emmading.com
info@emmading.com

Two-sample t-tests

Lesson Structure

📌 **Interview Questions**

- Calculate the mean difference and the standard error of mean difference for a two-sample t-test
- Implement two-sample t-tests in Python/R.

Independent Samples

💡 X_1 and X_2 are random samples from two independent populations.

- Each population is (approximately) normally distributed.
- The samples within each population are independent of each other.

▼ **Independent/unpaired samples**

There is no relationship between the subjects in each sample.

- Subjects in the first group cannot also be in the second group
- No subject in either group can influence subjects in the other group
- No group can influence the other group

▼ **Hypotheses**

$H_0: \mu_1 = \mu_2,$

$H_1: \mu_1 \neq \mu_2.$

▼ **Steps to conduct an independent t-test**

- Calculate difference among sample means
- Sum the variances of the samples and calculate the equal or unequal variances

  Two scenarios:

  - Equal variances
  - Unequal variances

- Calculate t-statistic and compare it with t-critical value
- If |t-statistic| > t-critical value, reject H_0.

▼ Two-Sample T-test with Equal Variances

When the two groups' variances are believed to be equal, we use a t-test based on a pooled variance estimate.

▼ t-statistic

$$T = \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2,\alpha/2}$$

Where $\bar{X}_i$ is the sample mean of sample i, for $i \in \{1, 2\}$, and the pooled sample variance s_p^2 is defined by $s_p^2 \ = \frac{SS_1+SS_2}{n_1+n_2-2}$ with $SS_i = \sum_{j=1}^{n_i} \left(X_j - \bar{X}_i\right)^2$.

▼ Confidence interval

A level $(1-\alpha) * 100\%$ confidence interval for the difference between two means:

$$\bar{X}_1 - \bar{X}_2 \pm t_{n_1+n_2-2,\alpha/2} * s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

▼ Welch's T-test with Unequal Variances

If the two variances are not similar (one is more than twice of the other) but the other assumptions for the 2-sample t-test hold, then we can use Welch's t-test, which employs an unpooled standard error.

▼ t-statistic

Unpooled standard error: $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{df,\alpha/2}$$

where $s_1 = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{1i}-\bar{X}_1\right)^2}{n_1-1}}$, $s_2 = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{2i}-\bar{X}_2\right)^2}{n_2-1}}$

and

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

▼ **E.g. Compare the mean heights of men in two cities.**

We sample $n_1 = 19$ and $n_2 = 23$ men respectively. We simulate some data and run the test in Python as follows. We assume that the two populations have different variances.

$H_0 : \mu_1 = \mu_2,$

$H_1 : \mu_1 \neq \mu_2.$

```python
# simulate n height sample points
n1 = 19
sigma1 = 1.97
mu_1 = 172
x1 = stats.norm.rvs(loc=mu_1, scale=sigma1, size=n1, random_state=1)

n2 = 23
sigma2 = 2.2
mu_2 = 169
x2 = stats.norm.rvs(loc=mu_2, scale=sigma2, size=n2, random_state=1)

# Test if mean height difference is close to mu_0 using a t-test with 95% CL
mu_0 = 0
unpooled_stdev = (x1.var() / n1 + x2.var() / n2)**0.5
observed_t_score = (x1.mean() - x2.mean() - mu_0) / unpooled_stdev

df = (x1.var()/n1+x2.var()/n2)**2 / ((x1.var()/n1)**2/(n1-1)+(x2.var()/n2)**2/(n2-1))
critical_t_score = stats.t.ppf(0.975, df)
print('observed_t_score = ', observed_t_score)
print('critical_t_score = ', critical_t_score)
```

```
observed_t_score =  3.9666234546425345
critical_t_score =  2.021720124238159
```

We are in the rejection region so there is a statistically significant difference between the height of men in the two cities.

▼ **Confidence interval**

A level $(1-\alpha) * 100\%$ confidence interval for the difference between two means:

$$\bar{X}_1 - \bar{X}_2 \pm t_{df,\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where $s_1 = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{1i}-\bar{X}_1\right)^2}{n_1-1}}$, $s_2 = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{2i}-\bar{X}_2\right)^2}{n_2-1}}$ and $df = \frac{(s_1^2/n_1+s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1-1)+(s_2^2/n_2)^2/(n_2-1)}$.

▼ **E.g. Confidence interval for the difference of heights of men.**

```
# Build CI at 95% CL for the difference in population mean heights
margin_of_error = critical_t_score * unpooled_stdev
lower = x1.mean()-x2.mean() - margin_of_error
upper = x1.mean()-x2.mean() + margin_of_error
confidence_interval = (lower, upper)
print(confidence_interval)
```

```
(1.3864646114468822, 4.2689147185413105)
```

Dependent Samples

Dependent/paired samples: each data point in one sample is uniquely paired to a

data point in the second sample. Examples:

- Change over time (longitudinal study), measure one variable at one point in time and measure the same variable of the same sample at a later point in time.
- Pre-test vs. post-test, to see whether there is a significant effect due to the treatment.

▼ Advantages

- Controls for individual differences

▼ Disadvantages

- Carry-over effects - second treatment can be affected by first treatment
- Order may influence results - which treatment is first can affect the results

▼ Hypotheses

Let $D_1, ..., D_n$ be a small random sample (n <= 30) of the differences in pairs.

$H_0 : \mu_D = \mu_0$,

$H_1 : \mu_D \neq \mu_0$.

Almost always $\mu_0 = 0$.

▼ Steps to conduct dependent t-test

- Calculate mean difference of the pairwise differences
- Calculate standard deviation of the pairwise differences
- Calculate t-statistic and compare it with t-critical value
- If |t-statistic| > t-critical value, reject H_0.

▼ t-statistic

Under H_0

$$T = \frac{\bar{D} - \mu_0}{s_D / \sqrt{n}} \sim t_{n-1}$$

$\bar{D}$ is the average of the differences. μ_0 is the average of the differences under the null hypothesis, and s_D is the standard deviation of sample differences.

▼ E.g. Evaluate the effect of stretching on height.

You measure the height of a group of people who do not stretch regularly. They then start stretching regularly for a year. You measure their heights again and take the difference. You want to see if there is a statistically significant effect, i.e. if the population mean differs from zero.

$H_0 : \mu_1 = \mu_2,$

$H_1 : \mu_1 \neq \mu_2.$

```
# simulate n height sample points
n = 20
sigma = 0.001
population_mean = 0
x = stats.norm.rvs(loc=population_mean, scale=sigma, size=n, random_state=1)

mu_0 = 0
observed_t_score = (x.mean() - mu_0) / (x.std() / n**0.5)
critical_t_score = stats.t.ppf(0.975, n-1)
print('observed_t_score = ', observed_t_score)
print('critical_t_score = ', critical_t_score)
```

```
observed_t_score =  -0.5423002591242909
critical_t_score =  2.093024054408263
```

So we fail to reject the null hypothesis. There is probably no change in height after the additional stretching.

▼ Confidence interval

A level $(1-\alpha) * 100\%$ confidence interval for the mean difference μ_D is given by

$$\bar{D} \pm t_{n-1,\alpha/2} * \frac{s_D}{\sqrt{n}}$$

If the sample size is large, then the confidence interval for the mean difference μ_D is

$$\bar{D} \pm z_{\alpha/2} * \frac{s_D}{\sqrt{n}}$$