Inferences About Two Means: Independent Samples

Independent Samples with σ_1 and σ_2 Unknown and Not Assumed Equal

Definitions

Two samples are independent if the sample values selected from one population are not related to or somehow paired or matched with the sample values from the other population.

Two samples are dependent if the sample values are *paired*. (That is, each pair of sample values consists of two measurements from the same subject (such as before/after data), or each pair of sample values consists of matched pairs (such as husband/wife data), where the matching is based on some inherent relationship.)

Notation

μ_1 = population mean

σ_1 = population standard deviation

n_1 = size of the first sample

$\bar{x}_1$ = sample mean

s_1 = sample standard deviation

Corresponding notations for μ_2, σ_2, s_2, $\bar{x}_2$ and n_2 apply to population 2.

Hypothesis Test for Two Means: Independent Samples

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

(where $\mu_1 - \mu_2$ is often assumed to be 0)

Hypothesis Test - cont

Test Statistic for Two Means: Independent Samples

Degrees of freedom: we use this simple and conservative estimate:

df = smaller of $n_1 - 1$ and $n_2 - 1$.

***P*-values:** Refer to student t table.

Critical values: Refer to student t table.

Confidence Interval Estimate of $\mu_1 - \mu_2$: Independent Samples

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$$

where $E = t_{\alpha/2} \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

where df = smaller $n_1 - 1$ and $n_2 - 1$

Caution

Before conducting a hypothesis test, consider the context of the data, the source of the data, the sampling method, and explore the data with graphs and descriptive statistics. Be sure to verify that the requirements are satisfied.

Example:

A headline in *USA Today* proclaimed that "Men, women are equal talkers." That headline referred to a study of the numbers of words that samples of men and women spoke in a day. Given below are the results from the study. Use a 0.05 significance level to test the claim that men and women speak the same mean number of words in a day. Does there appear to be a difference?

Number of Words Spoken in a Day

Men		Women	
n_1	= 186	n_2	= 210
$\bar{x}_1$	= 15,668.5	$\bar{x}_2$	= 16,215.0
s_1	= 8632.5	s_2	= 7301.2

Example:

Requirements are satisfied: two population standard deviations are not known and not assumed to be equal, independent samples, simple random samples, both samples are large.

Step 1: Formalise hypotheses

Alternative hypothesis does not contain equality, null hypothesis does.

$$H_0: \mu_1 = \mu_2 \qquad H_a: \mu_1 \neq \mu_2$$

Proceed assuming $\mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$.

Example:

Step 2: Significance level is 0.05

Step 3: Use a *t* distribution

Step 4: Calculate the test statistic

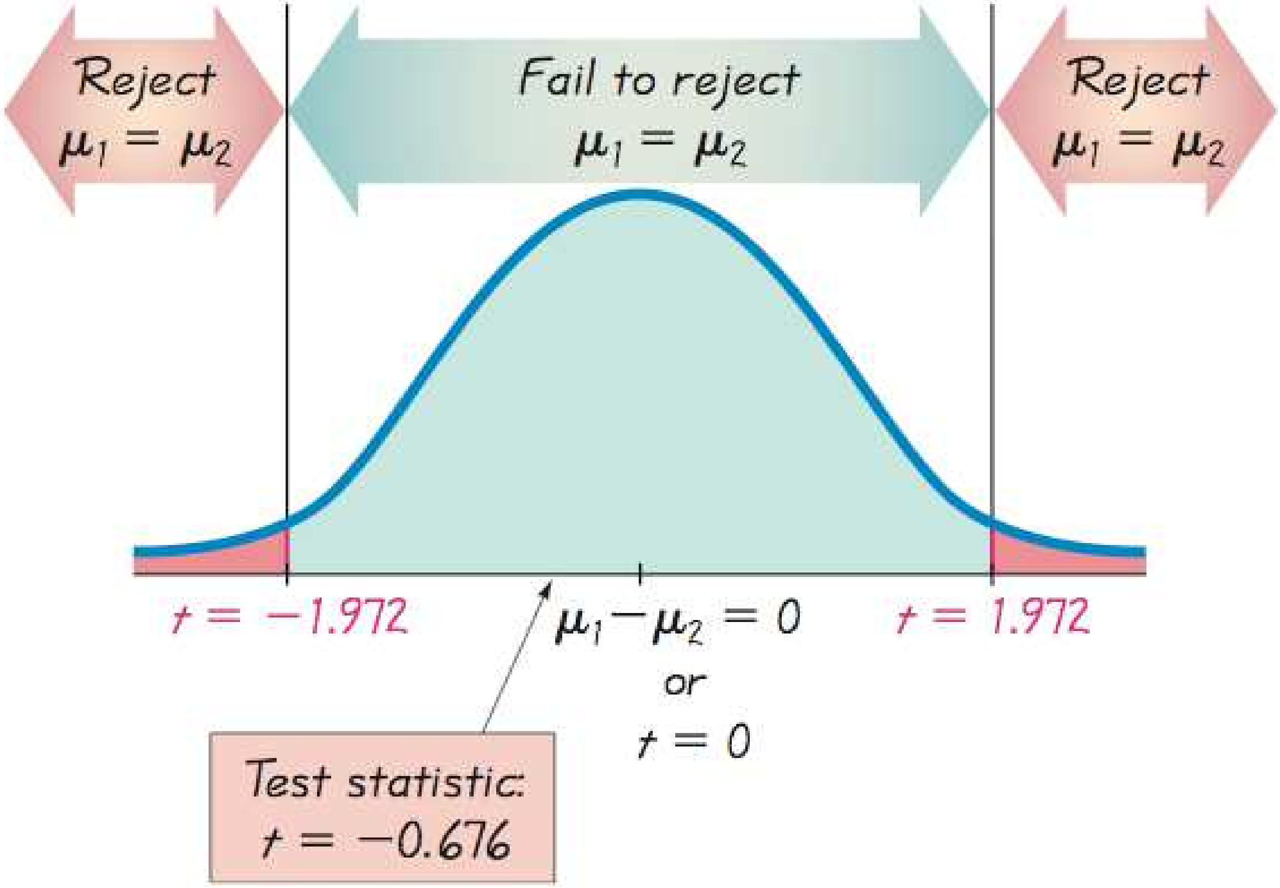
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{(15{,}668.5 - 16{,}215.0) - 0}{\sqrt{\frac{8632.5^2}{186} + \frac{7301.2^2}{210}}} = -0.676$$

Example:

Use Table A-3: area in two tails is 0.05, df = 185, which is not in the table, the closest value is

$t = \pm 1.972$



Example:

Step 5: **Because the test statistic does not fall within the critical region, fail to reject the null hypothesis:**
$\mu_1 = \mu_2$ **(or** $\mu_1 - \mu_2 = 0$**).**

There is not sufficient evidence to warrant rejection of the claim that men and women speak the same mean number of words in a day. There does not appear to be a significant difference between the two means.

Example:

Using the sample data given in the previous Example, construct a 95% confidence interval estimate of the difference between the mean number of words spoken by men and the mean number of words spoken by women.

Number of Words Spoken in a Day			
Men		Women	
$n_1 =$	186	$n_2 =$	210
$\bar{x}_1 =$	15,668.5	$\bar{x}_2 =$	16,215.0
$s_1 =$	8632.5	$s_2 =$	7301.2

Example:

Requirements are satisfied as it is the same data as the previous example.

Find the margin of Error, *E*; use $t_{\alpha/2}$ = 1.972

$$E = t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = 1.972\sqrt{\frac{8632.5^2}{186}+\frac{7301.2^2}{210}} = 1595.4$$

Construct the confidence interval use *E* = 1595.4 and $\bar{x}_1 = 15{,}668.5$ and $\bar{x}_2 = 16{,}215.0$.

$$\left(\bar{x}_1-\bar{x}_2\right)-E<\left(\mu_1-\mu_2\right)<\left(\bar{x}_1-\bar{x}_2\right)+E$$

$$-2141.9<\left(\mu_1-\mu_2\right)<1048.9$$

Independent samples assuming that $\sigma_1 = \sigma_2$ and Pool the Sample Variances.

Hypothesis Test Statistic for Two Means: Independent Samples and $\sigma_1 = \sigma_2$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}}$$

Where

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

and the number of degrees of freedom is df = $n_1 + n_2 - 2$

Confidence Interval Estimate of $\mu_1 - \mu_2$: Independent Samples with $\sigma_1 = \sigma_2$

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$$

where $$E = t_{\alpha/2} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

and number of degrees of freedom is df = $n_1 + n_2 - 2$

TEST t
two samples

Assuming equal variances

$$t_{n_1+n_2-2} = \frac{\bar{x}_1 - \bar{x}_2}{ES(\bar{x}_1 - \bar{x}_2)} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

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

$$df = (n_1 + n_2 - 2)$$

Not assuming equal variances

$$t_{\mathrm{MIN}(n_1-1,n_2-1)} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$df = \mathrm{MIN}(n_1 - 1, n_2 - 1)$$

Strategy

Unless instructed otherwise, use the following strategy:

Assume that σ_1 and σ_2 are unknown, do not assume that $\sigma_1 = \sigma_2$, and use the test statistic and confidence interval given in Part 1 of this lecture.

Exercise

Data Set 14 “Passive and Active Smoke” includes measures of cotinine (ng/mL) in subjects from different groups. Cotinine is produced when nicotine is absorbed by the body, so cotinine is a good indicator of nicotine. Listed below are the summary statistics from a group of smokers and another group of subjects who do not smoke but are exposed to environmental tobacco smoke at home or work.

Use a 0.05 significance level to test the claim that the population of smokers has a higher mean cotinine level than the nonsmokers exposed to smoke.
Do smokers appear to have higher of levels of cotinine than nonsmokers who are exposed to smoke?
If so, estimate, with 90% of confidence, the increased amount of cotinine in smokers, with respect to non smokers exposed to smoke

Smokers	n = 40, $\bar{x}$ = 172.5 ng/mL, s = 119.5 ng/mL
Nonsmokers Exposed to Smoke	n = 40, $\bar{x}$ = 60.6 ng/mL, s = 138.1 ng/mL

Exercise - assuming equal variances

$H_0: \mu_S = \mu_{NS}$
$H_1: \mu_S > \mu_{NS}$

α=0.05 Critical value $t_{40-1+40-1=78,0.95}$=1.664

$$s = \sqrt{\frac{39 * 119.5^2 + 39 * 138.1^2}{39 + 39}} = 129.14$$

$$t = \frac{172.5 - 60.6}{129.14\sqrt{\frac{1}{40}+\frac{1}{40}}} = 3.875$$

Reject H_0

P-value<0.005

It appears that smoking is associated with higher levels of cotinine than nonsmokers exposed to smoke.

$$IC90\%: (172.5 - 60.6) \pm 1.664 * 129.14\sqrt{\frac{1}{40} + \frac{1}{40}} = [63.9; 159.9]\text{ng/mL}$$

Exercise – not assuming equal variances

$H_0: \mu_S = \mu_{NS}$
$H_1: \mu_S > \mu_{NS}$
α=0.05 Critical value $t_{MIN(40-1,40-1)=39,0.95}$=1.685

$$t = \frac{172.5 - 60.6}{\sqrt{\frac{119.5^2}{40} + \frac{138.1^2}{40}}} = 3.875$$

Reject H_0

P-value<0.005

There is sufficient evidence to support the claim that the population of smokers has a higher mean cotinine level than the nonsmokers exposed to smoke.

$$IC90\%: (172.5 - 60.6) \pm 1.685\sqrt{\frac{119.5^2}{40} + \frac{138.1^2}{40}} = [63.2; 160.6]\text{ng/mL}$$

We are 90% confident that the limits of 63.2 ng/mL and 160.6 ng/mL actually do contain the difference between the two population means. Because those limits do not contain 0, this confidence interval suggests that the mean cotinine level of smokers is greater than the mean cotinine level of nonsmokers.