$$s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (y_{i2} - \bar{y}_2)^2$$

and so

$$s_p^2 = \frac{\sum_{i=1}^{n_1} (y_{i1} - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (y_{i2} - \bar{y}_2)^2}{n_1 + n_2 - 2}$$

Notes on Pooling Variance Estimates

NOTE!!!

- The pooling must be done on the variances. It does not work to pool the standard deviations in the same way. That is,

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

$$\begin{aligned} s_p = \sqrt{s_p^2} &= \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \\ &\neq \frac{(n_1 - 1)\sqrt{s_1^2} + (n_2 - 1)\sqrt{s_2^2}}{n_1 + n_2 - 2} \end{aligned}$$

Recap: Underlying Assumptions

1. Two samples are independent and observations within each sample are independent (random samples).
2. $\bar{Y}_1 \sim N(\mu_1, \sigma_1^2/n_1)$.
3. $\bar{Y}_2 \sim N(\mu_2, \sigma_2^2/n_2)$.
4. $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (common variance in the two groups).
5. Items 2 and 3 above imply that the difference $\bar{Y}_1 - \bar{Y}_2$ is also normally distributed with mean $E(\bar{Y}_1 - \bar{Y}_2) = \mu_1 - \mu_2$ and variance $\sigma_1^2/n_1 + \sigma_2^2/n_2$; if item 4 is true, the variance is $\sigma^2(1/n_1 + 1/n_2)$.

Recap: Underlying Assumptions

6. $$\frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

7. Even if items 2 and 3 are not true, the *t*-test is often robust enough to be used as an approximation, depending on the sample size (related to the Central Limit Theorem).

Independent Samples: Case III

Case III: $\sigma_1^2 \neq \sigma_2^2$, both unknown.

- It is still true that

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1 - \mu_2,\ \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

- However, it does NOT turn out that

$$t = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \text{ also has a } t \text{ distribution.}$$

Independent Samples: Case III

- The best we can do is *approximate* the true distribution with a t-distribution with the following degrees of freedom:

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

- That is,

$$\nu = \frac{(a+b)^2}{\frac{a^2}{n_1-1} + \frac{b^2}{n_2-1}}$$

where

$$a = \frac{s_1^2}{n_1} \text{ and } b = \frac{s_2^2}{n_2}.$$

Independent Samples: Case III

- $100(1-\alpha)\%$ (approximate) confidence interval on $\mu_1 - \mu_2$ is

$$(\bar{Y}_1 - \bar{Y}_2) \pm t_{\nu,\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where the degrees of freedom are estimated by ν.

Independent Samples: Case III

How to decide if $\sigma_1^2 = \sigma_2^2$ is reasonable?

1) 'Rule of thumb': if $s_1^2/s_2^2 > 3$ or $< 1/3$, assume inequality.

2) Do a special test of $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_A : \sigma_1^2 \neq \sigma_2^2$ using what is known as an F-test.

Independent Samples: Case III, Example

- Suppose that you wish to compare the characteristics of tuberculosis meningitis in patients infected with HIV and those who are not infected.
- In particular, you would like to determine whether the two populations have the same mean age.
- A sample of 37 infected patients has mean age $\bar{y}_1 = 27.9$ years and standard deviation $s_1 = 5.6$ years.
- A sample of 19 patients who are not infected has mean age $\bar{y}_2 = 38.8$ years and standard deviation $s_2 = 21.7$ years.

Independent Samples: Case III, Example

(a) Test the null hypothesis that the two populations of patients have the same mean age at the 0.05 significance level.

Independent Samples: Case III, Example

Test statistic:

$$
\begin{aligned}
t &= \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \\
&= \frac{(27.9 - 38.8) - (0)}{\sqrt{\frac{5.6^2}{37} + \frac{21.7^2}{19}}} \\
&= \frac{-10.9}{\sqrt{0.8476 + 24.78}} \\
&= \frac{-10.9}{5.062} \\
&= -2.15
\end{aligned}
$$

Independent Samples: Case III, Example

- Compare the -2.15 to a t-distribution with the following degrees of freedom:

$$
\begin{aligned}
\nu &= \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} \\
&= \frac{(0.85 + 24.78)^2}{\frac{0.85^2}{36} + \frac{24.78^2}{18}} \\
&= \frac{656.90}{0.02 + 34.11} \\
&= 19.24
\end{aligned}
$$

Independent Samples: Case III, Example

- Use d.f. $= 19$.
- Compare to the lower 0.025 percentile of a $t(19)$ distribution, which is -2.093.
- Since -2.15 is more extreme than -2.093 $(0.01 < p/2 < 0.025 \Rightarrow 0.02 < p < .05)$, we reject the null hypothesis at the 0.05 significance level.
- Conclusion: We conclude that we have sufficient evidence to claim that the mean age of those not infected with HIV is significantly higher than the mean age of those that are HIV positive.

Independent Samples: Case III, Example

(b) Do you expect a 95% confidence interval for the true difference in population means would contain the value 0? Why or why not?

F-Test for the Comparison of Two (Independent) Variances

- To test the hypothesis:

$$H_0 : \sigma_1^2 = \sigma_2^2 \;\; vs. \;\; H_A : \sigma_1^2 \neq \sigma_2^2$$

- Use the test statistic

$$F = \frac{s_1^2}{s_2^2}, \text{ or } F = \frac{s_2^2}{s_1^2}$$

- Under H_0, this F statistic follows an F-distribution with df_N and df_D degrees of freedom.

- If $F = s_1^2/s_2^2$ then
  $df_N =$ numerator degrees of freedom $= n_1 - 1$.
  $df_D =$ denominator degrees of freedom $= n_2 - 1$.

F-Test for the Comparison of Two (Independent) Variances

- For the two-sided test, compare that test statistic to the upper $\alpha/2$ cut-point of an F-distribution.
- To cut down on the size of the published F tables, the typical procedure is to define our F with the larger of the two sample variances in the numerator (still use the $\alpha/2$ cut-point).
- This is only one use of the F table. There are other types of F tests; some of these will use only the upper α cut-point.