Chapter 19: Comparing Two Populations Means using Independent Samples

Given two (quantitative) populations, we often want to compare the population means by considering their difference. In this chapter, we will make inferences about the value of the parameter $\mu_1 - \mu_2$ based upon samples which are **independent** of each other.

↳ want $\bar{y}_1$ and $\bar{y}_2$ to be independent random variables: $Var(\bar{y}_1 - \bar{y}_2) = Var(\bar{y}_1) + Var(\bar{y}_2)$

	Population Mean	Population Standard Deviation	Sample Size	Sample Mean	Sample Standard Deviation
Population 1	μ_1	σ_1	n_1	$\bar{y}_1$	s_1
Population 2	μ_2	σ_2	n_2	$\bar{y}_2$	s_2

use t-model

Typically, both σ_1 and σ_2 are unknown (and will have to be estimated), however, in some situations, we may believe that they are equal. In this case, we can use **pooling** to estimate the common variance (and common standard deviation).

$\sigma_1^2 = \sigma_2^2$ estimate with pooled variance s_p^2

We will consider two t-methods for making inferences about $\mu_1 - \mu_2$:

two sample
- **Nonpooled t-method:** σ_1 and σ_2 are not assumed to be equal.
  ↳ often $\sigma_1 \neq \sigma_2$
- **Pooled t-method:** $\sigma_1 = \sigma_2$.

Criterion: When should we assume equal variances/standard deviations? (**Equal Variance Assumption**)

$\sigma_1^2 = \sigma_2^2$

If the larger of the two sample standard deviations is less than twice the smaller of the two sample standard deviations, that is,

for this course use →

$$\frac{\text{larger } s}{\text{smaller } s} < 2$$

we will assume equal variances and use the pooled t-method for building confidence intervals and conducting hypothesis tests.

We may also want to consider:

- **Similar Spreads Condition:** making side-by-side boxplots of the data for each sample and comparing the IQRs. (Check that spreads are not wildly different)
- sample sizes. (often want sample sizes to be roughly the same.)
- if it is reasonable to assume the variances are equal.

  For example, suppose we want to conduct a randomized experiment to test the effectiveness of a home remedy by comparing it to a placebo. If we believe that both do nothing, then it seems reasonable to assume both treatment groups have the same mean and standard deviation post treatment. (Variance)

Note: In the case that we assume equal variances (and so equal standard deviations), we can use either the pooled t-method or the nonpooled t-method, but the pooled t-method is **stronger** under this condition. However, if we mistakingly use a pooled t-method when the variances are not equal, this might result in serious errors.

Assumptions and Conditions:

a) **Independence Assumption:**

   i) **Independent Responses Assumption:** within each group, the data sampled should come from independently responding individuals.

   ii) **Independent Groups Assumption:** the two samples, one from each population, must be independent of each other.

   iii) **Randomization Condition:** for each sample, the data should be drawn independently, using random selection from the population (or come from a completely randomized comparative experiment).

   iv) **10% Condition:** if the data are sampled without replacement, the sample size should not exceed 10% of the population.

b) **Normal Population Assumption:** we require either

   i) Both samples are large: $n_1 \geq 30$ and $n_2 \geq 30$, or

   ii) **Nearly Normal Condition:** Both of the population distributions are approximately Normal.

if one or both < 30, need ii)

The Sampling Distribution of the Difference Between Two Sample Means

When the relevant assumptions are made and conditions are met, the sampling distribution of $\bar{y}_1 - \bar{y}_2$ from two independent samples has:

$\bar{y}_1, \bar{y}_2$ independent random Variables

	Nonpooled t-Method $\sigma_1 \neq \sigma_2$	Pooled t-Method $\sigma_1 = \sigma_2 = \sigma_p$
Mean $\mu_{\bar{y}_1 - \bar{y}_2}$	$\mu_1 - \mu_2$	$\mu_1 - \mu_2$
Standard deviation $\sigma_{\bar{y}_1 - \bar{y}_2}$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\sigma_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
Standard error $SE(\bar{y}_1 - \bar{y}_2)$	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	pooled estimate $s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ where $s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$

If both of the population distributions are (approximately) Normal or both sample sizes are large ($n_1 \geq 30$ and $n_2 \geq 30$), then the sampling distribution of $\bar{y}_1 - \bar{y}_2$ is (approximately) normal.

Note: For the nonpooled t-method, we will use $df = \text{Min}(n_1 - 1, n_2 - 1)$, but software will use a t-distribution with

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

Two-Sample t-Interval for the Difference between Means

Recall: a confidence interval has the form

$$\text{point estimate} \pm \text{margin of error}$$
$$= \text{point estimate} \pm \left(\text{critical value} \times \text{standard error of the estimate}\right)$$

$\bar{y}_1 - \bar{y}_2$ $\quad t^*_{df}$ $\quad SE(\bar{y}_1 - \bar{y}_2)$

When the relevant assumptions are made and conditions are met:

a) if we assume $\sigma_1 \neq \sigma_2$, then we use the ***t*-method without pooling**, in which case a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is:

$$(\bar{y}_1 - \bar{y}_2) \pm t^*_{df}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where t^*_{df} is the critical value corresponding to the $100(1-\alpha)\%$ confidence level based on $df = \text{Min}(n_1 - 1, n_2 - 1)$ (or the complicated formula above).

$0 \in CI$? $\quad CI < 0$? $\quad CI > 0$?

b) if we assume $\sigma_1 = \sigma_2$, then we use the ***t*-method with pooling**, in which case a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is:

$$(\bar{y}_1 - \bar{y}_2) \pm t^*_{df}\, s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where s_p is the pooled standard deviation

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

and t^*_{df} is the critical value corresponding to the $100(1-\alpha)\%$ confidence level based on $df = n_1 + n_2 - 2$.

Hypothesis Test for the Difference between Means: Two-Sample t-Test

Two-Sample Nonpooled t-Test:

A hypothesis test for $\mu_1 - \mu_2$ using the t-method without pooling has five steps:

1. **Assumptions/Conditions:**
   - σ_1 and σ_2 are both unknown and $\sigma_1 \neq \sigma_2$. (not assumed to be equal)
   - Data collected using randomization.
   - The two groups are independent of each other.
   - Within each group, we have independent responses from the individuals.
   - Either **both** $n_1 \geq 30$ and $n_2 \geq 30$, or **both** population distributions are approximately Normal.

2. **Hypotheses:**

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$

often $\Delta_0 = 0$

choose one

$$H_A : \left\{ \begin{array}{ll} \mu_1 - \mu_2 \neq \Delta_0 & \text{(two-tailed test)} \\ \mu_1 - \mu_2 < \Delta_0 & \text{(lower-tailed test)} \\ \mu_1 - \mu_2 > \Delta_0 & \text{(upper-tailed test)} \end{array} \right.$$

3. **Test Statistic:**

if $\Delta_0 = 0 \Rightarrow \bar{y}_1 - \bar{y}_2$

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2 - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

assuming H_0 true

where t_0 follows a t-distribution with $df = \text{Min}(n_1 - 1, n_2 - 1)$ (or the complicated formula above).

4. ***P*-value:** Compute one of the following using software or the t-table, where $df = \text{Min}(n_1 - 1, n_2 - 1)$ (or the complicated formula above).

Test	P-value
Two-tailed Test	$2P(t_{df} > \|t_0\|)$
Lower-tailed Test	$P(t_{df} < t_0)$
Upper-tailed Test	$P(t_{df} > t_0)$

If using a t-table, we will likely only be able to find a range in which the P-value lies.

5. **Conclusion:** Given a significance level α,
   - if P - value $\leq \alpha$, we reject H_0 at level α
   - if P- value $> \alpha$, we do not reject H_0 at level α

Two-Sample Pooled *t*-Test:

A hypothesis test for $\mu_1 - \mu_2$ using the t-method with pooling has five steps:

1. **Assumptions/Conditions:**
   - σ_1 and σ_2 are both unknown, but we assume $\sigma_1 = \sigma_2$.
   - Data collected using randomization.
   - The two groups are independent of each other.
   - Within each group, we have independent responses from the individuals.
   - Either **both** $n_1 \geq 30$ and $n_2 \geq 30$, or **both** population distributions are approximately Normal.

2. **Hypotheses:**

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$

often $\Delta_0 = 0$

$$\begin{aligned} &\mu_1 - \mu_2 \neq \Delta_0 && \text{(two-tailed test)} \\ H_A : &\mu_1 - \mu_2 < \Delta_0 && \text{(lower-tailed test)} \\ &\mu_1 - \mu_2 > \Delta_0 && \text{(upper-tailed test)} \end{aligned}$$

if $\Delta_0 = 0$
$\bar{y}_1 - \bar{y}_2$

3. **Test Statistic:**

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2 - \Delta_0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

where

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

assuming H_0 true

and where t_0 follows a t-distribution with $df = n_1 + n_2 - 2$.

4. ***P*-value:** Compute one of the following using software or the t-table, where $df = n_1 + n_2 - 2$.

Test	P-value
Two-tailed Test	$2P(t_{df} > \lvert t_0 \rvert)$
Lower-tailed Test	$P(t_{df} < t_0)$
Upper-tailed Test	$P(t_{df} > t_0)$

If using a t-table, we will likely only be able to find a range in which the P-value lies.

5. **Conclusion:** Given a significance level α,

- if P - value $\leq \alpha$, we reject H_0 at level α
- if P- value $> \alpha$, we do not reject H_0 at level α

Example: Surgeons wanted to determine if a new surgical method reduced the operative time relative to the standard method for a common surgical procedure. To conduct a randomized comparative experiment for the two surgical methods, they randomly selected 20 patients requiring this surgery and then randomly assigned 14 patients to the new method and 6 patients to the standard method. The operative times were recorded (in minutes). The summary statistics are given below:

	New Method (pop 1)	Standard Method (pop 2)
Sample size	14	6
Sample Mean	394.64	468.33
Sample Standard Deviation	84.75	38.17

$$\frac{\text{larger } s}{\text{smaller } s} = \frac{84.75}{38.17} = 2.22 \geq 2$$

$\Rightarrow$ assume $\sigma_1 \neq \sigma_2$ and use nonpooled t-test

a) At the 5% significance level, do the data provide sufficient evidence to conclude that the mean operative time is less with the new method than with the standard method?

1. **Assumptions/Conditions:**

- assume $\sigma_1 \neq \sigma_2$
- random samples (assume independence)
- assume independent groups
- Sample sizes are small, so assume both population distributions are approximately Normal.

2. **Hypotheses:**

$$H_0: \mu_1 - \mu_2 = 0$$
$$H_A: \mu_1 - \mu_2 < 0$$

lower-tailed test

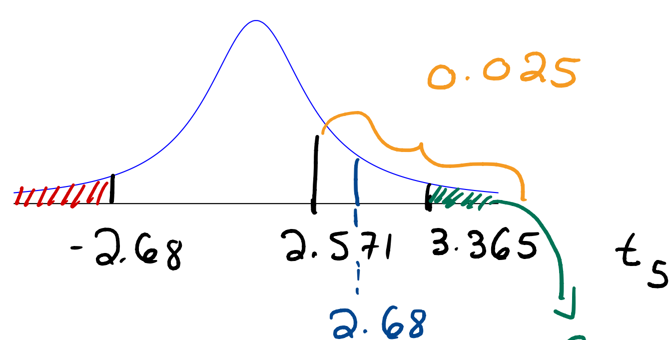
3. **Test Statistic:**

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{394.64 - 468.33}{\sqrt{\frac{(84.75)^2}{14} + \frac{(38.17)^2}{6}}}$$

$$= -2.68$$

4. ***P*-value:**

$$df = \min(n_1 - 1, n_2 - 1)$$
$$= \min(13, 5)$$
$$= 5$$



Using t-table with df = 5,

$$0.01 < \text{P-value} = P(t_5 < -2.68) < 0.025$$
$$\leq 0.05$$

Two sample T hypothesis test:

μ_1 : Mean of New Method
μ_2 : Mean of Standard Method
$\mu_1 - \mu_2$: Difference between two means
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 < 0$
(without pooled variances)

messy formula

Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	-73.690476	27.492134	17.83231	-2.6804204	0.0077

≤ 0.05

$\bar{y}_1 - \bar{y}_2$ $SE(\bar{y}_1 - \bar{y}_2)$

5. **Conclusion:**

Since P-value $\leq \alpha = 0.05$, we reject Ho at the 0.05 significance level, that is, there is sufficient statistical evidence to conclude that the mean operative time is less with the new method than with the standard method.

b) Construct a 90% confidence interval for the difference in mean operative time $\mu_1 - \mu_2$.

$$(\bar{y}_1 - \bar{y}_2) \pm t_5^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$= (394.64 - 468.33) \pm (2.015) \sqrt{\frac{(84.75)^2}{14} + \frac{(38.17)^2}{6}}$$

$$= -73.69 \pm 55.398$$

$$= (-129.09, -18.29)$$

Since $0 \notin (-129.09, -18.29)$, we are 90% confident that $\mu_1 < \mu_2$, that is, we are 90% confident that the new method relative to the standard method reduces the mean operative time by 18.29 minutes to 129.09 minutes.

Two sample T confidence interval:

μ_1 : Mean of New Method
μ_2 : Mean of Standard Method
$\mu_1 - \mu_2$: Difference between two means
(without pooled variances)

$$df = \min(n_1 - 1, n_2 - 1) = \min(13, 5) = 5$$

$$t_5^* = 2.015$$

90% confidence interval results:

Difference	Sample Diff.	Std. Err.	DF	L. Limit	U. Limit
$\mu_1 - \mu_2$	-73.690476	27.492134	17.83231	-121.38788	-25.993069

$(-121.39, -25.99) \longrightarrow$ narrower

Example: Researchers wanted to compare the sense of direction of adult males and adult females, in particular, their frame of reference. They randomly selected 30 adult males and 30 adult females and took them to an unfamiliar park. One at a time, each person was separated from the group, given a pointer attached to a large 360° protractor, and was asked to point south. The number degrees that each person was off from south was recorded. The summary statistics are given below:

	Males (pop 1)	Females (pop 2)
Sample size	30	30
Sample Mean	37.6	55.8
Sample Standard Deviation	38.49	48.26

$$\frac{\text{larger } s}{\text{smaller } s} = \frac{48.26}{38.49} = 1.25 < 2$$

$\Rightarrow$ assume $\sigma_1 = \sigma_2$ and use pooled t-test

a) At the 1% significance level, do the data provide enough statistical evidence to conclude that, on average, males have a better sense of direction and, in particular, a better frame of reference than females?

1. **Assumptions/Conditions:**

- assume $\sigma_1 = \sigma_2$
- random samples (assume independence)
- assume independent groups
- both $n_1 = 30 \geq 30$ and $n_2 = 30 \geq 30$

2. **Hypotheses:**

$$H_0: \mu_1 - \mu_2 = 0$$
$$H_A: \mu_1 - \mu_2 < 0$$

lower-tailed test

3. **Test Statistic:**

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

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

$$= \sqrt{\frac{(30-1)(38.49)^2 + (30-1)(48.26)^2}{30+30-2}}$$

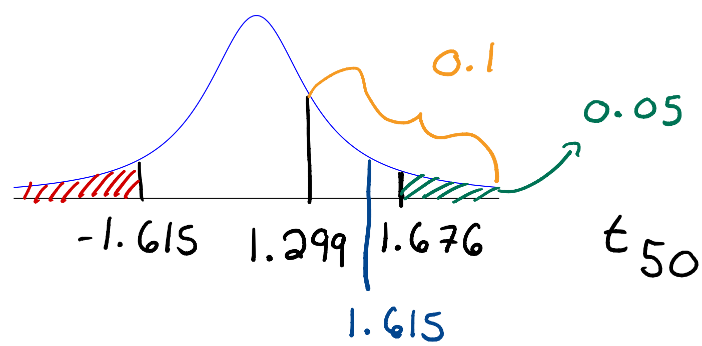
$$= 43.649$$

$$= \frac{37.6 - 55.8}{43.649\sqrt{\frac{1}{30} + \frac{1}{30}}}$$

$$= -1.615$$

4. ***P*-value:**

$$df = n_1 + n_2 - 2$$
$$= 30 + 30 - 2$$
$$= 58 \longrightarrow \text{use } 50$$



Using t-table,

$$0.01 < 0.05 < P\text{-value} = P(t_{58} < -1.615) < 0.1$$

Two sample T hypothesis test:

μ_1 : Mean of Males
μ_2 : Mean of Females
$\mu_1 - \mu_2$: Difference between two means
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 < 0$
(with pooled variances)

Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	-18.2	11.269734	58	-1.614945	0.0559

> 0.01

5. **Conclusion:**

Since P-value $> \alpha = 0.01$, we do not reject H_0 at the 0.01 significance level, that is, there is not enough statistical evidence to conclude that, on average, males have a better sense of direction and, in particular, a better frame of reference than females.

b) Construct a 98% confidence interval for the difference $\mu_1 - \mu_2$.

$$(\bar{y}_1 - \bar{y}_2) \pm t^*_{df} \, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= (37.6 - 55.8) \pm (2.403)(43.649)\sqrt{\frac{1}{30} + \frac{1}{30}}$$

$$= -18.2 \pm 27.082$$

$$= (-45.28, 8.88) \ni 0$$

possible

$\mu_1 = \mu_2$

$df = n_1 + n_2 - 2$

$= 30 + 30 - 2$

$= 58$

$t^*_{50} = 2.403$

Two sample T confidence interval:

μ_1 : Mean of Males
μ_2 : Mean of Females
$\mu_1 - \mu_2$: Difference between two means
(with pooled variances)

98% confidence interval results:

Difference	Sample Diff.	Std. Err.	DF	L. Limit	U. Limit
$\mu_1 - \mu_2$	-18.2	11.269734	58	-45.161457	8.761457

$(-45.16, 8.76)$