STAT 200

Elementary Statistics

9.4 - Comparing Two Independent Means

Printer-friendly version (https://onlinecourses.science.psu.edu/stat200/print/book/export/html/60)

Two independent means are compared using an **independent t test.**

1. Check any necessary assumptions and write null and alternative hypotheses.

There are two assumptions for the following test of comparing two independent means: (1) the two samples are independent and (2) each sample is randomly sampled from a population that is approximately normally distrbuted.

Below are the possible null and alternative hypothesis pairs:

Research Question	**Are the means of group 1 and group 2 different?**	**Is the mean of group 1 greater than the mean of group 2?**	**Is the mean of group 1 less than the mean of group 2?**
Null Hypothesis, H_0	$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \leq 0$	$\mu_1 - \mu_2 \geq 0$
Alternative Hypothesis, H_a	$\mu_1 - \mu_2 \neq 0$	$\mu_1 - \mu_2 > 0$	$\mu_1 - \mu_2 < 0$
Type of Hypothesis Test	Two-tailed, non-directional	Right-tailed, directional	Left-tailed, directional

2. Calculate an appropriate test statistic.

This will be a *t* test statistic. The calculations for these test statistics can get quite involved. Below you are presented with the formulas that are used, however, in real life these calculations are performed using statistical software (e.g., Minitab Express).

Recall that test statistics are typically a fraction with the numerator being the difference observed in the sample and the denominator being the standard error.

The standard error of the difference between two means is different depending on whether or not the standard deviations of the two groups are similar.

Pooled Standard Error Method (Similar Standard Deviations)

If the two standard deviations are similar (neither is more than twice of the other), then the pooled standard error is used:

Pooled standard error

$$s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

s_p = pooled standard deviation

Pooled standard deviation

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

Test statistic for independent means (pooled)

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

Degrees of freedom for independent means (pooled)

$$df=n_1+n_2-2$$

Unpooled Standard Error Method (Differing Standard Deviations)

If the two standard deviations are not similar (one is more than twice of the other), then the unpooled standard error is used:

Unpooled standard error

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

Test statistic for independent means (unpooled)

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

The degrees of freedom are found using a complicated approximation formula. You won't have to do that calculation "by hand" because Minitab Express will compute it for you, but is done by:

Degrees of freedom for independent means (unpooled)

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

NOTE: If one is performing hand calculations using the unpooled method, then choice of degrees of freedom can be made by choosing the smaller of $n_1 - 1$ and $n_2 - 1$.

3. Determine a *p* value associated with the test statistic.

The *t* test statistic found in Step 2 is used to determine the *p* value.

4. Decide between the null and alternative hypotheses.

If $p \leq \alpha$ reject the null hypothesis. If $p > \alpha$ fail to reject the null hypothesis.

5. State a "real world" conclusion.

Based on your decision in Step 4, write a conclusion in terms of the original research question.

Example: Cholesterol (Unpooled)

Cholesterol levels are measured for 28 heart attack patients (2 days after their attacks) and 30 other hospital patients who did not have a heart attack. The response is quantitative so we compare means. It is thought that cholesterol levels will be higher for the heart attack patients, so a one-sided alternative hypothesis is used.

For the 28 heart attack patients, the mean cholesterol level was 253.9 with a standard deviation of 47.7. For the 30 other hopital patients who did not have a heart attach, the mean cholesterol level was 193.1 with a standard deviation of 22.3.

Step 1: The two groups are indpendent of one another. We can assume that the populations are approximately normally distributed. The standard deviations of the two samples are not similar; the standard deviation for the attach group is more than twice of that for the control group. Therefore, we should use unpooled methods when computing the standard error in Step 2.

Let's call the heart attack patients group 1 and the control patients group 2:

$$H_0 : \mu_1 - \mu_2 \leq 0$$
$$H_a : \mu_1 - \mu_2 > 0$$

Method

μ_1: mean of Sample 1
μ_2: mean of Sample 2
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Sample 1	28	253.900	47.700	9.014
Sample 2	30	193.100	22.300	4.071

Estimation for Difference

Difference	95% Lower Bound for Difference
60.800	44.113

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 > 0$

T-Value	DF	P-Value
6.15	37	<0.0001

Minitab Express output that can be used for Steps 2-3

Step 2:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{253.9 - 193.1}{\sqrt{\frac{47.7^2}{28} + \frac{22.3^2}{30}}} = 6.15$$

$$df = \frac{(\frac{47.7^2}{28}+\frac{22.3^2}{30})^2}{\frac{1}{28-1}\left(\frac{47.7^2}{28}\right)^2+\frac{1}{30-1}\left(\frac{22.3^2}{30}\right)^2} = \frac{(81.260+16.576)^2}{244.565+9.457} = \frac{9572.018}{254.040} = 37.679$$

Or, using the Minitab output, the test statistic is given in last line of output as $t = 6.15$, degrees of freedom given as 37.

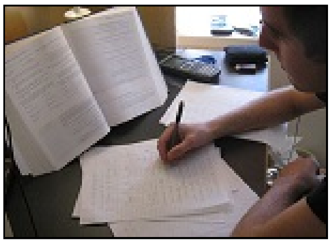
Step 3: p value is given as < 0.0001.

Since we are interested in a one-sided test (>), the p-value can be found by the area to the right of 6.15 in a t-distribution with $df = 37$. We could use t-table (https://onlinecourses.science.psu.edu/stat200/sites/onlinecourses.science.psu.edu.stat200/files/T-table.pdf) to find this p value range.

Step 4: $p \leq .05$ there we reject the null hypothesis in favor of the alternative hypothesis.

Step 5: We decide that the mean cholesterol is higher for those who have recently had a heart attack.

Example: Studying (Pooled)



Hours spent studying per week were reported by students in a class survey. Students who say they usually sit in the front were compared to students who say they usually sit in the back.

For the 99 students who reported that they usually sit in the front, the mean was 16.4 hours with a standard deviation of 10.85 hours. For the 94 students who reported that they usually sit in the back, the mean was 10.9 hours with a standard deviation of 8.41 hours.

Step 1: The two groups are indpendent of one another. We can assume that the populations are approximately normally distributed. The standard deviations of the two samples are similar.

Let's say that students who usually sit in the front are group 1 and students who usually sit in the back are group 2.

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_a : \mu_1 - \mu_2 \neq 0$$

Method

μ_1: mean of Sample 1
μ_2: mean of Sample 2
Difference: $\mu_1 - \mu_2$

Equal variances are assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Sample 1	99	16.400	10.850	1.090
Sample 2	94	10.9000	8.4100	0.8674

Estimation for Difference

Difference	Pooled StDev	95% CI for Difference
5.50	9.74	(2.73, 8.27)

Test

Null hypothesis	H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis	H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
3.92	191	0.0001

Minitab Express output can be used for Steps 2-3

Step 2:

$$s_p = \sqrt{\frac{(99-1)10.85^2 + (94-1)8.41^2}{99+94-2}} = 9.739$$

$$t = \frac{(16.4-10.9)-0}{9.739\sqrt{\frac{1}{99}+\frac{1}{94}}} = \frac{5.5}{1.402} = 3.923$$

Or, using Minitab the test statistic is given in last line of output as $t = 3.923$, degrees of freedom given as 191. The degrees of freedom are found by $n_1 + n_2 - 2$.

Again, we would have to first calculate these sample standard deviations so we would know whether to select in Minitab the "Assume Equal Variances".

Step 3: p value is given as 0.0001.

Since we were interested in the two-tailed test the p value is the area to the right of +3.923

and the area to left of -3.923 in a t-distribution with df = 191. Again we could use t-table (https://onlinecourses.science.psu.edu/stat200/sites/onlinecourses.science.psu.edu.stat200/files/T-table.pdf) and we would find that $p < .001$

Step 4: $p \leq .05$, therefore we reject the null hypothesis in favor of the alternative hypothesis.

Step 5: We decide that the mean time spent studying is different for the two populations.

‹ 9.3 - Comparing Two Independent Proportions (/stat200/node/61) up (/stat200/node/57) 9.5 - Comparing Paired Means › (/stat200/node/62)