Two-sample problems for population means

BPS chapter 19

Objectives (BPS chapter 19)

Comparing two population means

- Two-sample *t* procedures
- Examples of two-sample *t* procedures
- Using technology
- Robustness again
- Details of the *t* approximation
- Avoid the pooled two-sample *t* procedures
- Avoid inference about standard deviations
- The *F* test for comparing two standard deviations

Conditions for inference comparing two means

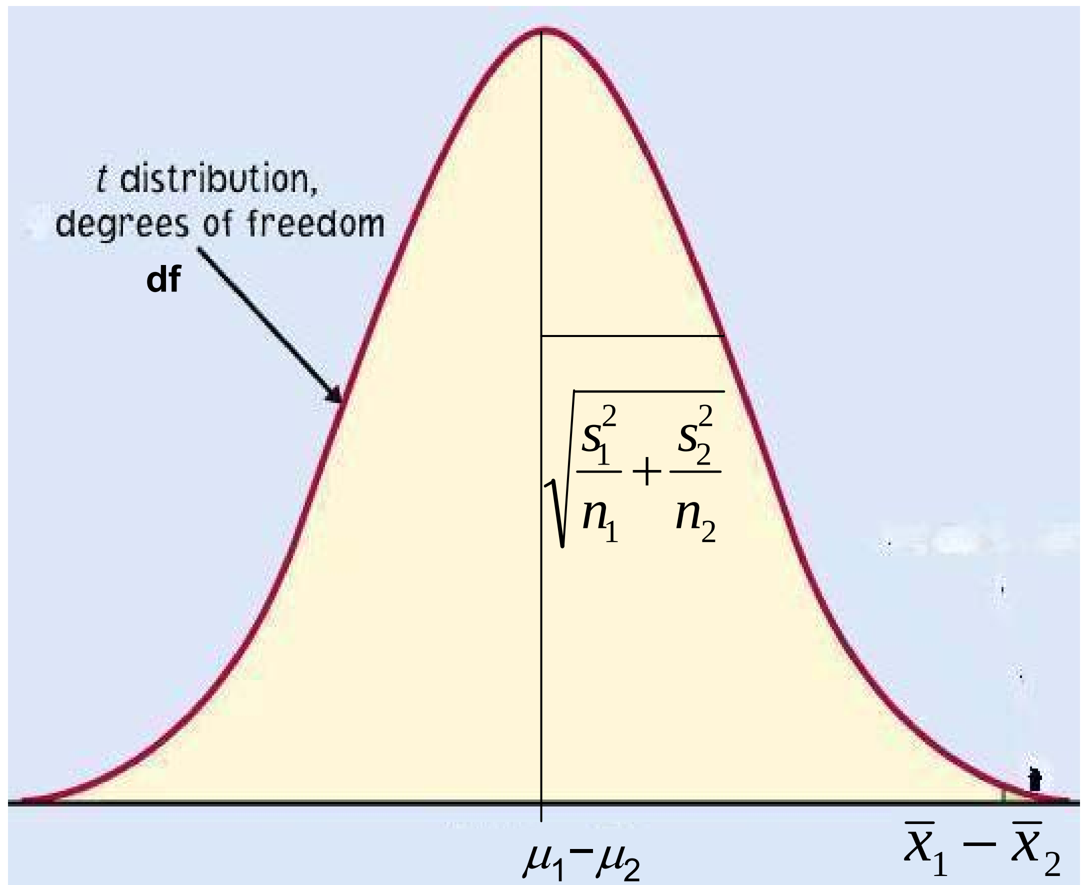
We have **two independent SRSs** (simple random samples) coming from two distinct populations (like men vs. women) with (μ_1, σ_1) and (μ_2, σ_2) unknown.

Both populations should be Normally distributed. However, in practice, it is enough that the two distributions have similar shapes and that the sample data contain no strong outliers.

The two-sample t statistic follows approximately the t distribution with a standard error SE (spread) reflecting variation from both samples:

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Conservatively, the degrees of freedom is equal to the smallest of (n_1−1, n_2−1).



Two-sample *t*-test

The null hypothesis is that both population means μ_1 and μ_2 are equal, thus their difference is equal to zero.

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

with either a one-sided or a two-sided alternative hypothesis.

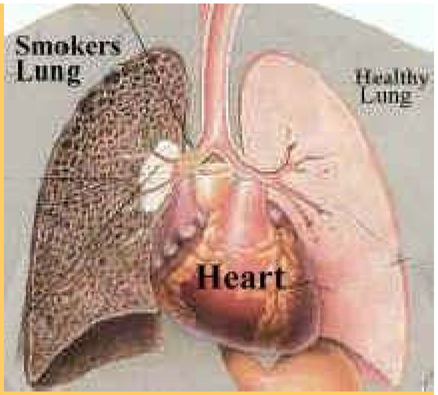
We find how many standard errors (SE) away from $(\mu_1 - \mu_2)$ is $(\bar{x}_1 - \bar{x}_2)$ by standardizing with t:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SE}$$

Because in a two-sample test H_0 poses $(\mu_1 - \mu_2) = 0$, we simply use

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

with df = smallest($n_1 - 1, n_2 - 1$)



Does smoking damage the lungs of children exposed to parental smoking?

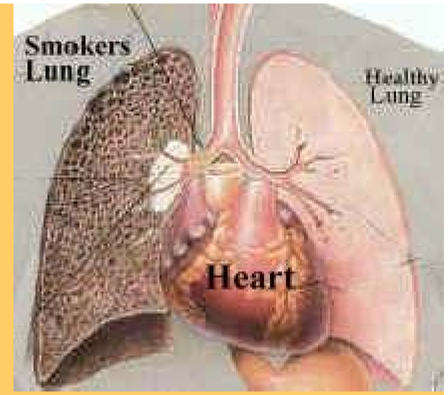
Forced Vital Capacity (FVC) is the volume (in milliliters) of air that an individual can exhale in 6 seconds.

FVC was obtained for a sample of children not exposed to parental smoking and a group of children exposed to parental smoking.

Parental smoking	FVC $\bar{x}$	s	n
Yes	75.5	9.3	30
No	88.2	15.1	30

We want to know whether parental smoking decreases children's lung capacity as measured by the FVC test.

Is the mean FVC lower in the population of children exposed to parental smoking?



H_0: $\mu_{smoke} = \mu_{no}$ <=> $(\mu_{smoke} - \mu_{no}) = 0$

H_a: $\mu_{smoke} < \mu_{no}$ <=> $(\mu_{smoke} - \mu_{no}) < 0$ (one sided)

The difference in sample averages follows approximately the *t* distribution: $t\left(0, \sqrt{\frac{s_{smoke}^2}{n_{smoke}} + \frac{s_{no}^2}{n_{no}}}\right)$, $df \;\; 29$

We calculate the t statistic:

$$t = \frac{\bar{x}_{smoke} - \bar{x}_{no}}{\sqrt{\frac{s_{smoke}^2}{n_{smoke}} + \frac{s_{no}^2}{n_{no}}}} = \frac{75.5 - 88.2}{\sqrt{\frac{9.3^2}{30} + \frac{15.1^2}{30}}}$$

Parental smoking	FVC $\bar{x}$	s	n
Yes	75.5	9.3	30
No	88.2	15.1	30

$$t = \frac{-12.7}{\sqrt{2.9 + 7.6}} \approx -3.9$$

In Table C, for df 29 we find:

$|t| > 3.659$ => $p < 0.0005$ (one-sided)

It's a very significant difference, we reject H_0.

Lung capacity is significantly impaired in children of smoking parents.

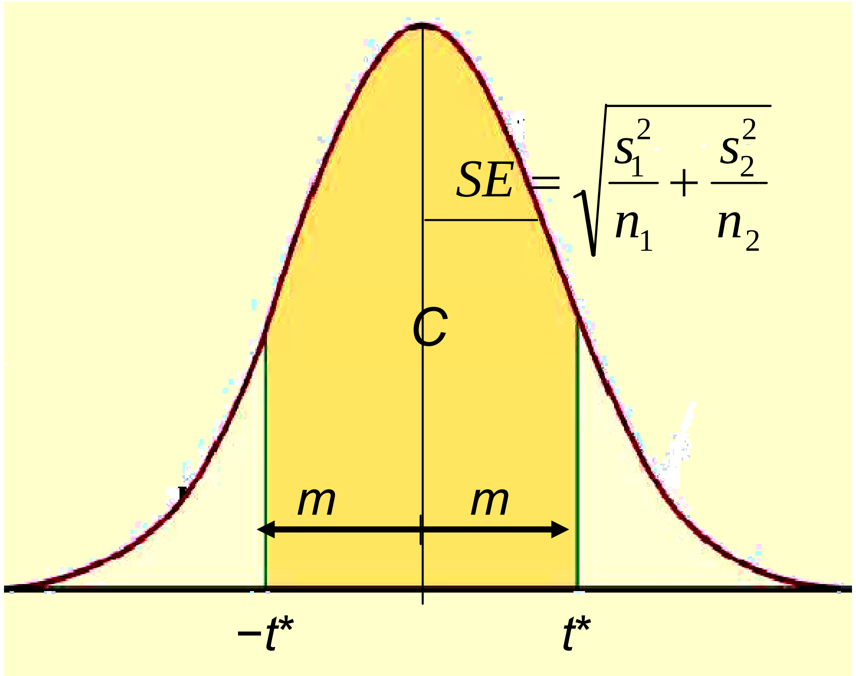
Two sample *t*-confidence interval

Because we have two independent samples we use the difference between both sample averages ($\bar{x}_1 - \bar{x}_2$) to estimate $(\mu_1 - \mu_2)$.

Practical use of *t*: *t**

- C is the area between −*t** and *t**.
- We find *t** in the line of Table C for df = smallest $(n_1 - 1; n_2 - 1)$ and the column for confidence level C.
- The margin of error *m* is:

$$m = t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = t^* SE$$



Common mistake!!!

A common mistake is to calculate a one-sample confidence interval for μ_1 and then check whether μ_2 falls within that confidence interval, or vice versa.

This is WRONG because the variability in the sampling distribution for two independent samples is more complex and must take into account variability coming from both samples—hence the more complex formula for the standard error.

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Can directed reading activities in the classroom help improve reading ability? A class of 21 third-graders participates in these activities for 8 weeks while a control classroom of 23 third-graders follows the same curriculum without the activities. After the 8 weeks, all children take a reading test (scores in table).

Treatment group				Control group			
24	61	59	46	42	33	46	37
43	44	52	43	43	41	10	42
58	67	62	57	55	19	17	55
71	49	54		26	54	60	28
43	53	57		62	20	53	48
49	56	33		37	85	42	

Group	n	$\bar{x}$	s
Treatment	21	51.48	11.01
Control	23	41.52	17.15

95% confidence interval for ($\mu_1 - \mu_2$), with df = 20 conservatively → t^* = 2.086:

$$CI: \ (\bar{x}_1 - \bar{x}_2) \pm m; \quad m = t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 2.086 * 4.31 \approx 8.99$$

With 95% confidence, (mu1 – mu2) falls within 9.96 ± 8.99 or 1.0 to 18.9.

20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Robustness

The two-sample statistic is the most robust when both sample sizes are equal and both sample distributions are similar. But even when we deviate from this, two-sample tests tend to remain quite robust.

As a guideline, a combined sample size ($n_1 + n_2$) of 40 or more will allow you to work even with the most skewed distributions.

Details of the two-sample *t* procedures

The **true value of the degrees of freedom** for a two-sample *t*-distribution is quite lengthy to calculate. That's why we use an approximate value, df = smallest($n_1 - 1$, $n_2 - 1$), which errs on the conservative side (often smaller than the exact).

Computer software, though, gives the exact degrees of freedom — or the rounded value — for your sample data.

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

95% confidence interval for the reading ability study using the more precise degrees of freedom:

$$df = \frac{\left(\frac{11.01^2}{21} + \frac{17.15^2}{23}\right)^2}{\frac{1}{20}\left(\frac{11.01^2}{21}\right)^2 + \frac{1}{22}\left(\frac{17.15^2}{23}\right)^2}$$

$$= \frac{344.486}{9.099} = 37.86$$

$$m = t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$m = 2.024 * 4.31 \approx 8.72$$

30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423
	50%	60%	70%	80%	90%	95%	96%	98%

Table C Confidence level C

Excel

t-Test: Two-Sample Assuming Unequal Variances

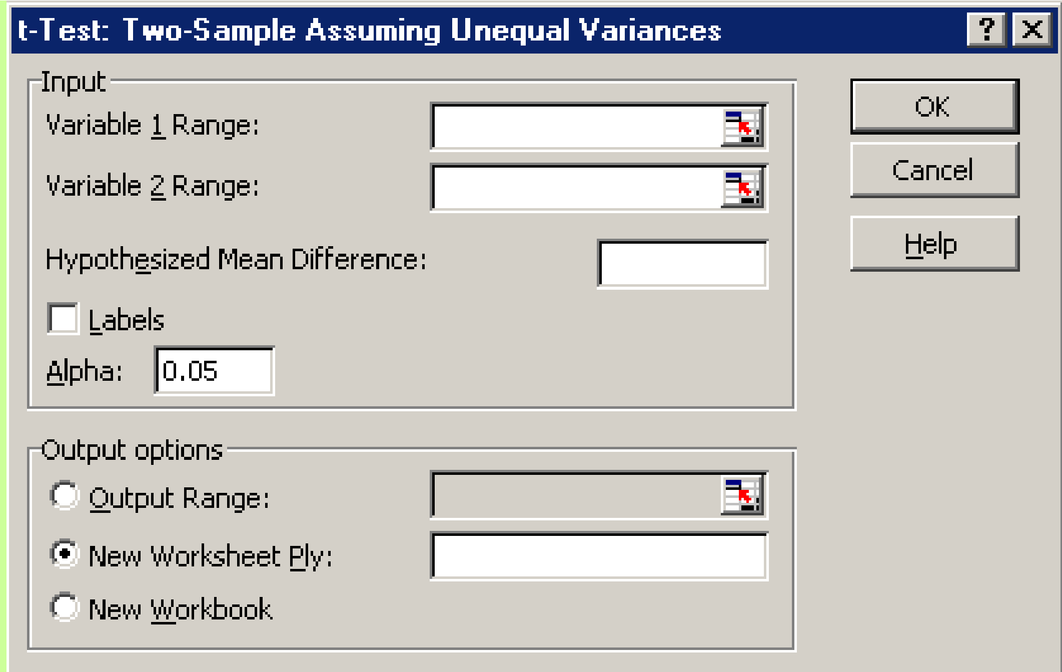
	Treatment group	*Control group*
Mean	51.476	41.522
Variance	121.162	294.079
Observations	21	23
Hypothesized Mean Difference	-	
df	38	
t Stat	2.311	
P(T<=t) one-tail	0.013	
t Critical one-tail	1.686	
P(T<=t) two-tail	0.026	
t Critical two-tail	2.024	t*

SPSS

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
ReadingScore	Equal variances assumed	2.362	.132	2.267	42	.029	9.95445	4.39189	1.09125	18.81765
	Equal variances not assumed			2.311	37.855	.026	9.95445	4.30763	1.23302	18.67588

Excel



menu/tools/data_analysis →

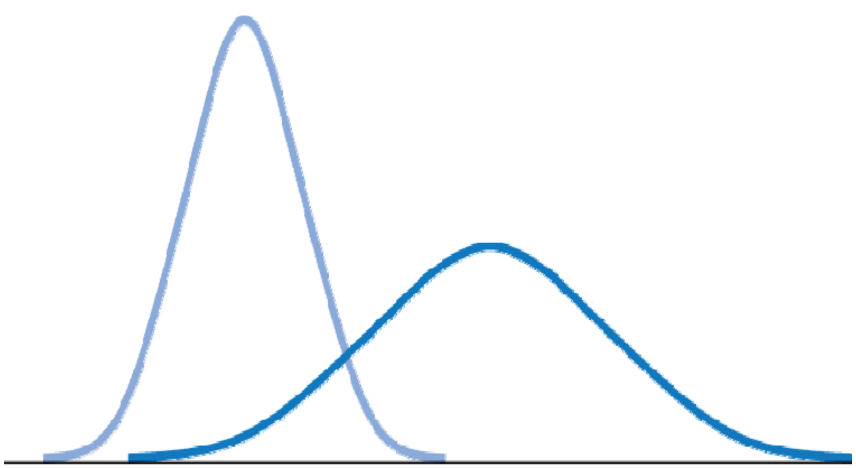
or

=TTEST(array1,array2,tails,type)

- *Array1* is the first data set.
- *Array2* is the second data set.
- *Tails* specifies the nature of the alternative hypothesis (1: one-tailed; 2: two-tailed).
- *Type* is the kind of *t*-test to perform (1: paired; 2: two-sample equal variance; 3: two-sample unequal variance).

Two-sample test assuming equal variance

There are two versions of the two-sample *t*-test: one **assuming equal variance ("pooled two-sample test")** and one **not assuming equal variance ("unequal" variance)** for the two populations. You may have noticed slightly different formulas and degrees of freedom.



Two normally distributed populations with unequal variances

The pooled (equal variance) two-sample *t*-test was often used before computers because it has exactly the *t* distribution for degrees of freedom $n_1 + n_2 - 2$.

However, the assumption of equal variance is hard to check, and thus ***the unequal variance test is safer***.

Which type of test? One sample, paired samples, two samples?

- Comparing vitamin content of bread, immediately after baking versus 3 days later (the same loaves are used on day one and 3 days later).
- Comparing vitamin content of bread, immediately after baking versus 3 days later (tests made on independent loaves).
- Average fuel efficiency for 2005 vehicles is 21 miles per gallon. Is average fuel efficiency higher in the new generation "green vehicles?"
- Is blood pressure altered by use of an oral contraceptive? Comparing a group of women not using an oral contraceptive with a group taking it.
- Review insurance records for dollar amount paid after fire damage in houses equipped with a fire extinguisher versus houses without one. Was there a difference in the average dollar amount paid?

Comparing two standard deviations

It is also possible to compare two population standard deviations σ_1 and σ_2 by comparing the standard deviations of two SRSs. However, the procedures are not robust at all against deviations from normality.

When s_1^2 and s_2^2 are sample variances from independent SRSs of sizes *n*1 and *n*2 drawn from normal populations, the *F*-statistic $F = s_1^2 / s_2^2$

has the *F* distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom when H_0: $\sigma_1 = \sigma_2$ is true.

The *F*-value is then compared with critical values from Table D for the *P*-value with a one-sided alternative; this *P*-value is doubled for a two-sided alternative.