# CHAPTER 8 – Hypothesis testing two-sample inference

Monday, 15 December 2025 1:23 PM

* We are comparing a parameter of 2 samples of 2 different populations

Studies
- longitudinal (follow up)
- cross sectional

* A longitudinal (follow up) study is where the same group is observed repeatedly over along period of time to see how things change

* A cross-section study is a study which uses 2 samples of one population and compares them

Samples
- Paired (dependent): the two sets of data are linked or come from the same individuals
- independent: the 2 groups are completely unrelated and people in each group are different

(if it could be taken from the same population or different populations)

* In this chapter we are making hypothesis testing for 2 samples

- paired (dependent) → Paired t test
- independent → un Paired t test

we use t test

---

**Paired t test** → the parameter being tested is the difference of means $\Delta$

→ we do a test on a sample in the D distribution ($\bar{d}$ belong to D) = measurement$_1$ − measurement$_2$

this is how a Paired t-test reduces the problem to a one-Sample t-test on the differences

* * * for the Population of differences:-
- the mean is $\Delta$
- the variance is $\sigma_d^2$

for a Sample
- the mean is $\bar{d}$
- the variance is $S_d^2$

* why do we make a distribution of the differences and find its mean and variance?
↳ to test the hypotheses about the parameter $\mu_d$

- Critical value method
- P-value method

* $T.S = \dfrac{\bar{d} - \Delta}{S_d/\sqrt{n}}$ (the population difference; under $H_0: \Delta = 0$ we cancel it out)

$$T.S = \frac{\bar{d}}{S_d/\sqrt{n}}$$

---

✱ Critical value method :)

① decide the hypotheses
$H_0: \Delta = 0$
$H_1: \Delta \neq 0$ → 2 sided, $\Delta > 0$, $\Delta < 0$

② we find the test statistic

③ find the critical values

$$t_{n-1,\,1-\alpha/2} \quad , \quad -t_{n-1,\,1-\alpha/2}$$

for a 2-sided test

④ Make a decision

two-sided

$-t_{n-1,1-\alpha/2} \leq t.s \leq t_{n-1,1-\alpha/2}$ ⇒ accept $H_0$

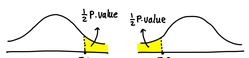
$t.s < -t_{n-1,1-\alpha/2}$ or $t.s > t_{n-1,1-\alpha/2}$ ⇒ reject $H_0$




✱ P-value method ⇒ P.value $< \alpha$ ⇝ reject $H_0$; P.value $> \alpha$ ⇝ accept $H_0$

two-sided

$t.s > 0$: P.value $= 2P(t > t.s)$
$t.s \leq 0$: P.value $= 2P(t < t.s)$




* * *

✱ Finding a confidence interval for the True difference $\Delta$ in 2 Paired Samples

$$(1-\alpha)\%\ CI = \bar{d} \pm t_{n-1,\,1-\alpha/2} * \frac{S_d}{\sqrt{n}}$$

$\bar{d}$: mean of the Sample of differences; $S_d$: standard deviation of a sample of the D distribution

---

⚠ $\Delta > 0$ ↳ After − Before; $\Delta < 0$ ↳ Before − After

in 2-Sided tests it doesn't matter which to use / we only care for the difference ⇓ both After−Before give the same answer

---

**un Paired t test** → the parameter being tested here is the difference between the means of 2 samples $\bar{X}_1 - \bar{X}_2$

→ We make a $\bar{X}_1 - \bar{X}_2$ normally distributed distribution with mean $\mu_1 - \mu_2$ and variance $\sigma^2 * \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$

the variance has to be the same for both Samples even if it was unknown

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

→ under $H_0$ condition where $\mu_1 = \mu_2$ then

$$\bar{X}_1 - \bar{X}_2 \sim N\left(0,\ \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

↳ to get the standardized form

$$\frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$$

σ is known here

* if $\sigma$ is unknown then it is estimated using the samples variances $S_1$ and $S_2$ by the pooled estimate of the variance

$$S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

---

* The test statistic which is going to be used to test the hypotheses $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$

$$t.s = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S = \sqrt{(n_1-1)S_1^2 + (n_2-1)S_2^2 / (n_1+n_2-2)}$$

---

✱ The Critical value method

① decide the hypotheses $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$

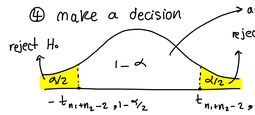
② find the test statistic

③ find the critical values
* note that the degree of freedom here is $n_1+n_2-2$

$$-t_{n_1+n_2-2,\,1-\alpha/2} \quad , \quad t_{n_1+n_2-2,\,1-\alpha/2}$$

for a 2-sided test

④ make a decision



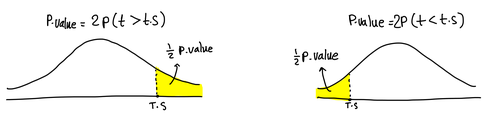

$t.s < -t_{n_1+n_2-2,\,1-\alpha/2}$ OR ⇒ reject $H_0$

$t.s > t_{n_1+n_2-2,\,1-\alpha/2}$

$-t_{n_1+n_2-2,\,1-\alpha/2} \leq t.s \leq t_{n_1+n_2-2,\,1-\alpha/2}$ ⇒ accept $H_0$

✱ P-value method ⇒ P.value $> \alpha$ ⇝ accept $H_0$; P.value $< \alpha$ ⇝ reject $H_0$

$t.s > 0$: P.value $= 2P(t > t.s)$
$t.s \leq 0$: P.value $= 2P(t < t.s)$




---

✱ Finding a confidence interval for the difference $\bar{X}_1 - \bar{X}_2$

$$(1-\alpha)\%\ CI = \bar{X}_1 - \bar{X}_2 \pm t_{n_1+n_2-2,\,1-\alpha/2} * S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$S = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

if the CI interval is wide then it isn't accurate and a much larger sample is needed to accurately assess the true mean difference

---

■ E.G.

**Hypertension** Suppose a sample of eight 35- to 39-year-old nonpregnant, premenopausal OC users who work in a company and have a mean systolic blood pressure (SBP) of 132.86 mm Hg and sample standard deviation of 15.34 mm Hg are identified. A sample of 21 nonpregnant, premenopausal, non-OC users in the same age group are similarly identified who have mean SBP of 127.44 mm Hg and sample standard deviation of 18.23 mm Hg. What can be said about the underlying mean difference in blood pressure between the two groups?

$n_1$

the hypotheses ⇒ $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$
$H_0: \mu_1 - \mu_2 = 0$ vs $H_1: \mu_1 - \mu_2 \neq 0$

1 using the critical value method

$$T.S = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (\text{it doesn't matter if you take } \bar{X}_1 - \bar{X}_2 \text{ or } \bar{X}_2 - \bar{X}_1)$$

$$S = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

$$= \frac{132.86 - 127.44}{\sqrt{\frac{1}{8} + \frac{1}{21}}\sqrt{\frac{7*(15.34)^2 + 20*(18.23)^2}{8+21-2}}} = \frac{5.42}{(17.50)(0.4154)} = 0.74$$

→ Critical values

$$-t_{n_1+n_2-2,\,1-\alpha/2} \quad , \quad t_{n_1+n_2-2,\,1-\alpha/2}$$

$$-t_{27,\,0.975} \quad , \quad t_{27,\,0.975}$$

$$-2.052 \quad , \quad 2.052$$

→ T.S > Critical value
$0.74 > -2.052$ ∴ accept $H_0$

2 Using the P-value method

→ T.S > 0

$$\text{P-value} = 2P(t > t.S) = 2P(t > 0.74)$$

$$\Rightarrow 0.4 < 2P(t < 0.74) < 0.5$$

$0.05 <$ P-value ∴ accept $H_0$

3 Finding a 95% CI

$$\bar{X}_1 - \bar{X}_2 \pm t_{n_1+n_2-2} * S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$132.86 - 127.44 \pm 2.052 * 17.52 * \sqrt{\frac{1}{8} + \frac{1}{21}}$$

the 95% CI

⚠ if the CI doesn't contain 0 then we reject $H_0: \Delta = 0$ : $\bar{X}_1 - \bar{X}_2 = 0$

if the CI contains 0 we accept $H_0: \Delta = 0$ : $\bar{X}_1 - \bar{X}_2 = 0$