
HYPOTHESIS TESTING II

TESTS ON MEANS

Sorana D. Bolboacă

OBJECTIVES

- Significance value vs p-value
- Parametric vs non-parametric tests
- Tests on means:

SIGNIFICANCE LEVEL VS. p-VALUE



Materials
and Methods

- Significance level (α) = property of a statistical procedure and takes a fixed value.
 - Usually take a value equal to 0.05
- p-value = random variable whose value depends upon the composition of the individual sample



Results

SIGNIFICANCE LEVEL VS. p-VALUE

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4129321/>

n Res

Statistical Analysis

Sample size in our study was calculated by using method described by S. Freestone et al., for sample size estimation for short term trails of antihypertensive drugs [6]. The power of our study was 90% at 0.05% level of significance. Considering the loss to follow up, total 90 patients were enrolled in the study. Qualitative data such as sex, patient's habits were analyzed by using chi-square test. Data on adverse effects was analysed using z-test for difference between two proportions. Quantitative data was analysed using z-test for difference between means. p-value < 0.05 was taken as significant. p-value < 0.001 was taken as highly significant, p-value >0.05 was considered insignificant.

lin Diagn Res

Effect of nebivolol and atenolol on Heart rate

S. No	Parameters	Nebivolol	Atenolol	p-value
1	Before treatment	78.05±5.839	76.55±5.33	p>0.05
2	After treatment	63.53±3.86	59.0±3.271	—
3	Mean reduction in heart rate	14.51±4.69	17.55±5.06	p<0.05
4	p-Value	p<0.001	p<0.001	

PARAMETRIC & NON-PARAMETRIC

	Parametric	Non-Parametric
Assumed distribution	Normal	Any
Assumed variance	Homogenous	Any
Type of data	Ratio or Interval	Ordinal or Nominal
Central measure	Mean	Median
Dispersion measure	Standard deviation	(Q1; Q3)

	Parametric	Non-Parametric
2 independent groups	Independent t-test	Mann-Whitney test
2 dependent groups	Paired t-test	Wilcoxon test
> 2 groups	ANOVA	Kruskal-Wallis test Friedman's ANOVA
Correlation	Pearson	Spearman, Kendall, etc.
...

HYPOTHESIS TESTING FOR A MEAN

- Hypothesis testing via CI
- One and two-sided hypothesis tests

HYPOTHESIS

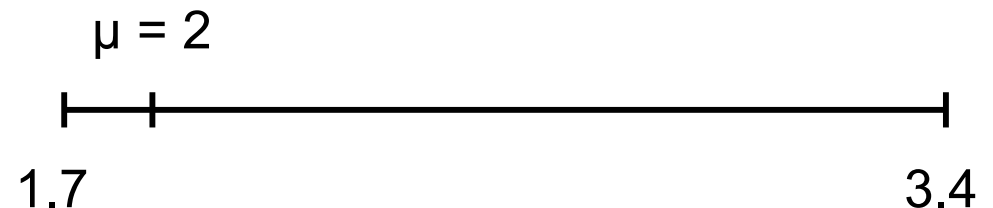
Null	H_0	Not significantly different (= symbol)
Alternative	H_A/H_1	Significantly different <ul style="list-style-type: none">■ two-sided: \neq symbol■ one-sided: $<$ symbol OR $>$ symbol

HYPOTHESIS TESTING VIA CI

- $n=33$, $m = 2.6$, $s = 1.2$, $SE = 0.21$
- 95% CI for the average number exams failed by medical students in the first year of study is $(1.7, 3.4)$. Based on this confidence interval, do these data support the hypothesis that medical students on average failed on average more than 2 exams?

H_0	$\mu = 2$	Medical student in first year of study failed 2 exams, on average
H_A	$\mu > 2$	Medical student in first year of study failed more than 2 exams, on average

Always about population parameter, never about population statistics



HYPOTHESIS TESTING VIA CI

- $n=33, m = 2.6, s = 1.2, SE = 0.21$
- $P(\text{observed or more extreme outcome} | H_0 \text{ true})$
- $P(m > 2.6 | H_0: \mu = 2)$
- Test statistic: $Z = (2.6 - 2) / 0.21 = 2.8571$
- $p\text{-value} = P(Z > 2.8571) = 0.0021$
 - $p\text{-value}$ = the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis was true.
 - If $p\text{-value} < 0.05$ we say that it would be very unlikely to observe the data if the null hypothesis were true, and hence reject H_0 .
 - If the $p\text{-value} > 0.05$ we say that it is likely to observe the data even if the null hypothesis were true, and hence fail to reject H_0 .

HYPOTHESIS TESTING ON A SINGLE MEAN

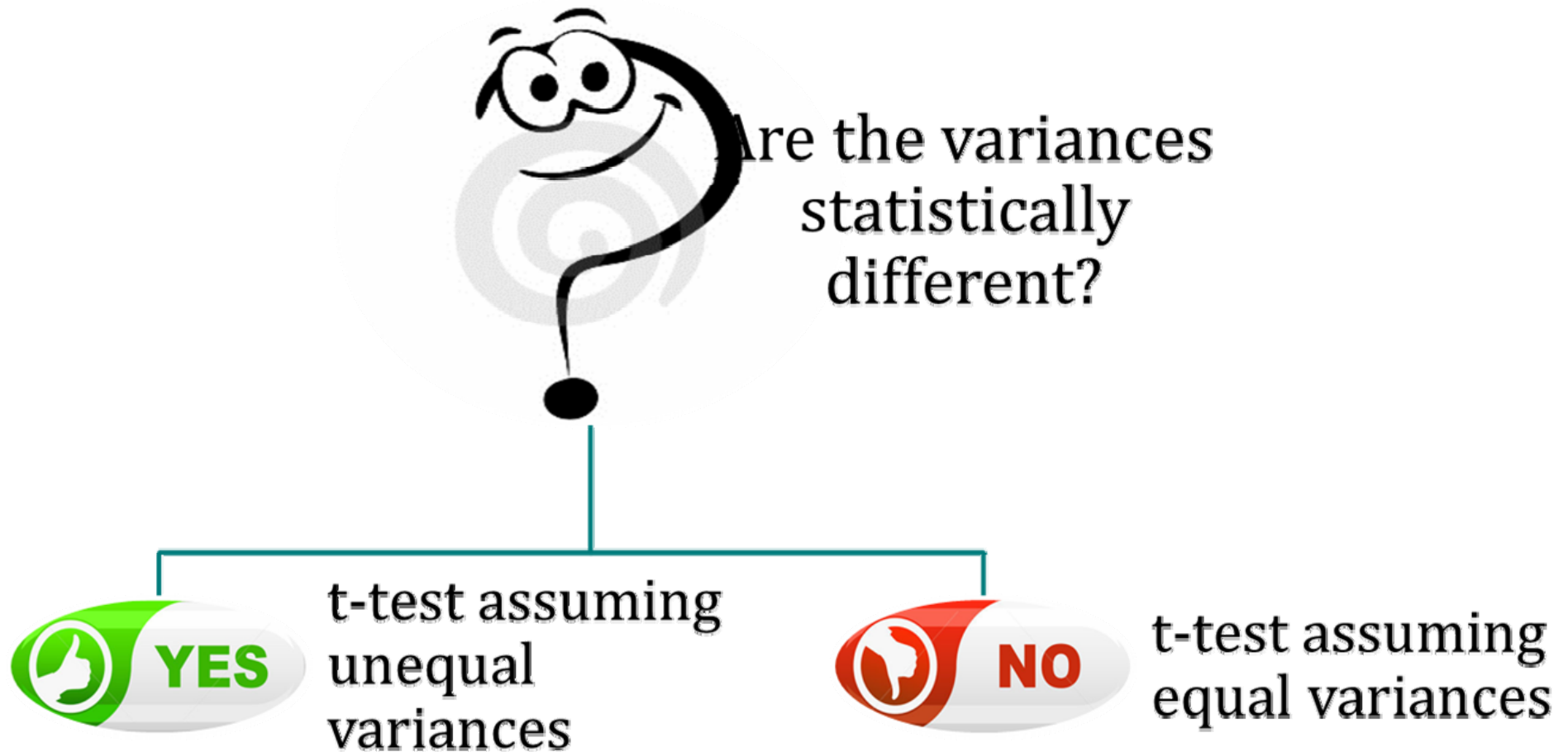
1. Hypotheses: $H_0: \mu = \text{null value}$ & $H_A: \mu \neq \text{null value}$
2. Calculate the point estimator
3. Check conditions:
 - Independence: observations are independent by each others
 - Sample size: $n \geq 30$
4. Draw sampling distribution, shade p-value, calculate test statistic: $Z = (\bar{m} - \mu) / (s / \sqrt{n})$
5. Make a decision:
 - $p\text{-value} < \alpha \rightarrow$ reject H_0 (data provide convincing evidence for H_A)
 - $p\text{-value} > \alpha \rightarrow$ fail to reject H_0 (data do not provide convincing evidence for H_A)

INDEPENDENT SAMPLES: ARE TWO MEANS THE SAME?

Total sample size	Subgroup sample size	Equal Variance	Unequal Variances
Large size ($n > 50$ or $n > 100$) or σ 's known	~ equal very different	Z-test	Rank-sum test
Small size	~ equal very different	t-test	Rank-sum test

Assumptions: ❶ the observations are independent from each other; ❷ the samples are drawn from a normal distribution (use a Rank-test when this assumption is violated); ❸ the standard deviation of samples are not statistically different by each other (apply an unequal variance form of the means test or a rank test).

INDEPENDENT SAMPLES T-TEST



Z AND T TESTS TO COMPARE A SAMPLE MEAN WITH A POPULATION MEAN

Z test

When? Population standard deviation is known OR $n > 50$ (100)

Hypotheses:

■ $m = \mu (H_0)$ vs. $m \neq \mu (H_1)$

Significance level ($\alpha = 0.05$) → critical value with $n-1$ df (degree of freedom)

Test statistics:

$z = (m - \mu) / (\sigma / \sqrt{n})$ where σ = standard deviation, n = sample size

t-test

When? Unknown standard deviation OR $n < 50$ (100)

Hypotheses:

■ $\mu_1 = \mu_2 (H_0)$ vs. $\mu_1 \neq \mu_2 (H_1)$

Significance level ($\alpha = 0.05$) → critical value with $n-1$ df (degree of freedom)

Test statistics:

$t = (m - \mu) / (s / \sqrt{n})$ where s = standard deviation, n = sample size

Z AND T TESTS TO COMPARE A SAMPLE MEAN WITH A POPULATION MEAN - **PROBLEM 1**

Z-test

- $\mu = 0 (H_0)$ vs $\mu \neq 0 (H_1)$
- $\alpha = 0.05$
- $\sigma = 1.75$
- $n = 15$
- $m = 3.87$
- $Z_{\text{crit}} = 1.96$

Z = ?

Conclusion ?

T-test

- $\mu = 0 (H_0)$ vs $\mu \neq 0 (H_1)$
- $\alpha = 0.05$
- $s = 2.50$
- $n = 15$
- $m = 3.87$
- $t_{\text{crit}} = 2.145$

Z = ?

Conclusion ?

Z AND T TESTS TO COMPARE TWO SAMPLE MEANS

Z test

When? Population standard deviation is known OR $n > 50$ (100)

Hypotheses:

■ $\mu_1 = \mu_2$ (H_0) vs. $\mu_1 \neq \mu_2$ (H_1)

Significance level ($\alpha = 0.05$)
→ critical value with $n_1 + n_2 - 1$ df (degree of freedom)

Test statistics:

$z = (m_1 - m_2) / \sigma_d$, where $\sigma_d =$ population standard error ($\sigma_d = \sigma \cdot \sqrt{1/n_1 + 1/n_2}$)

t-test

When? Unknown standard deviation OR $n < 50$ (100)

Hypotheses:

■ $\mu_1 = \mu_2$ (H_0) vs. $\mu_1 \neq \mu_2$ (H_1)

Significance level ($\alpha = 0.05$)
→ critical value with $n_1 + n_2 - 1$ df (degree of freedom)

Test statistics:

$z = (m_1 - m_2) / s_d$, where $s_d =$ standard error

$$s_d = \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \cdot \left[\frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}\right]}$$

T TESTS TO COMPARE TWO SAMPLE MEANS

Age and prostate cancer t-test

- Negative biopsy: $n_1=206$, $m_1=66.59$ years old, $s_1=8.21$
- Positive biopsy: $n_2=95$, $m_2=67.14$ years old, $s_2=7.88$
- $\sigma = 8.10$ ($n=301$)
- $\alpha = 0.05 \rightarrow t_{\text{critic}} = 1.96$
- $sd = \sqrt{((1/206+1/95) \times ((204 \cdot 8.21^2 + 94 \cdot 7.88^2) / (205+95-2)))} = 1.0055$
- $t = (m_1 - m_2) / sd = (66.59 - 67.14) / 1.0055 = -0.5470$ (p-value = 0.582)
- $-1.96 \leq -0.5470 \leq 1.96 \rightarrow$ we failed to reject the H_0 (The mean age of subjects with positive biopsy is not significantly different by the mean age of subjects with negative biopsy)
- For samples > 100 the difference between Z and t-statistic is negligible while the p-values are identical

STUDENT T-TEST FOR COMPARING TWO MEANS (UNKNOWN AND EQUAL VARIANCES)

- **Null hypothesis:** Means difference of the two populations is not significantly different by zero.
- **Alternative hypothesis** for two-tailed test: Means difference of the two populations is significantly different by equal.
- **Assumptions:**
 - The variables in the two samples are normal distributed
 - The variances are equal
- If these two assumptions are not satisfied the test loss its validity.
- If the variances of populations are known the Z test is applied (is most powerful)

STUDENT T-TEST FOR COMPARING TWO MEANS (UNKNOWN AND EQUAL VARIANCES)

- **Degree of freedom (df):**
 - $df = n_1 + n_2 - 2$
- **Significance level:** $\alpha = 0.05$
- **Critical region** for two-tailed test

$$\left(-\infty; -t_{n_1+n_2-2; \frac{\alpha}{2}} \right) \cup \left(t_{n_1+n_2-2; \frac{\alpha}{2}}; +\infty \right)$$

- **Statistics**

$$t = \frac{m_1 - m_2}{\sqrt{s \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

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

STUDENT T-TEST FOR COMPARING TWO MEANS (UNKNOWN AND EQUAL VARIANCES)

- We want to study whether there is a significant difference between the amount of blood uric acid in women from urban and rural. In a sample of 16 women aged between 30 and 50 years in urban areas, average uric acid was 5 mg/100 ml, with a variance of 2 mg/100 ml. An average equal to 4 mg/100 ml with a variance of 2 mg/100 ml was obtained on a sample of 16 women aged 30 to 50 years in rural areas.

STUDENT T-TEST FOR COMPARING TWO MEANS (UNKNOWN AND EQUAL VARIANCES)

- **Data:**
 - $n_1 = 16; n_2 = 16$
 - $m_1 = 5; m_2 = 4$
 - $s^2 = 2$
- **Null hypothesis:** There is no significant difference between the two samples means.
- **Alternative hypothesis** for two-sided test: There is a significant difference between the two samples means.
- **Degree of freedom:**
 $df = n_1 + n_2 - 2 = 16 + 16 - 2 = 30$
- **Significance level:**
 $\alpha = 0.05.$
- **Critical region** for bilateral test:
 $(-\infty; -t_{n_1+n_2-2; 0.025}] \cup [t_{n_1+n_2-2; 0.025}; +\infty)$
 $(-\infty; -2.04] \cup [2.04; +\infty)$

STUDENT T-TEST FOR COMPARING TWO MEANS (UNKNOWN AND EQUAL VARIANCES)

$$s = \sqrt{\frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(16 - 1) \cdot 2 + (16 - 1) \cdot 2}{16 + 16 - 2}} = \sqrt{\frac{60}{30}} = 1.41$$

$$t = \frac{m_1 - m_2}{\sqrt{s \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{5 - 4}{\sqrt{1.41 \cdot \left(\frac{1}{16} + \frac{1}{16} \right)}} = \frac{1}{\sqrt{1.41 \cdot 0.25}} = \frac{1}{\sqrt{0.3525}} = \frac{1}{0.5937} = 1.68$$

Conclusion:

- Statistical: The null hypothesis is failed to be rejected since the statistics did not belongs to the critical region.
- Clinical: The serum level of uric acid is not significantly different in women from rural compared to those from urban areas.

STUDENT T-TEST FOR COMPARING TWO MEANS (UNKNOWN AND EQUAL VARIANCES)

spark.rstudio.com/minebocek/dist_calc/

Distribution Calculator

Distribution:
t

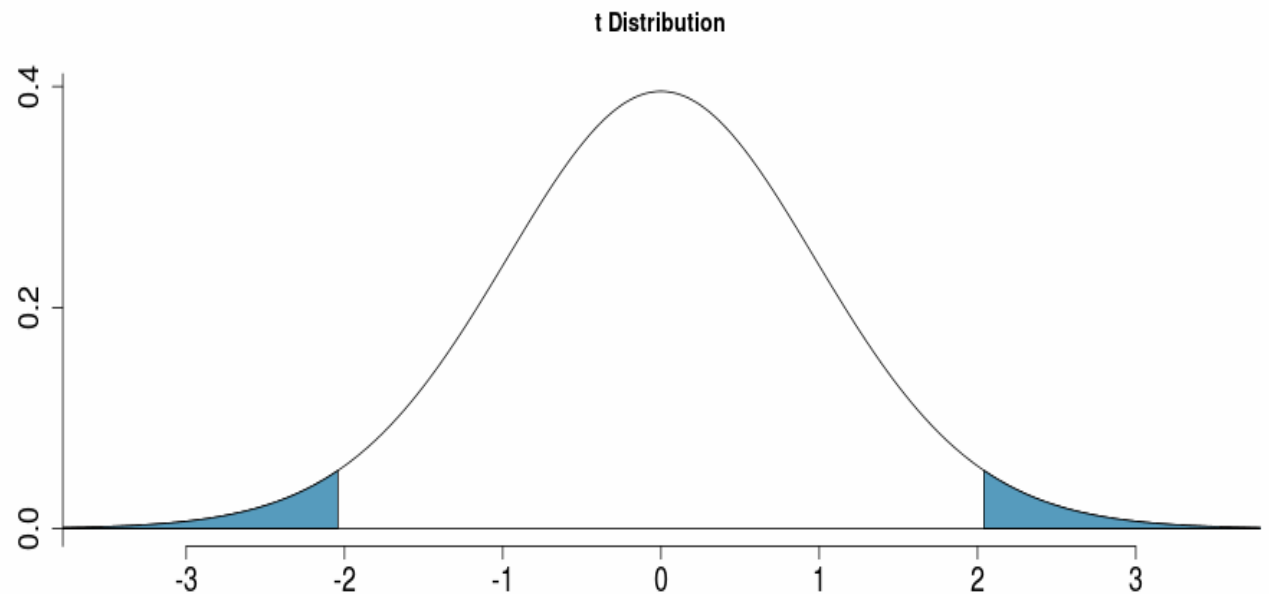
Degrees of freedom
1 30 50

Model:
 $P(X < a \text{ or } X > b)$

Find Area:
Both Tails

a
-6 -2.04 6

b
-6 2.04 6



$$P(X < -2.04 \text{ or } X > 2.04) = 0.0502$$

of asphalt binder in low, intermediate and high service temperatures.

4.2.1. The mean hypothesis test method

Due to test conditions, the hypothesis test of two populations with unknown variances and equal number of samples is used in this paper. The steps of this method are [29]:

1. $H_0 : \mu_1 - \mu_2 = 0$
2. $H_1 : \mu_1 - \mu_2 > 0$ or $H_1 : \mu_1 - \mu_2 < 0$
3. $T_{f, n_1+n_2-2} = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

where H_0 is the null hypothesis, H_1 is the alternative hypothesis, T is the test statistics, \bar{X}_1 and \bar{X}_2 are the mean value measure-

ments of two population samples, n_1 and n_2 are the numbers of each sample that has the equal quantity, and s_p^2 is the combined variance of the two samples which can be calculated as:

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

where s_1^2 and s_2^2 are the variances of two samples.

4. T has the t -distribution with $n - 1$ degree of freedom, which α is the significant level.
5. The acceptable area is defined as:

$$T < t_{1-\alpha}$$

PAIRED SAMPLES

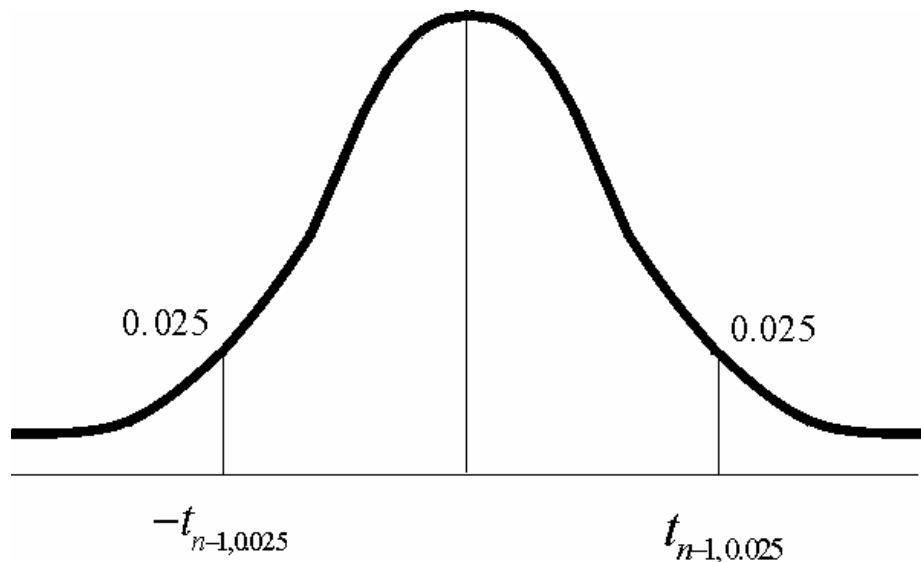
STUDENT T-TEST FOR COMPARING MEANS

- **Aim:** comparing the means of two paired samples on quantitative continuous variable (paired means the observation of the same quantitative variable before and after the action of a factor)
- **Assumptions:**
 - Individual observations from the first sample corresponds to a pair in the second sample
 - The differences between pairs of values are normally distributed.
- **Null hypothesis:** The mean of difference of paired data is not significantly different by zero.
- **Alternative hypothesis** for two-tailed test: The mean of difference of paired data is significantly different by zero.

STUDENT (T) FOR COMPARING MEANS OF PAIRED SAMPLES

- **Degrees of freedom (df):** $df = n - 1$
- **Significance level:** $\alpha = 0.05$
- **Critical region:**

$$\left(-\infty; -t_{n-1; \frac{\alpha}{2}}\right] \cup \left[t_{n-1; \frac{\alpha}{2}}; +\infty\right)$$



- **Statistics**

$$t = \frac{\bar{d}}{\frac{s}{\sqrt{n}}}$$

$$\bar{d} = \frac{(d_1 + d_2 + \dots + d_n)}{n}$$

- s = standard deviation of the differences
- n = sample size

PAIRED STUDENT T-TEST

Paired t Test

Denote the test statistic $\bar{d}/(s_d/\sqrt{n})$ by t , where s_d is the sample standard deviation of the observed differences:

$$s_d = \sqrt{\left[\frac{\sum_{i=1}^n d_i^2 - \left(\sum_{i=1}^n d_i \right)^2 / n}{n-1} \right]}$$

n = number of matched pairs

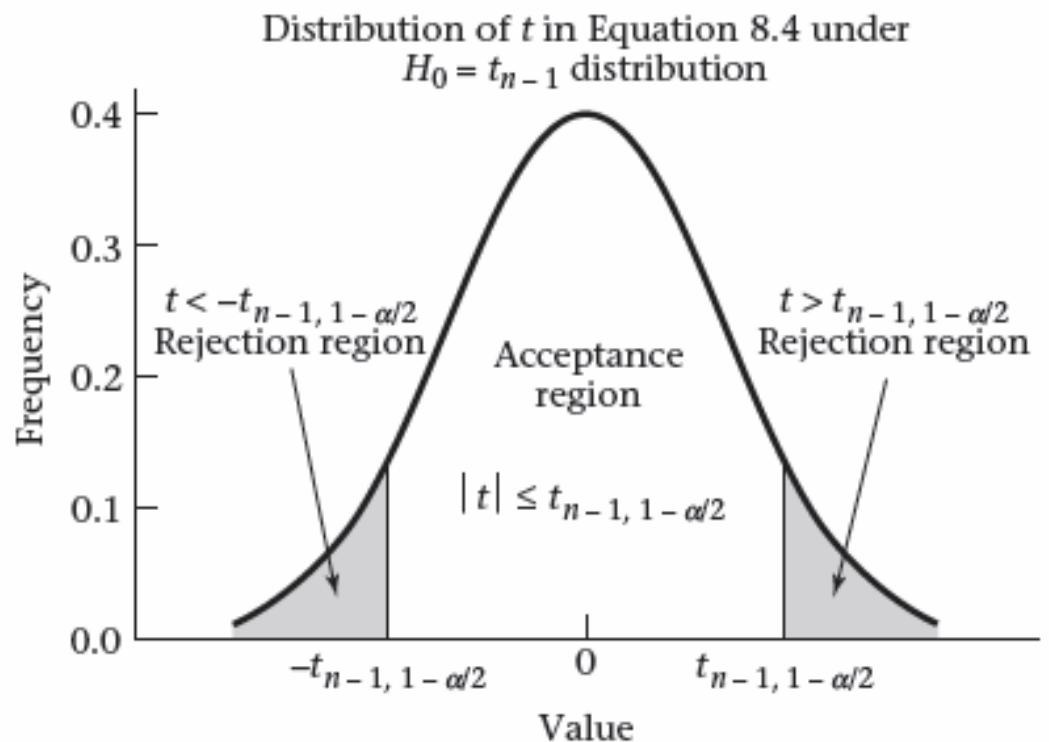
If $t > t_{n-1, 1-\alpha/2}$ or $t < -t_{n-1, 1-\alpha/2}$

then H_0 is rejected.

If $-t_{n-1, 1-\alpha/2} \leq t \leq t_{n-1, 1-\alpha/2}$

then H_0 is accepted. The acceptance and rejection regions are shown in Figure 8.1.

Figure 8.1 Acceptance and rejection regions for the paired t test



PAIRED STUDENT T-TEST

SBP levels (mm Hg) in 10 women while not using (baseline) and while using (follow-up) OCs

i	SBP level while not using OCs (x_{i1})	SBP level while using OCs (x_{i2})	d_i^*
1	115	128	13
2	112	115	3
3	107	106	-1
4	119	128	9
5	115	122	7
6	138	145	7
7	126	132	6
8	105	109	4
9	104	102	-2
10	115	117	2

$$^*d_i = x_{i2} - x_{i1}$$

Assume that the SBP of the i th woman is normally distributed at baseline with mean μ_1 and variance σ^2 and at follow-up with mean $\mu_1 + \Delta$ and variance σ^2 .

STUDENT T-TEST FOR COMPARING MEANS OF PAIRED SAMPLES

- **Null hypothesis:** There is no significant difference in systolic blood pressure before and after having used of oral contraceptives.
- **Alternative hypothesis** for two-tailed test: There is a significant difference in systolic blood pressure before and after having used of oral contraceptives.
- **Degrees of freedom:** $df = n - 1 = 10 - 1 = 9$

STUDENT T-TEST FOR COMPARING MEANS OF PAIRED SAMPLES

$$\bar{d} = \frac{13 + 3 - 1 + 9 + 7 + 7 + 6 + 4 - 2 + 2}{10} = \frac{48}{10} = 4.8$$

$$s = \sqrt{\frac{67.24 + 3.24 + 33.64 + (4.2)^2 + 2 \cdot 4.84 + 1.44 + 0.64 + 46.24 + 7.84}{10 - 1}} = \sqrt{\frac{187.60}{9}} = \sqrt{20.84} = 4.57$$

$$t = \frac{\bar{d}}{\frac{s}{\sqrt{n}}} = \frac{4.8}{\frac{4.57}{\sqrt{9}}} = \frac{4.8}{\frac{4.57}{3}} = \frac{4.8}{1.52} = 3.15$$

Conclusion (two-sided test):

- Statistical: The null hypothesis is rejected since the statistics belongs to critical region.
- Clinical: The use of oral contraceptives is associated to a significant increase in systolic blood pressure.

TESTS ON MEANS BY EXAMPLES

<http://www.hindawi.com/journals/tswj/2013/608683/tab2/>

Table 2: The values of blood pressures and heart rates prior to and after the treatment with metoprolol and nebivolol in the study

	Total (<i>n</i> = 60)	Metoprolol (<i>n</i> = 30)	Nebivolol (<i>n</i> = 30)	<i>P</i> value
Systolic blood pressure (mm/Hg)				
Before	152.1 ± 4.9	151.3 ± 3.6	152.8 ± 5.9	0.27**
After	136.4 ± 10.9	134.9 ± 10.5	137.8 ± 11.3	0.3**
<i>P</i> value	<0.001*	<0.001*	<0.001*	
Diastolic blood pressure (mm/Hg)				
Before	91.7 ± 3.9	91.1 ± 3.7	92.3 ± 4.1	0.25**
After	82.6 ± 7.1	82.4 ± 6.7	82.7 ± 7.3	0.87**
<i>P</i> value	<0.001*	<0.001*	<0.001*	
Heart rate (pulse/min)				
Before	75.7 ± 5.7	75.2 ± 5.7	76.2 ± 5.7	0.5**
After	70.2 ± 5.4	70.3 ± 5.8	70.1 ± 5.1	0.87**
<i>P</i> value	<0.001*	<0.001*	<0.001*	
Achieved targeted blood pressure (<i>n</i> , %)	37 (61.6)	19 (63.3)	18 (60)	0.5***

*Paired *t*-test; **independent samples *t*-test; ***chi-square test.

Table 3: The values of blood pressures and heart rates measured during rest, exercise, and recovery periods in the study population.

	Total (<i>n</i> = 60)	Metoprolol (<i>n</i> = 30)	Nebivolol (<i>n</i> = 30)	<i>P</i> value
Systolic blood pressure (mm/Hg)				
Rest	137.4 ± 13.7	135.3 ± 13.5	139.6 ± 13.8	0.22
Stage 1	152 ± 22.5	151.7 ± 20.9	152.3 ± 24.4	0.91
Stage 2	162.7 ± 27.1	160.8 ± 27.1	164.6 ± 27.5	0.61
Stage 3	175.3 ± 29.3	174.3 ± 29.7	176.2 ± 29.7	0.84
Recovery (3 min)	164.4 ± 27.4	167.1 ± 29.8	161.7 ± 25	0.46
Recovery (5 min)	156.5 ± 27.9	158.5 ± 27	154.5 ± 29	0.58
Diastolic blood pressure (mm/Hg)				
Rest	82.6 ± 9.9	83.7 ± 10.6	81.5 ± 9.3	0.39
Stage 1	76 ± 21.2	78.6 ± 21.5	73.3 ± 20.9	0.34
Stage 2	72.9 ± 18.7	76.7 ± 16.5	68.9 ± 20.2	0.12
Stage 3	71.5 ± 19.5	77.3 ± 11.5	66.7 ± 23.4	0.07
Recovery (3 min)	77.9 ± 17	81.8 ± 16.6	74.1 ± 16.7	0.08
Recovery (5 min)	81.9 ± 12.9	83.2 ± 14.1	80.7 ± 11.8	0.45
Heart rate per minute				
Rest	71.7 ± 5.5	72.1 ± 5.3	71.2 ± 5.9	0.63
Stage 1	106.7 ± 16.4	105.2 ± 17.2	108.1 ± 15.7	0.49
Stage 2	122.8 ± 17.2	119.6 ± 16.6	126.3 ± 17.6	0.15
Stage 3	139.5 ± 18.5	140.3 ± 18.1	138.9 ± 19.2	0.82
Recovery (3 min)	118.7 ± 16.3	117.2 ± 17.7	120.1 ± 14.9	0.5
Recovery (5 min)	90.3 ± 14.5	89.4 ± 15.2	91.2 ± 14	0.63

Thank you for your attention!

Study habits...

