Lecture 12: Comparison of population means

MVE055 / MSG810
Mathematical statistics and discrete mathematics

Moritz Schauer

Last updated October 19, 2021

GU & Chalmers University of Technology

Comparisons

Comparisons

A common situation is that you want to make comparisons between different samples. Examples of when this may be of interest include

- We want to compare performances of two designs.
- We want to investigate the effect of a new drug.

Comparisons

A common situation is that you want to make comparisons between different samples. Examples of when this may be of interest include

- We want to compare performances of two designs.
- We want to investigate the effect of a new drug.

Today we will examine two types of comparisons

- Independent samples (measurements of two populations)
- Paired samples (samples are pairs of related measurements)

Paired samples

Paired samples:

A common situation is that the measurements are made in pairs.

Paired samples:

A common situation is that the measurements are made in pairs. For example when you take different measurements of the same subjects, e.g. strength of the right arm and strength of the left arm.

Paired samples:

A common situation is that the measurements are made in pairs. For example when you take different measurements of the same subjects, e.g. strength of the right arm and strength of the left arm.

We set up a model which has n pairs of observations

$$X_1, Y_1, \quad X_2, Y_2, \quad \ldots, \quad X_n, Y_n.$$

For each measurement, we form the difference, which is assumed to be normally distributed:

$$D_i = X_i - Y_i \overset{iid}{\sim} \mathsf{N}(\mu_{\mathrm{diff}}, \sigma^2)$$

Paired samples:

A common situation is that the measurements are made in pairs. For example when you take different measurements of the same subjects, e.g. strength of the right arm and strength of the left arm.

We set up a model which has n pairs of observations

$$X_1, Y_1, \quad X_2, Y_2, \quad \ldots, \quad X_n, Y_n.$$

For each measurement, we form the difference, which is assumed to be normally distributed:

$$D_i = X_i - Y_i \overset{iid}{\sim} \mathsf{N}(\mu_{\text{diff}}, \sigma^2)$$

Summary: We test whether $H_0\colon \mu_{\text{diff}} = 0$ against an alternative. This is done as usual for normally distributed measurements with known or unknown variance.

Independent samples

Independent samples

Assume we have two independent samples from different populations:

- n_1 observations $X_1, X_2, \ldots, X_{n_1}$ from $\mathsf{N}(\mu_1, \sigma_1^2)$.
- Also n_2 observations $Y_1, Y_2, \ldots, Y_{n_2}$ from $\mathsf{N}(\mu_2, \sigma_2^2)$.

Summary: Build test/CI for $H :: \mu_1 - \mu_2$. We'll start with estimator $\bar{D} = \bar{X} - \bar{Y}$ of $\mu_1 - \mu_2$.

Paired or not

1. Compare pre-class (beginning of semester) and post-class (end of semester) scores of students. $\boxed{}$
2. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 12 women.
3. Compare artery thicknesses at the beginning of a study and after 2 years of taking Vitamin E for the same group of patients.
4. Measure the strength of the left arm vs right arm of each subject.
5. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 10 women.

Paired or not

1. Compare pre-class (beginning of semester) and post-class (end of semester) scores of students. Paired.
2. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 12 women.
3. Compare artery thicknesses at the beginning of a study and after 2 years of taking Vitamin E for the same group of patients.
4. Measure the strength of the left arm vs right arm of each subject.
5. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 10 women.

Paired or not

1. Compare pre-class (beginning of semester) and post-class (end of semester) scores of students. Paired.
2. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 12 women.
3. Compare artery thicknesses at the beginning of a study and after 2 years of taking Vitamin E for the same group of patients.
4. Measure the strength of the left arm vs right arm of each subject.
5. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 10 women.

Paired or not

1. Compare pre-class (beginning of semester) and post-class (end of semester) scores of students. Paired.
2. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 12 women. Independent.
3. Compare artery thicknesses at the beginning of a study and after 2 years of taking Vitamin E for the same group of patients.
4. Measure the strength of the left arm vs right arm of each subject.
5. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 10 women.

Paired or not

1. Compare pre-class (beginning of semester) and post-class (end of semester) scores of students. Paired.
2. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 12 women. Independent.
3. Compare artery thicknesses at the beginning of a study and after 2 years of taking Vitamin E for the same group of patients. $\boxed{}$
4. Measure the strength of the left arm vs right arm of each subject.
5. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 10 women.

Paired or not

1. Compare pre-class (beginning of semester) and post-class (end of semester) scores of students. Paired.
2. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 12 women. Independent.
3. Compare artery thicknesses at the beginning of a study and after 2 years of taking Vitamin E for the same group of patients. Paired.
4. Measure the strength of the left arm vs right arm of each subject.
5. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 10 women.

Paired or not

1. Compare pre-class (beginning of semester) and post-class (end of semester) scores of students. Paired.
2. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 12 women. Independent.
3. Compare artery thicknesses at the beginning of a study and after 2 years of taking Vitamin E for the same group of patients. Paired.
4. Measure the strength of the left arm vs right arm of each subject. $\boxed{}$
5. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 10 women.

Paired or not

1. Compare pre-class (beginning of semester) and post-class (end of semester) scores of students. Paired.
2. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 12 women. Independent.
3. Compare artery thicknesses at the beginning of a study and after 2 years of taking Vitamin E for the same group of patients. Paired.
4. Measure the strength of the left arm vs right arm of each subject. Paired.
5. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 10 women.

Paired or not

1. Compare pre-class (beginning of semester) and post-class (end of semester) scores of students. Paired.
2. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 12 women. Independent.
3. Compare artery thicknesses at the beginning of a study and after 2 years of taking Vitamin E for the same group of patients. Paired.
4. Measure the strength of the left arm vs right arm of each subject. Paired.
5. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 10 women.

Paired or not

1. Compare pre-class (beginning of semester) and post-class (end of semester) scores of students. Paired.
2. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 12 women. Independent.
3. Compare artery thicknesses at the beginning of a study and after 2 years of taking Vitamin E for the same group of patients. Paired.
4. Measure the strength of the left arm vs right arm of each subject. Paired.
5. Assess gender-related salary gap by comparing salaries of 10 randomly sampled men and 10 women. Still independent.

Example for $\boxed{}$ samples

You would like to know whether a new wheat variety yields a higher harvest than the existing variety. You select six fields that differ in fertility and climate, and divide each field into two parts in which each variety is grown.

Example for $\boxed{}$ samples

You would like to know whether a new wheat variety yields a higher harvest than the existing variety. You select six fields that differ in fertility and climate, and divide each field into two parts in which each variety is grown.

Field nr	1	2	3	4	5	6
Harvest sort 1, kg/ha	7529	8913	6534	6503	6896	8023
Harvest sort 2, kg/ha	7239	8726	6129	6351	6644	7711
Difference D_i	290	187	405	152	252	312

Example for $\boxed{}$ samples

You would like to know whether a new wheat variety yields a higher harvest than the existing variety. You select six fields that differ in fertility and climate, and divide each field into two parts in which each variety is grown.

Field nr	1	2	3	4	5	6
Harvest sort 1, kg/ha	7529	8913	6534	6503	6896	8023
Harvest sort 2, kg/ha	7239	8726	6129	6351	6644	7711
Difference D_i	290	187	405	152	252	312

We test $H_0 : \mu_{\text{diff}} = 0$ against $H_1 : \mu_{\text{diff}} \neq 0$ at level $\alpha = 0.05$.

Example for $\boxed{}$ samples

You would like to know whether a new wheat variety yields a higher harvest than the existing variety. You select six fields that differ in fertility and climate, and divide each field into two parts in which each variety is grown.

Field nr	1	2	3	4	5	6
Harvest sort 1, kg/ha	7529	8913	6534	6503	6896	8023
Harvest sort 2, kg/ha	7239	8726	6129	6351	6644	7711
Difference D_i	290	187	405	152	252	312

We test $H_0 : \mu_{\text{diff}} = 0$ against $H_1 : \mu_{\text{diff}} \neq 0$ at level $\alpha = 0.05$. We have $\bar{D} = 266.3$ and $s_D = 91$ and look up $t_{0.025}(5) = 2.57$

$$I_{\mu_{\text{diff}}} = (\bar{D} \pm t_{0.025}(5) \cdot s_D/\sqrt{6}) = (171, 362)$$

As $0 \notin I_{\mu_{\text{diff}}}$ we reject H_0.

Independent samples

Assume we have two independent samples

- n_1 observations $X_1, X_2, \ldots, X_{n_1}$ from $\mathsf{N}(\mu_1, \sigma_1^2)$.
- Also n_2 observations $Y_1, Y_2, \ldots, Y_{n_2}$ from $\mathsf{N}(\mu_2, \sigma_2^2)$.

We want to test wether μ_1 and μ_2 differ ($H_0 \colon \mu_1 = \mu_2$).

Independent samples

Assume we have two independent samples

- n_1 observations $X_1, X_2, \ldots, X_{n_1}$ from $\mathsf{N}(\mu_1, \sigma_1^2)$.
- Also n_2 observations $Y_1, Y_2, \ldots, Y_{n_2}$ from $\mathsf{N}(\mu_2, \sigma_2^2)$.

We want to test wether μ_1 and μ_2 differ ($H_0\colon \mu_1 = \mu_2$).

Introduce $\mu_{\text{diff}} = \mu_1 - \mu_2$ with estimator $\bar{D} = \bar{X} - \bar{Y}$. Test

$$
\begin{aligned}
&H_0\colon \mu_{\text{diff}} = 0, \\
&H_1\colon \mu_{\text{diff}} \neq 0 \qquad\qquad \text{(or against } H_1\colon \mu_{\text{diff}} > 0, \text{ or ...)}
\end{aligned}
$$

But what is the standard error??

3 cases

We distinguish between 3 cases:

Case 1: σ_1 and σ_2 are known.

Case 2: $\sigma_1 = \sigma_2 = \sigma$ where σ is unknown.

Case 3: σ_1 and σ_2 are unknown and not necessarily the same.

If the case is not known, we may first have to test whether $\sigma_1 = \sigma_2$ with the

Preliminary test:

$$H_0 : \frac{\sigma_1}{\sigma_2} = 1$$
$$H_1 : \frac{\sigma_1}{\sigma_2} \neq 1$$

Case 1: Known σ_1 and σ_2

If σ_1 and σ_2 are known it holds that

$$\mathrm{SE} = \mathrm{SE}(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

Case 1: Known σ_1 and σ_2

If σ_1 and σ_2 are known it holds that

$$\mathrm{SE} = \mathrm{SE}(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

In a hypothesis test we use that under H_0

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\mathrm{SE}} \sim \mathsf{N}(0,1)$$

with p-value $p = 2(1 - \Phi(|Z_{obs}|))$.

Case 1: Known σ_1 and σ_2

If σ_1 and σ_2 are known it holds that

$$\mathrm{SE} = \mathrm{SE}(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

In a hypothesis test we use that under H_0

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\mathrm{SE}} \sim \mathsf{N}(0,1)$$

with p-value $p = 2(1 - \Phi(|Z_{obs}|))$.

A confidence interval for $\mu_{\text{diff}} = \mu_1 - \mu_2$ is given by

$$I_{\mu_{\text{diff}}} = \left(\hat{\mu}_{\text{diff}} \pm z_{\alpha/2}\,\mathrm{SE}\right) = \left(\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

Case 2: $\sigma_1 = \sigma_2 = \sigma$ where σ unknown

Pooled estimate of variance

For 2 normally distributed samples $\mathsf{N}(\mu_j, \sigma^2), j = 1, 2$ an unbiased estimate of σ^2 is

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}. \quad \text{Step 1!}$$

With

$$\mathrm{SE} = \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \quad \text{Step 2!}$$

one has under H_0 that

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\mathrm{SE}} \sim t(n_1 + n_2 - 2)$$

Case 2: $\sigma_1 = \sigma_2 = \sigma$ where σ unknown

Pooled estimate of variance

For 2 normally distributed samples $\mathsf{N}(\mu_j, \sigma^2), j = 1, 2$ an unbiased estimate of σ^2 is

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}. \quad \text{Step 1!}$$

With

$$\mathrm{SE} = \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \quad \text{Step 2!}$$

one has under H_0 that

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\mathrm{SE}} \sim t(n_1 + n_2 - 2)$$

Confidence interval: $I_{\mu_{\mathrm{diff}}} = \left(\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}(n_1 + n_2 - 2)\,\mathrm{SE} \right)$.

Case 3: $\sigma_1 \neq \sigma_2$ unknown

Theorem

For two normally distributed samples

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

is approximately $t(df)$-distributed where

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

Case 3: $\sigma_1 \neq \sigma_2$ unknown

Theorem

For two normally distributed samples

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

is approximately $t(df)$-distributed where

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

We can now create confidence intervals and perform hypothesis tests in the same way as before:
$I_{\mu_{\text{diff}}} = \left(\hat{\mu}_{\text{diff}} \pm t_{\alpha/2}(f)\sqrt{s_1^2/n_1 + s_2^2/n_2}\right)$.

Example (Exercise 10.14)

To decide whether or not to purchase a new hand-held laser scanner for use in inventorying stock, tests are conducted on the scanner currently in use and on the new scanner. There data are obtained on the number of 7-inch bar codes that can be scanned per second:

new	old
$n_1 = 61$	$n_2 = 61$
$\bar{x}_1 = 40$	$\bar{x}_2 = 29$
$s_1^2 = 24.9$	$s_2^2 = 22.7$

Example (Exercise 10.14)

To decide whether or not to purchase a new hand-held laser scanner for use in inventorying stock, tests are conducted on the scanner currently in use and on the new scanner. There data are obtained on the number of 7-inch bar codes that can be scanned per second:

new	old
$n_1 = 61$	$n_2 = 61$
$\bar{x}_1 = 40$	$\bar{x}_2 = 29$
$s_1^2 = 24.9$	$s_2^2 = 22.7$

1. Find the pooled variance.
2. Find a 90% CI on $\mu_1 - \mu_2$.
3. Does the new laser appear to read more bar codes per second on the average?

1. Find the pooled variance.

1. Find the pooled variance.

$$s_2^p = \frac{60(24.9) + 60(22.7)}{120} = 23.8$$

1. Find the pooled variance.

$$s_2^p = \frac{60(24.9) + 60(22.7)}{120} = 23.8$$

2. Find a 90% CI on $\mu_1 - \mu_2$.

1. Find the pooled variance.

$$s_2^p = \frac{60(24.9) + 60(22.7)}{120} = 23.8$$

2. Find a 90% CI on $\mu_1 - \mu_2$.
t-distribution with $df = 120$. $t_{\alpha/2} = t_{0.05} = 1.658$ (note that the table does not give the values for degrees of freedom greater than 100 , use then an approximation). A 90% CI is therefore

$$(40 - 29 \pm 1.658\sqrt{23.8(1/61 + 1/61)}) = (9.54, 12.45)$$

1. Find the pooled variance.

$$s_2^p = \frac{60(24.9) + 60(22.7)}{120} = 23.8$$

2. Find a 90% CI on $\mu_1 - \mu_2$.
t-distribution with $df = 120$. $t_{\alpha/2} = t_{0.05} = 1.658$ (note that the table does not give the values for degrees of freedom greater than 100 , use then an approximation). A 90% CI is therefore

$$(40 - 29 \pm 1.658\sqrt{23.8(1/61 + 1/61)}) = (9.54, 12.45)$$

3. Does the new laser appear to read more bar codes per second on the average?

1. Find the pooled variance.

$$s_2^p = \frac{60(24.9) + 60(22.7)}{120} = 23.8$$

2. Find a 90% CI on $\mu_1 - \mu_2$.
t-distribution with $df = 120$. $t_{\alpha/2} = t_{0.05} = 1.658$ (note that the table does not give the values for degrees of freedom greater than 100 , use then an approximation). A 90% CI is therefore

$$(40 - 29 \pm 1.658\sqrt{23.8(1/61 + 1/61)}) = (9.54, 12.45)$$

3. Does the new laser appear to read more bar codes per second on the average?

Yes, since the interval does not contain 0 and is positive-valued.

Preliminary test: Comparison of variance

Denote with $F_\alpha(df_1, df_2)$ the α-quantile of the F-distribution. A confidence interval for σ_1^2/σ_2^2 is

$$I_{\sigma_1^2/\sigma_2^2} = \left[\frac{s_1^2/s_2^2}{F_{\alpha/2}(n_1-1, n_2-1)}, \frac{s_1^2/s_2^2}{F_{1-\alpha/2}(n_1-1, n_2-1)}\right]$$

Preliminary test: Comparison of variance

Denote with $F_\alpha(df_1, df_2)$ the α-quantile of the F-distribution. A confidence interval for σ_1^2/σ_2^2 is

$$I_{\sigma_1^2/\sigma_2^2} = \left[\frac{s_1^2/s_2^2}{F_{\alpha/2}(n_1-1, n_2-1)}, \frac{s_1^2/s_2^2}{F_{1-\alpha/2}(n_1-1, n_2-1)}\right]$$

Use for a hypothesis test $H_0\colon \sigma_1^2/\sigma_2^2 = 1$ (same as $H_0\colon \sigma_1^2 = \sigma_2^2$)
.