1 Inference for the difference between two population means $\mu_1 - \mu_2$

Consider a study for comparing the effect of two drugs on blood pressure. If μ_1 and μ_2 are the mean changes in the blood pressure for drug one and two, respectively, then here one might want to compare the two means based on sample data from the two populations including people being treated with one or the other drug. This would call for a method using the information in the two samples and estimating and testing properties for the two means.

When comparing two population means we need to distinguish the following two situations describing the type of samples we will base the comparison on.

1. independent samples – example: Comparing two drugs based on data on two different groups of individuals, each group receiving one of the drugs.

   The two samples under investigation are unrelated they are independent. (one measurement per individual)

2. paired/matched samples – example: Comparing two drugs based on data where one group of individuals first receives drug A and after the effect has worn of receives drug B. (two measurements for the same individual)

   In this case the samples on the effect of drug A and drug B are related through the individual. For every measurement in one sample you have a corresponding measurement in the second sample.

1.1 Paired samples

In order to control extraneous factors in some studies you can use paired samples. In this case for every individual in the sample from population 1 you find a matching individual from population 2. And the decision is made based on the resulting sample data. In this case we always get the same sample sizes for the two samples.

Example:

- Compare the resting pulse and pulse after exercise.

  To control for all other influences, you take both measurements for every individual observed resulting in two paired samples (before and after exercise).

We are interested in the difference in the population means $\mu_d = \mu_1 - \mu_2$.

For statistical inference the differences of the paired observations

$$x_d = \text{sample 1 value} - \text{sample 2 value}$$

are used.
Which then will create **one** sample of size n of measurements of pairwise differences.
$\bar{x}_d$ and s_d denote the mean and the standard deviation, respectively, for those differences.

For the distribution of $\bar{x}_d$

1. $\mu_{\bar{x}_d} = \mu 1 - \mu_2$, $\bar{x}_d$ is an unbiased estimator for $\mu_1 - \mu_2$
2. $\sigma_{\bar{x}_d} = \sigma/\sqrt{n}$, where σ is the population standard deviations of the pairwise differences.
3. If n is large than $\bar{x}_d$ is normally distributed.

So we get for the t-score

$$t = \frac{\bar{x}_d - (\mu_1 - \mu_2)}{s_d/\sqrt{n}}$$

is t-distributed with $df = n - 1$, if n is large or the pairwise differences come from a normal distribution.

These facts lead to the following

Paired t-Confidence Interval for $\mu_1 - \mu_2$

Assumption: n is large or the population distribution of differences is approximately normal.

The C% Confidence Interval for $\mu_d = \mu_1 - \mu_2$:

$$\bar{x}_d \pm t^*_{n-1}\frac{s_d}{\sqrt{n}}$$

and t^*_{n-1} is the C critical value of the t-distribution with $n-1$ degrees of freedom (Table D).

Reminder: For a given C, t^*_{n-1} is chosen such that $P(-t^*_{n-1} < t - score < t^*_{n-1}) = C$.

Paired t-Test for Comparing Two Population Means

1. **Hypotheses:**

Test type	
Upper tail	$H_0 : \mu_d \leq d_0 \Leftrightarrow \mu_1 - \mu_2 \leq d_0$ versus $H_a : \mu_d > d_0 \Leftrightarrow \mu_1 - \mu_2 > d_0$
Lower tail	$H_0 : \mu_d \geq d_0 \Leftrightarrow \mu_1 - \mu_2 \geq d_0$ versus $H_a : \mu_d < d_0 \Leftrightarrow \mu_1 - \mu_2 < d_0$
Two tail	$H_0 : \mu_d = d_0 \Leftrightarrow \mu_1 - \mu_2 = d_0$ versus $H_a : \mu_d \neq d_0 \Leftrightarrow \mu_1 - \mu_2 \neq d_0$

Assumption: Random sample of differences, and n is large or the population distribution of the differences is approximately normal.

Test statistic:

$$t_0 = \frac{\bar{x}_d - d_0}{s_d/\sqrt{n}}$$

with $n-1$ df.

2. **P-value:**

Test type	**P-value**
Upper tail	$P(t > t_0)$
Lower tail	$P(t < t_0)$
Two tail	$2 \cdot P(t > abs(t_0))$

3. **Decision:**

   If P-value$\leq \alpha$ reject H_0,
   If P-value$> \alpha$ do not reject H_0.

4. **Context**

Example:
The effect of exercise on the amount of lactic acid in the blood was examined.
Blood lactate levels were measured in eight males before and after playing three games of racquetball.

Player	**Before**	**After**	**Difference**
1	13	18	-5
2	20	37	-17
3	17	40	-23
4	13	35	-22
5	13	30	-17
6	16	20	-4
7	15	33	-18
8	16	19	-3

This data results in $\bar{x}_d = -13.63$, $s_d = 8.28$, $n = 8$
Lets test if the decrease in mean lactate levels is significant at a significance level of 0.05. That is

1. $$H_0 : \mu_b - \mu_a \geq 0 \;\; vs. \;\; H_a : \mu_b - \mu_a < 0$$

   .

   where μ_b (μ_a) is the mean lactate level before (after) three games of racquetball.

   $\alpha = 0.05$.

2. **Assumption:** The sample is a random sample and it is appropriate to assume that the difference in lactate level is normal distributed.

3. **Test statistic:** with $d_0 = 0$

   $$t = \frac{\bar{x}_d - d_0}{s_d/\sqrt{n}} = \frac{-13.63}{8.28/\sqrt{8}} = -4.65597$$

   with 7 df.

4. **P-value:** Since we perform a lower tail test the P-value $= P(t < t_0) = P(t > abs(t_0))$.

   Use table D from the text book. Focus on row with $df = 7$, observe that $abs(t_0) = 4.65$ falls between 4.029 and 4.785. Therefore 0.001<P-value<0.0025.

5. **Decision:** Since P-value$< 0.0025 < \alpha = 0.05$, we reject H_0 and accept H_a.

6. **Result:** The data provide sufficient evidence that the mean lactate level after three games of racquet ball is significantly lower than before at a significance level of 0.05.

Lets give an estimate (95% Confidence interval) for the reduction in the mean lactate level through three games of racquetball in males.

$$\bar{x}_d \pm t^*_{n-1}\frac{s_d}{\sqrt{n}} = -13.63 \pm 2.365\frac{8.28}{\sqrt{8}} = -13.63 \pm 6.938$$

or $(-20.568; -6.696)$. $t^*_{n-1} = 2.365$.
Based on the sample data, we can be 95 % confident that the mean decrease in lactate level is between 6.692 and 20.568 after three racquetball games.

Example: A company wanted to study the effect of using a large computer screen (42-inch) versus a small sized (15-inch) screen.
They asked volunteers to perform a certain task on both screens and measured the time it took the volunteers to complete the task for each screen size (resulting in paired observations).
Assume they only asked five volunteers (this is unreasonable, but we want a small example for class). The data is reported below

Volunteer	Small Screen	Large Screen	Difference=Small-Large
1	122	111	11
2	131	116	15
3	127	113	14
4	123	119	4
5	132	121	11

These data results in $\bar{x}_d = 11$, $s_d = 4.301$, $n = 5$
Is at significance level of 5% the mean time to complete the task significantly shorter for the large screen?
If this is correct $\mu_d = \mu_s - \mu_l$ would be greater than zero.

1. $$H_0 : \mu_d \leq 0 \quad vs. \quad H_a : \mu_d > 0$$

   .

   $\alpha = 0.05$.

2. **Assumption:** The sample is a random sample and it is appropriate to assume that the difference in lactate level is normal distributed.

3. **Test statistic:** with $d_0 = 0$

   $$t = \frac{\bar{x}_d - d_0}{s_d/\sqrt{n}} = \frac{11}{4.301/\sqrt{5}} = 5.72$$

   with 4 df.

4. **P-value:** Since we perform an upper tail test the P-value $= P(t > t_0)$.

   Use table D from the text book. Focus on row with $df = 4$, observe that 5.72 falls between 5.598 and 7.173. Therefore 0.001< P-value<0.0025.

5. **Decision:** Since P-value$< 0.0025 < \alpha = 0.05$ reject H_0 and accept H_a.

6. **Result:** At significance level of 5% the data provide sufficient evidence that the mean time to complete the task is smaller for the large than for the small screen.

Lets give an estimate (95% Confidence interval) for the in the mean time to complete the task:

$$\bar{x}_d \pm t^*_{n-1}\frac{s_d}{\sqrt{n}} = 11 \pm 2.776\frac{4.301}{\sqrt{5}} = 11 \pm 5.34$$

or $[5.66, 16.34]$. $t^*_4 = 2.776$.
Based on the sample data we can be 95 % confident that the mean difference in time to complete the task using the different screens falls between 5.7 and 16.3 seconds.
Since zero does fall below the confidence interval, we can be 95% confident that the mean difference in time to complete the task using the different screens exceeds zero, and therefore conclude that on average it takes longer to complete the task on the smaller screen.

1.2 Independent Random Samples

In this section we will see how to perform statistical tests for the difference of the means $\mu_1 - \mu_2$ and how to calculate confidence interval for the difference based **on two independent samples**.

The point estimate for $\mu_1 - \mu_2$ that comes first to mind is the difference of the sample means $\bar{x}_1 - \bar{x}_2$.
In order to do inferential statistics using this difference we have to investigate the distribution of this statistic.

Sample Distribution of $\bar{x}_1 - \bar{x}_2$ from two independent samples.

- For the mean: $\mu_{\bar{x}_1-\bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$, so that $\bar{x}_1 - \bar{x}_2$ is an unbiased estimator for $\mu_1 - \mu_2$.

- For the variance:
$$\sigma^2_{\bar{x}_1-\bar{x}_2} = \sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

- For the standard deviation:
$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- If n_1 and n_2 are both large or both populations are normal distributed then the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is (approximately) normal.

The t-statistic

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

Mathematical results tell us that t is approximately t-distributed with (approximate) degrees of freedom of

$$\text{df} = min(n_1 - 1, n_2 - 1)$$

$min(n_1 - 1, n_2 - 1)$ is the smaller of the two numbers $n_1 - 1$ and $n_2 - 1$.
This is all the information we need to put together a

Two–sample t-Confidence Interval for Comparing Two Population Means

Assumption: Independent random samples and n_1 and n_2 are large or both populations are approximately normal distributed.

The Cx100% Confidence Interval for $\mu_1 - \mu_2$:

$$(\bar{x}_1 - \bar{x}_2) \pm t^*_{df}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

with

$$\text{df} = min(n_1 - 1, n_2 - 1)$$

and t^*_{df} is the critical value of the t-distribution with the given number of degrees of freedom (Table D).

Two–sample t-Test for Comparing Two Population Means

1. **Hypotheses**

Test type	
Upper tail	$H_0 : \mu_1 - \mu_2 \leq d_0$ versus $H_a : \mu_1 - \mu_2 > d_0$
Lower tail	$H_0 : \mu_1 - \mu_2 \geq d_0$ versus $H_a : \mu_1 - \mu_2 < d_0$
Two tail	$H_0 : \mu_1 - \mu_2 = d_0$ versus $H_a : \mu_1 - \mu_2 \neq d_0$

2. **Assumption:** Independent random samples, n_1 and n_2 are large or both populations are approximately normal distributed.

3. **Test statistic:**

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

with

$$\text{df} = min(n_1 - 1, n_2 - 1).$$

4. **P-value:**

Test type	**P-value**
Upper tail	$P(t > t_0)$
Lower tail	$P(t < t_0)$
Two tail	$2 \cdot P(t > abs(t_0))$

5. **Decision:**

   if P-value$\leq \alpha$ reject H_0,
   if P-value$> \alpha$ do not reject H_0.

6. **Context**

Example :
A company wants to show, that a vitamin supplement decreases the recovery time from the common cold. They selected randomly 70 adults with a cold. 35 of those are randomly selected to receive the vitamin supplements (treatment group) the remaining 35 patients receive a placebo pills (control group). The data on the recovery time for both samples is shown below.

population	1 no vitamin	2 vitamin
sample size	35	35
sample mean	6.9	5.8
sample standard deviation	2.9	1.2

Now test the claim of the company: $H_0 : \mu_1 - \mu_2 \leq 0$ versus $H_a : \mu_1 - \mu_2 > 0$ at a significance level of α=0.05.

Assumption: Independent random samples and the sample sizes are sufficiently large.

Test statistic: with $d_0 = 0$

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{6.9 - 5.8}{\sqrt{\frac{2.9^2}{35} + \frac{1.2^2}{35}}} = \frac{1.1}{0.53} = 2.07$$

and

$$\text{df} = 35 - 1$$

P-value approach: We found $t_0 = 2.07$ for 34 df, use df=30 in table D, observe that t_0 falls between $t_{0.025} = 2.042$ and $t_{0.02} = 2.147$, so the p-value falls between 0.01 and 0.02.

$$p - value \leq 0.02 < 0.05$$

therefore the p-value is less than $\alpha = 0.05$. The test is significant at significance level 0.05, we can reject H_0.

Decision: At significance level of 5% the data provide sufficient evidence that vitamins decrease the mean recovery time for common colds.

The 95% Confidence Interval for the difference in the mean recovery time for treatment and control group $\mu_1 - \mu_2$.
The degrees of freedom are 34:

$$\bar{x}_1 - \bar{x}_2 \pm t^*_{df}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(6.9 - 5.8) \pm 2.042 \cdot 0.53$$

$$1.1 \pm 1.082$$

or (0.018 ; 2.182). We are 95% confident that the difference in mean recovery time falls between 0.018 and 2.882 days (between less than half an hour and two days and 4 hours).
Also the 95% confidence interval lies entirely above 0. So that 0 is with a confidence of 95% less than $\mu_1 - \mu_2$. We can state with confidence 0.95 that the recovery time without vitamin treatment takes longer than with vitamin treatment.