Inferences about two population means: independent samples.

Usually, we will be testing the null hypothesis

$$H_0 : \mu_1 = \mu_2,$$

where μ_1 and μ_2 are the means of populations 1 and 2, respectively.

Requirements:

(i) Two ***independent*** simple random samples.

(ii) The samples come from populations with (approximately) normal distributions ***or*** both sample sizes are 'large': $n_1 > 30$ and $n_2 > 30$. (Or both).

Comment: This test is *'robust against departures from normality'*, i.e., the results are reliable as long as the data are not extremely far from normal and there are no outliers.

Notation:

- $\overline{x}_1$, s_1 and n_1 are the sample mean, sample standard deviation and sample size of the first sample
- $\overline{x}_2$, s_2 and n_2 are the sample mean, sample standard deviation and sample size of the second sample.
- $m =$ the ***smaller of*** $(n_1 - 1)$ and $(n_2 - 1)$.

  We will use m as a conservative estimate for the number of d.f.

Procedure:

1. Formulate Null and Alternative hypotheses:

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 < \mu_2 \quad \text{(left tail)} \quad \text{OR}$$

$$H_a : \mu_1 > \mu_2 \quad \text{(right tail)} \quad \text{OR}$$

$$H_a : \mu_1 \neq \mu_2 \quad \text{(two tails)}$$

2. Choose significance level $\alpha = P(\text{type I error})$

3. Calculate the (observed) test statistic:

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

(*) **Distribution of test statistic:** t-distribution with m degrees of freedom. (There are fancier estimates for the number of d.f., but this estimate is conservative and easy to remember.)

4. Decision Criterion.

(a) *Traditional method:* Reject H_0 if ...

$$H_a:\ \mu_1 < \mu_2 \implies t^* < -t_\alpha$$
$$H_a:\ \mu_1 > \mu_2 \implies t^* > t_\alpha$$
$$H_a:\ \mu_1 \neq \mu_2 \implies |t^*| > t_{\alpha/2}$$

(b) *P-value method:* Reject H_0 if ...

$$H_a:\ \mu_1 < \mu_2 \implies P = P(t < t^*) < \alpha$$
$$H_a:\ \mu_1 > \mu_2 \implies P = P(t > t^*) < \alpha$$
$$H_a:\ \mu_1 \neq \mu_2 \implies P = P(|t| > |t^*|) < \alpha$$

Comment: Statistical software computes P-values for t-distributions. If you are relying on a t-table for your hypothesis tests, use the traditional (critical value) method.

5. Summarize conclusions.

Example. The effects of alcohol on motor and visual skills. Ethanol was given to the members of the treatment group (sample 1) and a placebo was given to the members of the control group (sample 2), both groups were given motor and visual tests and errors were recorded. Use the sample data below to test the hypothesis that alcohol consumption results in more errors.

(*) Sample 1: $\overline{x}_1 = 4.20$, $s_1 = 2.20$ and $n_1 = 22$.

(*) Sample 2: $\overline{x}_2 = 1.71$, $s_2 = 0.72$ and $n_2 = 22$.

1. Hypotheses: $H_0 : \mu_1 = \mu_2$ and $H_a : \mu_1 > \mu_2$.

The alternative hypothesis is that the mean number of errors for inebriated people (μ_1) is larger than the mean number of errors for sober people (μ_2). The null hypothesis says that alcohol has no effect and the means are equal.

2. Significance level: $\alpha = 0.05$

3. Test statistic:

$$t^* = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{4.2 - 1.71}{\sqrt{\frac{2.2^2}{22} + \frac{0.72^2}{22}}} \approx 5.045.$$

$m = 22 - 1 = 21$ d.f.

4. Rejection criterion: this is a right-tailed test ($H_a : \mu_1 > \mu_2$) and the 1-tail critical value for $\alpha = 0.05$ and 21 d.f. is $t_{0.05} = 1.721$. The rejection criterion is $t^* > t_{0.025}$, which is satisfied in this case, so we reject H_0.

5. Conclusion: There is sufficient sample evidence to support the claim that alcohol consumption increases the mean number of errors in visuomotor skill tests (like driving a car).

Confidence interval for the difference of two population means (independent samples).

(*) Point estimate for $\mu_1 - \mu_2$: $\overline{x}_1 - \overline{x}_2$.

(*) Margin of error: $E_\alpha = t_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

We find the critical t value using $m = \min(n_1 - 1, n_2 - 1)$ d.f.

(*) Confidence interval for $\mu_1 - \mu_2$:

$$(\overline{x}_1 - \overline{x}_2) \pm E_\alpha.$$

Example. 99%-Confidence interval for the difference in mean number of errors for inebriated and sober people (using the same data as before):

(*) $E_{0.01} = t_{0.005} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \overbrace{2.831}^{21\ d.f.} \cdot \sqrt{\frac{2.2^2}{22} + \frac{0.72^2}{22}} \approx 1.397.$

(*) Confidence interval for difference in means:

$$(4.20 - 1.71) \pm 1.397 = 2.49 \pm 1.397 = \ (1.093, 3.887)$$

Inferences about two population means: matched pairs.

Requirements:

(i) The sample data consist of ***matched pairs***.

This means that we have two sets of sample data, where each observation in the first sample is paired with a matching observation from the second sample. Then 'matching' reflects a dependence between the two observations. For example:

- The weights of individuals before (sample 1) and after (sample 2) participating in a weight-loss program.
- The number of years of education of husbands (sample 1) and their wives (sample 2).

(ii) The sample is a simple random sample (of matched pairs).

(iii) The sample size is 'large', i.e., $n > 30$ matched pairs ***or*** the *differences between the pairs of values* come from a population with an approximately normal distribution (or both).

Notation:

- d = the difference between the two values in a single matched pair.
- μ_d = the mean value of the differences d for the population of all matched pairs.
- $\overline{d}$ = the mean value of the differences d for a sample of matched pairs.
- s_d = the standard deviation of the differences d for the sample of matched pairs.
- n = sample size.

Hypothesis tests for matched pairs.

Procedure:

1. Formulate Null and Alternative hypotheses:

$$H_0 : \mu_d = \mu_0 \quad \text{(usually } \mu_0 = 0\text{)}$$
$$H_a : \mu_d < \mu_0 \quad \text{(left tail)} \quad \text{OR}$$
$$H_a : \mu_d > \mu_0 \quad \text{(right tail)} \quad \text{OR}$$
$$H_a : \mu_d \neq \mu_0 \quad \text{(two tails)}$$

2. Choose significance level $\alpha = P(\text{type I error})$

3. Calculate the (observed) test statistic:

$$t^* = \frac{\overline{d} - \mu_0}{\frac{s_d}{\sqrt{n}}}$$

(*) **Distribution of test statistic:** t-distribution with $n-1$ degrees of freedom.

4. Decision Criterion.

(a) *Traditional method:* Reject H_0 if ...

$$H_a:\ \mu_d < \mu_0 \implies t^* < -t_\alpha$$
$$H_a:\ \mu_d > \mu_0 \implies t^* > t_\alpha$$
$$H_a:\ \mu_d \neq \mu_0 \implies |t^*| > t_{\alpha/2}$$

(b) *P-value method:* Reject H_0 if ...

$$H_a:\ \mu_d < \mu_0 \implies P = P(t < t^*) < \alpha$$
$$H_a:\ \mu_d > \mu_0 \implies P = P(t > t^*) < \alpha$$
$$H_a:\ \mu_d \neq \mu_0 \implies P = P(|t| > |t^*|) < \alpha$$

Comment: Statistical software computes P-values for t-distributions. If you are relying on a t-table for your hypothesis tests, use the traditional (critical value) method.

5. Summarize conclusions.

Confidence interval for mean difference between matched pairs:

(*) Point estimate: $\overline{d}$.

(*) Margin of error (for $(1-\alpha)\cdot 100\%$) confidence interval:

$$E_\alpha = t_{\alpha/2} \cdot \frac{s_d}{\sqrt{n}}$$

(*) Confidence interval:

$$\overline{d} \pm E_\alpha.$$

Example. A simple random sample of 12 girls (12-16) were surveyed and the difference between their self-reported heights and their measured heights were recorded.

Summary statistics:

(*) $n = 12$ (11 d.f.)

(*) $\overline{d} = -0.475$ inches

(*) $s_d = 1.981$ inches

95%-confidence interval for the mean of (rep. height)−(meas. height):

The margin of error is $E_{0.05} = t_{0.025} \cdot \frac{s_d}{\sqrt{n}} = \overbrace{2.201}^{11\ d.f.} \cdot \frac{1.981}{\sqrt{12}} \approx 1.259$, so a 95%-confidence interval for the mean difference is

$$-0.475 \pm 1.259 = (-1.734, 0.784).$$