Hypothesis testing: comparig means

- Prof. Giuseppe Verlato

Unit of Epidemiology & Medical Statistics – Department of Diagnostics & Public Health – University of Verona

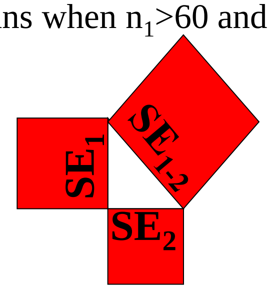
COMPARING MEANS

1) Comparing sample mean to population mean

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \qquad t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

2) Comparison of two sample means when $n_1>60$ and $n_2>60$

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



This equation is useful to understand test development, but it not used in everyday practice, equation 3 is used instead.

3)Comparison between the means of two small sample ($n_1<60$ and $n_2<60$)

$$s_p^2 = s^2_{pooled} = \frac{SSq_1 + SSq_2}{n_1 + n_2 - 2}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{SSq_1+SSq_2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

4) Comparing paired data

$$\left.\begin{array}{l} x_{11} - x_{21} = d_1 \\ x_{12} - x_{22} = d_2 \\ x_{13} - x_{23} = d_3 \\ x_{14} - x_{24} = d_4 \\ x_{15} - x_{25} = d_5 \end{array}\right\} \begin{array}{l} \bar{d} \\ s_d \end{array}$$

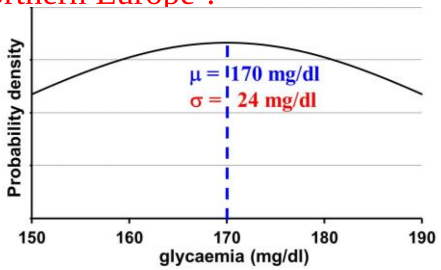
$$t = \frac{\bar{d} - 0}{s_d / \sqrt{n}}$$

Hypothesis testing: comparing a sample mean ($\bar{x}$) to a population mean (μ)

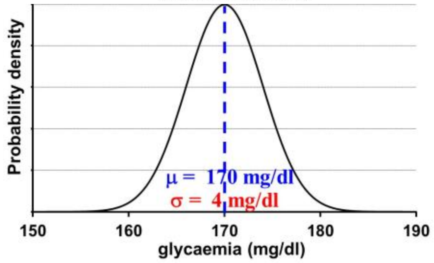
Let's assume that in Northern Europe glycaemia of type 2 diabetic patients is equal to 170 ± 24 mg/dl ($\mu \pm \sigma$).

A mean glycaemia of 161 mg/dl is recorded in 36 Italian patients.

Is glycaemia of type 2 diabetic patients the same in Italy as in Northern Europe ?



Distribution of glycaemia in type 2 diabetic patients from Northern Europe



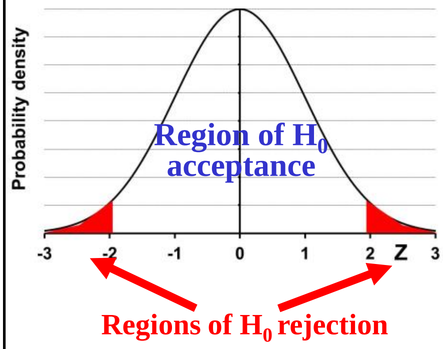
Corresponding distribution of sample means for n=36

($\sigma = 24/\sqrt{36} = 24/6 = 4$ mg/dl)

N.B. Since n>30, the distribution of sample mean is normal irrespective of the distribution of the original variable.

Hypotheses $\begin{cases} H_0: \mu_{it} = \mu_0 \\ H_1: \mu_{it} \neq \mu_0 \end{cases}$

Significance level
[P(type I error)] = 5%



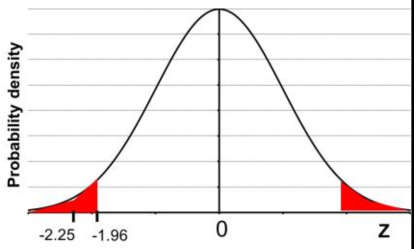
z-test is used, based on the z-distribution (normal standard deviate)

Significance level	10%	**5%**	1%
Critical threshold (absolute value)	1.645	**1.96**	2.576

There are two regions of rejection, hence the test is two-tailed.

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{161-170}{24/\sqrt{36}} = \frac{-9}{4} = -2.25$$

| observed z | > threshold
2.25 1.96



H_0 is rejected.

Glycaemia of Italian diabetic patients differs from glycaemia of North European diabetic patients (*P*=0.024).

Let's assume that in Northern Europe glycaemia in type 2 diabetic patients is normally distributed with mean 170 mg/dl. The variability of glycaemia is unknown.

Glycaemia in a sample of 25 Italian diabetic patients is 155 ± 20 mg/dl (mean ± standard deviation).

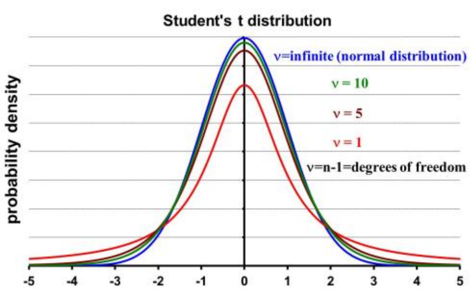
Is glycaemia of type 2 diabetic patients the same in Italy as in Northern Europe ?

s (sample standard deviation) is taken as an estimate for σ (population standard deviation):

Standard Error = $s/\sqrt{n}$ = $20/\sqrt{25}$ = 20/5 = 4

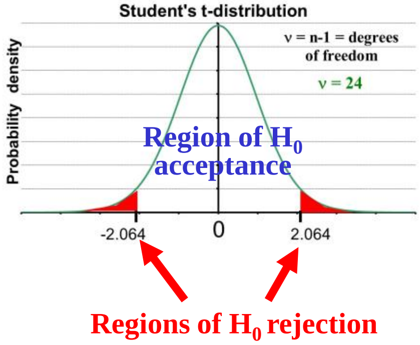
However estimating σ from s introduces an additional source of sample variability: both mean and standard deviation vary from one sample to another.

To cope with this problem, the z-distribution is replaced by the t-distribution, which presents larger variability.



Hypotheses $\begin{cases} H_0: \mu_{it} = \mu_0 \\ H_1: \mu_{it} \neq \mu_0 \end{cases}$

Significance level
[P(type I error)] = 5%

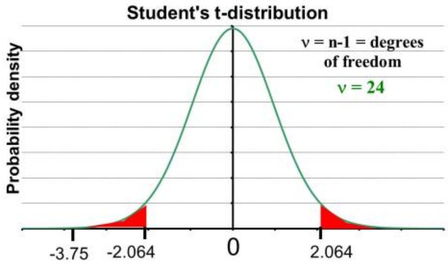


Assuming α=5%, with n-1=24
critical threshold = 2.064

There are two regions of rejection, hence the test is two-tailed.

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{155-170}{20/\sqrt{25}} = \frac{-15}{4} = -3.75$$

| observed t | > threshold
3.75 2.064



H_0 is rejected.
Glycaemia of Italian diabetic patients differs from glycaemia of North European diabetic patients (*P*=0.001).

Hypothesis testing: comparison of two sample means ($\bar{x}_1$ and $\bar{x}_2$)

Example: Twenty hypertensive patients are randomly assigned to two different treatments. The first group is administered a diuretic drug, while the second group is given a beta-blocker. The next table shows the heart rate (in beats/min), recorded in each patient after two weeks of treatment:

Diuretic	80	86	88	82	88	87	77	72	88	79
Beta-blocker	70	65	66	76	68	66	71	72	69	82

Does heart rate significantly differ between the two groups ?

Student's t-test (for unpaired data)

$$\begin{cases} H_0: \mu_{diur} = \mu_{B-} \\ H_1: \mu_{diur} \neq \mu_{B-} \end{cases}$$

two-tailed test

$$\begin{cases} H_0: \mu_{diur} \leq \mu_{B-} \\ H_1: \mu_{diur} > \mu_{B-} \end{cases}$$

one-tailed test

Significance level = 5%

Degrees of freedom = $n_1 + n_2 - 2 = 10+10-2 = 18$

Critical threshold = $t_{18,\ 0.025} = 2.101$

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{SE_{\bar{x}1-\bar{x}2}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{(1/n_1+1/n_2)\,(SSq_1+SSq_2)/\,(n_1+n_2-2)}}$$

	Σx	Σx^2	n	$\bar{x}$	SSq	Var	SD
Diuretic	827	68675	10	82.7	282.1	31.34	5.599
Beta-blocker	705	49947	10	70.5	244.5	27.17	5.212

Assumptions:
1) Heart rate is normally distributed
2) Variance does not differ between the two groups

$$t = \frac{|82.7-70.5|}{\sqrt{(1/10+1/10)\,(282.1+244.5)/\,(10+10-2)}} = \frac{12.2}{\sqrt{5.85}} = \frac{12.2}{2.42} = 5.043$$

observed t (5.043) > critical t value (2.101)

⇓

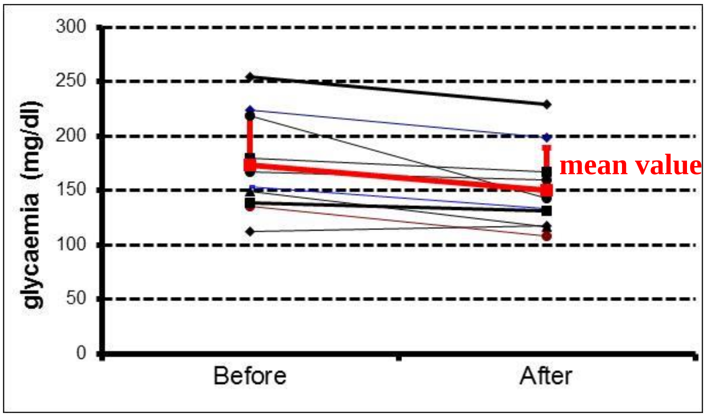
H_0 is rejected ($P<0.001$)

Hypothesis testing:
Student’s t test for paired data

An oral hypoglycaemic drug is given to a sample of diabetic patients. Glycaemic values (in mg/dl), recorded in each patient before and after treatment, are presented in the following table:

Patient	1	2	3	4	5	6	7	8	9	10
Before	224	179	149	254	112	135	218	153	167	138
After	198	167	116	229	117	108	143	133	159	131

Was glycaemia significantly affected by treatment ?



Patient	Before	After	Difference
1	224	198	-26
2	179	167	-12
3	149	116	-33
4	254	229	-25
5	112	117	5
6	135	108	-27
7	218	143	-75
8	153	133	-20
9	167	159	-8
10	138	131	-7

$\Sigma d = -228$

$\Sigma d^2 = 9426$

$\bar{d} = -22{,}8$

$s_d = 21{,}67$

(Student's t-test for paired data)

$$\begin{cases} H_0: \delta = 0 \\ H_1: \delta \neq 0 \end{cases}$$

Two-tailed test

Significance level = 5%

Degrees of freedom = n_1 - 1 = 10-1 = 9

Critical threshold = $t_{9,\ 0.025}$ = 2.262

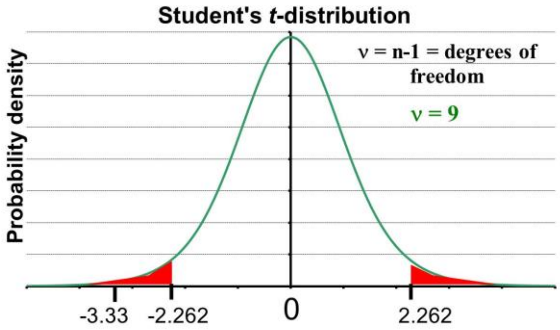
$$t = \frac{\bar{d}-0}{s_d/\sqrt{n}} = \frac{-22.8}{21.67/\sqrt{10}} = \frac{-22.8}{6.853} = -3.327$$

| observed t | > critical t value

3.327 2.262

H_0 is rejected.

Glycaemia significantly decreased in type 2 diabetic patients after administering an oral hypoglycaemic drug (*P*=0.009).



Comparing means: computing sample size

Computing sample size to achieve enough power to detect possible difference between two means

$$n > 2 \left[\frac{(z_\alpha + z_\beta)\,\sigma}{\delta}\right]^2$$

where

n = number of subjects in each group

$z_\alpha = 1.96$ for alpha = 5%

$z_\beta = 0.842, 1.282, 1.645$ for power = 80, 90, 95%

σ = standard deviation, derived from pilot studies or current literature

$\delta = \bar{x}_1 - \bar{x}_2$ = minimal clinically-important difference

Example: The reference drug decreases systolic pressure by 25 mmHg. To represent a real improvement, the new drug should reduce systolic pressure by at least 30 mmHg (i.e. 5 mmHg more). Standard deviation of pressure changes is estimated to be 10 mmHg from previous studies. Alpha is set at 5% and power at 90%, hence $z_{\alpha/2} = 1.96$ e $z_\beta = 1.282$.

$$n > 2 \left[\frac{(z_\alpha + z_\beta)\,\sigma}{\delta}\right]^2$$

$$n > 2 \left[\frac{(1.96 + 1.28)\,10}{5}\right]^2$$

$$n > 84.06 \qquad n \geq 85$$

85 subjects per group are needed to achieve 90% power.