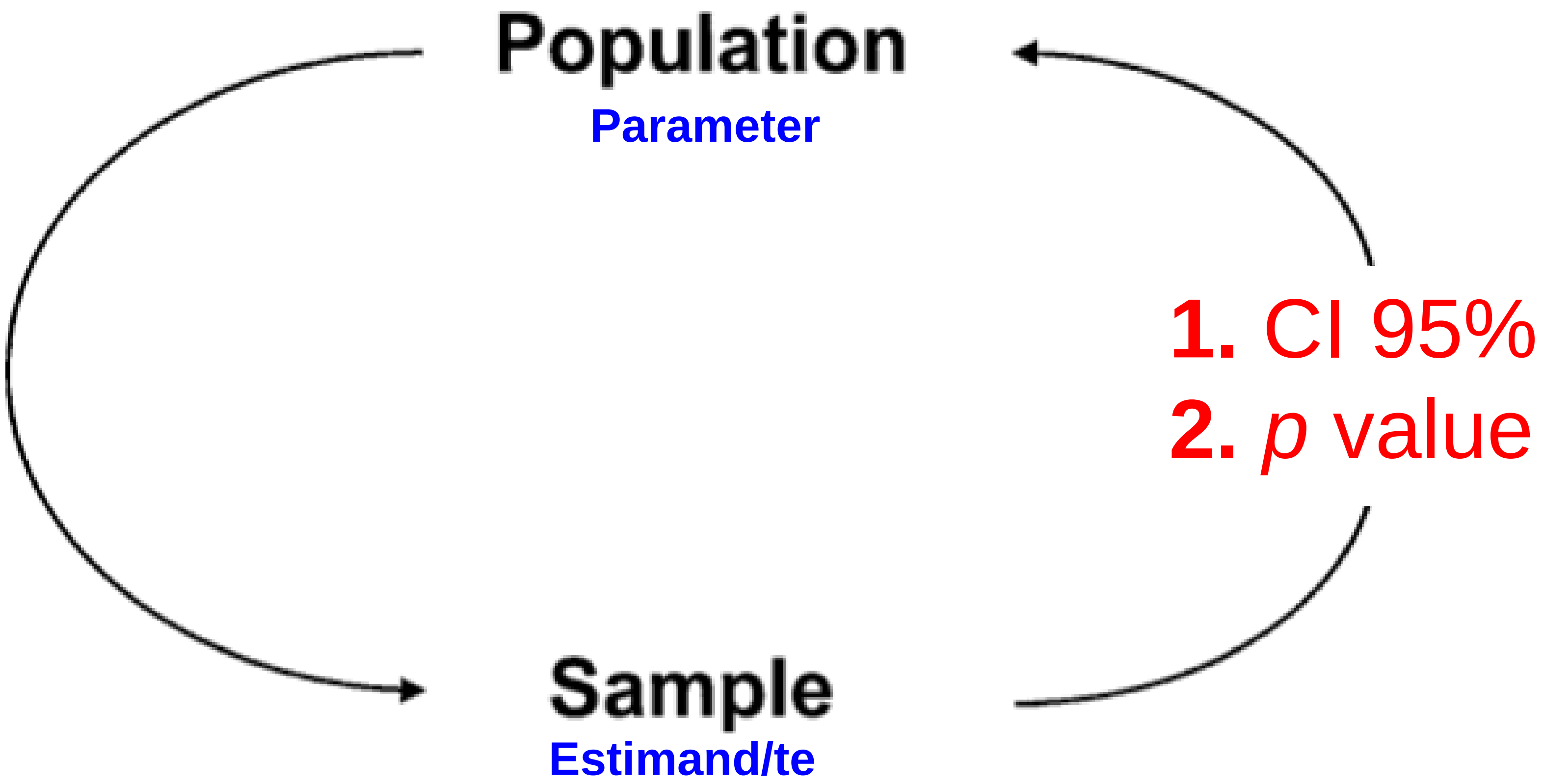
ANALYSIS OF CONTINUOUS OUTCOMES

Jose Andres Calvache MD MSc PhD
Department of Anesthesiology, Universidad del Cauca, Colombia
Department of Anesthesiology, Easmus University Medical Centre, Rotterdam, The Netherlands
@jacalvache


Population
Parameter
1. CI 95%
2. p value
Sample
Estimand/te

Test for a simple group

- N = [General population living in Popayán]
  - What is the average systolic blood pressure in the population?
  - I develop my study and select a sample (n)

- H_0 : $\mu = \mu_0$
- H_1 : $\mu \neq \mu_0$

I'm going to contrast the average of my sample against a certain value (population μ_0 *)*

This will allow me to decline or not reject H_0

Test for a simple group

- There are two ways to approach this contrast
- If n ≥ 30, there are virtally no differences between the two

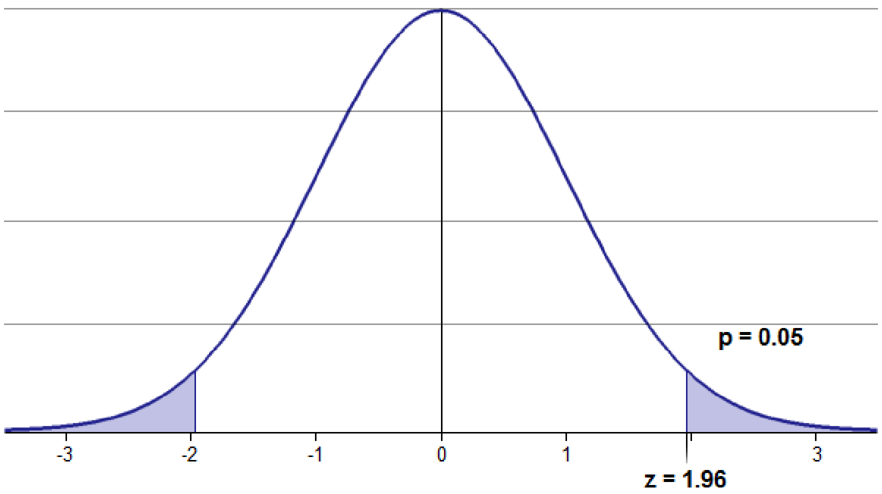
$$Z = \frac{X - \mu}{SEM} = \frac{X - \mu}{s/\sqrt{n}}$$

$$t = \frac{X - \mu}{s/\sqrt{n-1}}$$

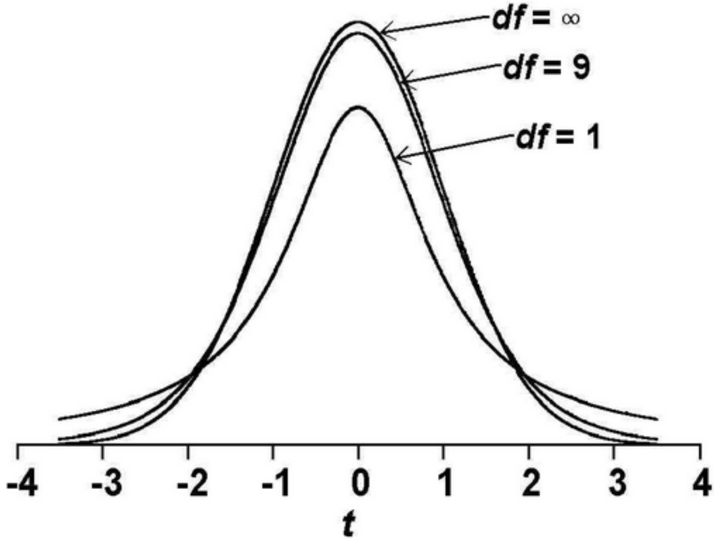
Z Test

T test
Student T test

Z test



T test / Student T test



- Either, results presented as a p value
- What represents the evidence in my sample for rejecting H0

Example,

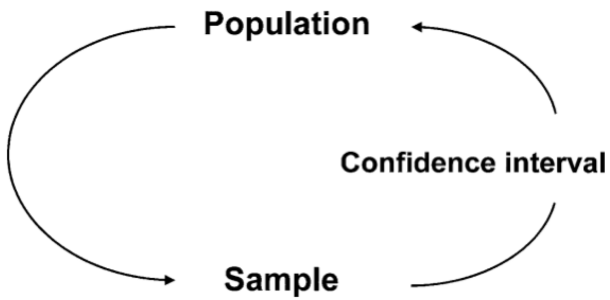
- I think the SBP of the population is 120 mmHg !
- I select my sample,
  - H_0 : $\mu = 120$
  - H_1 : $\mu \neq 120$

X = 130 mmHg

- I apply one of the tests described (Z, t-test)
- P-value = 0,01
- There is enough evidence in my sample to reject H0

Example,

- X = 130 mmHg versus 120 mmHg



$$IC\ 95\% = X \pm 1.96\ (\mathit{SEM})$$

IC 95% = [126 – 135]

Test to compare independent means

- I have samples that come from two groups independent of each other
- Cohort study, to study cardiovascular risk at a given site
  - I am interested to know if the BMI of men and women in my study is different (differs significantly)
    - Group of men in follow-up
    - Group of women in follow-up

	BMI	**SD**	**n**	**SEM**
Man	25.53	2.44	74	0.28
Women	27.18	5.08	95	0.52

Test para comparar medias independientes

$H_0: \ \mu_H = \mu_M$

$H_1: \ \mu_H \neq \mu_M$

$H_0: \ \mu_H - \mu_M = 0$

$H_1: \ \mu_H - \mu_M \neq 0$

$X_H - X_M = 25.53 - 27.18 =$ **-1.65** ***p* = 0.005**

$$SEM_{X_1 - X_2} = \sqrt{(SEM_1)^2 + (SEM_2)^2}$$

$$SEM_{X_H - X_M} = \sqrt{(SEM_H)^2 + (SEM_M)^2}$$

$$SEM_{X_H - X_M} = \sqrt{(0.28_H)^2 + (0.52_M)^2} = 0.59$$

IC 95% = -1.65 ± 1.96 (SEM_{diff}) = -1.65 ± (0.59) = **[-2.81 - -0.49]**

Conclusion

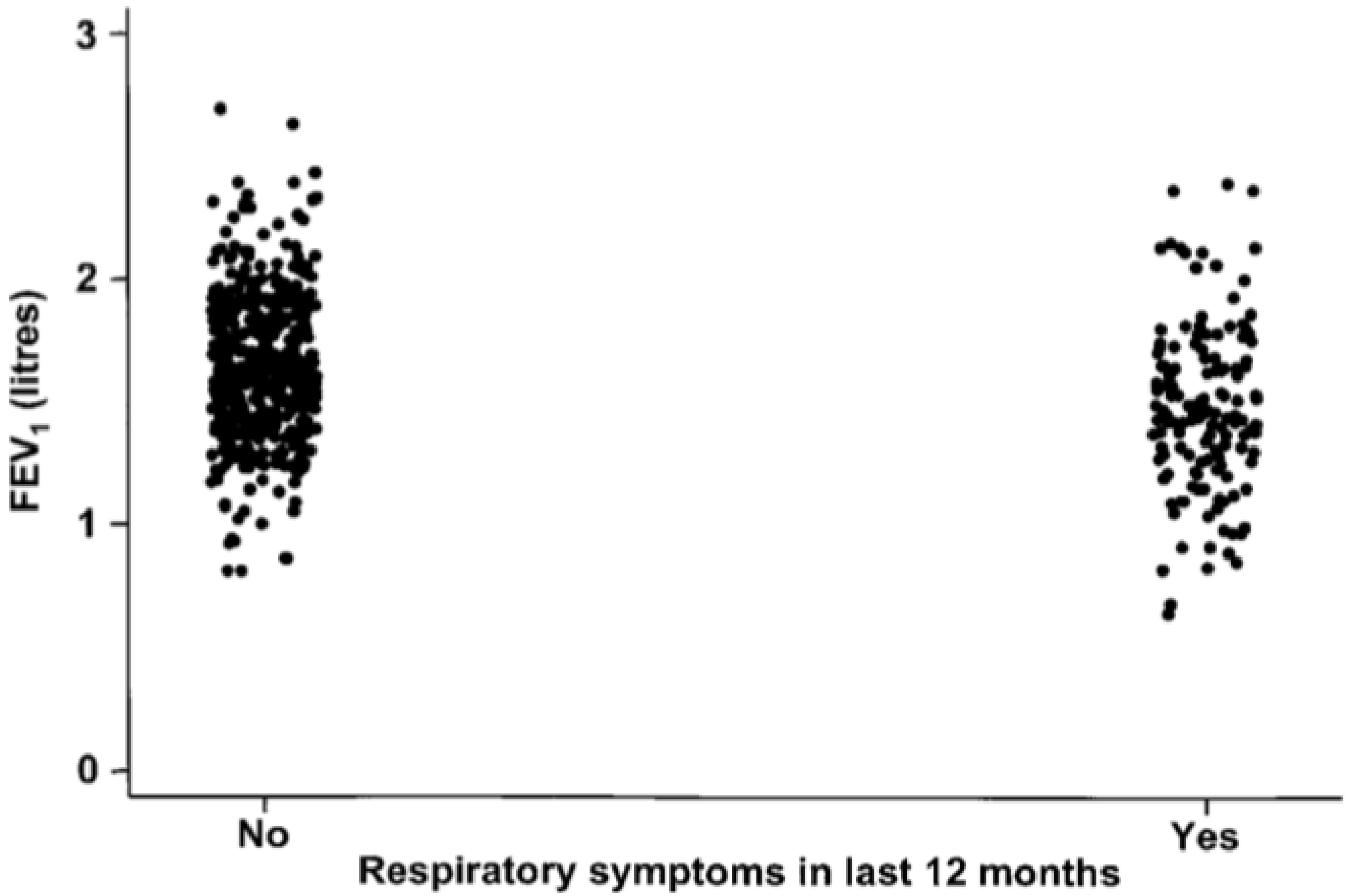
- There is a statistically significant difference in the BMI of men and women

- Such a difference is : –1.69 kg/m^2 IC95% [-2.81- -0.49] (*p*=0.005)

- *On average men have a BMI 1.69 kg/m2 lower than women, but such a difference can be between -2.81 to -0.049*

- *This CI does not contain zero, so it suggests a real difference at the population level in the BMI of men and women.*

Scatter plot showing the relationship between FEV1 and respiratory symptoms in 636 children living in a deprived suburb of Lima, Peru.



Assumptions of the T test

- Single group
  1. The distribution of the variable at the population level is normal with an unknown variance

- Two independent groups
  - The distribution of variables in the population is normal
  - The variances of the variables are equal (*Levene test*)

	IMC	DE	n	SEM
Hombre	25.53	2.44	74	0.28
Mujeres	27.18	5.08	95	0.52

Test to compare related means

- In this case, the two samples to be compared are paired.
  - Assess blood pressure before and after treatment in a group of subjects (the same before and after)
  - A group of students are evaluated at the beginning and end of the academic semester
  - A numerical variable is measured in a group of subjects several times over time (e.g., hemoglobin)

Applying the concept of the t-test for independent samples here is incorrect

Test to compare related means

- 10 obese subjects underwent a diet for 1 month. The concentration of a pro-inflammatory substance in the blood (IL-10) was measured before and after

	Before X_0	After X_1	$X_0 - X_1$
1	6.1	5.2	0.9
2	6.0	7.9	−1.9
3	8.2	3.9	4.3
4	7.6	4.7	2.9
5	6.5	5.3	1.2
6	5.4	7.4	−2.0
7	6.9	4.2	2.7
8	6.7	6.1	0.6
9	7.4	3.8	3.6
10	5.8	7.3	−1.5

$\mathbf{X_{diff} = 1.08}$

Test to compare related means

$$H_0: \quad \mu_{diff} = 0$$
$$H_1: \quad \mu_{diff} \neq 0$$

$\mathbf{X_{diff} = 1.08}$ $\quad\quad$ $SEM_{diff} = 0.73$

Paired t test : P-value = 0.168

1.08 CI 95% [-0.57 – 2.73]

- We do not have enough evidence to reject H_0
- In other words, there is no evidence that the intervention significantly reduces the value of IL-10
- The CI contains 0

To investigate whether smoking reduces lung function, forced vital capacity (FVC, a test of lung function) was measured in 100 men aged 25–29, of whom 36 were smokers and 64 non-smokers.

Table 7.1 Results of a study to investigate the association between smoking and lung function.

Group	Number of men	Mean FVC (litres)
Smokers (1)	$n_1 = 36$	$\bar{x}_1 = 4.7$
Non-smokers (0)	$n_0 = 64$	$\bar{x}_0 = 5.0$

X(smok) – X(non-smok) = 4,7 – 5,0 = **- 0,3 litres.**

$$\text{s.e.} = \sqrt{(\text{s.e.}_1^2 + \text{s.e.}_0^2)} = \sqrt{(0.1^2 + 0.075^2)} = 0.125 \text{ litres}$$

The 95% confidence interval for the population difference in mean FVC is therefore:

$$\begin{aligned} 95\%\ \text{CI} &= -0.3 - (1.96 \times 0.125) \text{ to } -0.3 + (1.96 \times 0.125) \\ &= -0.545 \text{ litres to } -0.055 \text{ litres} \end{aligned}$$

$$z = \frac{\bar{x}_1 - \bar{x}_0}{\text{s.e.}} = \frac{\bar{x}_1 - \bar{x}_0}{\sqrt{(\sigma_1^2/n_1 + \sigma_0^2/n_0)}} \qquad z = \frac{-0.3}{0.125} = -2.4$$

$$p = 0.0164 \; (= 0.0082 + 0.0082)$$

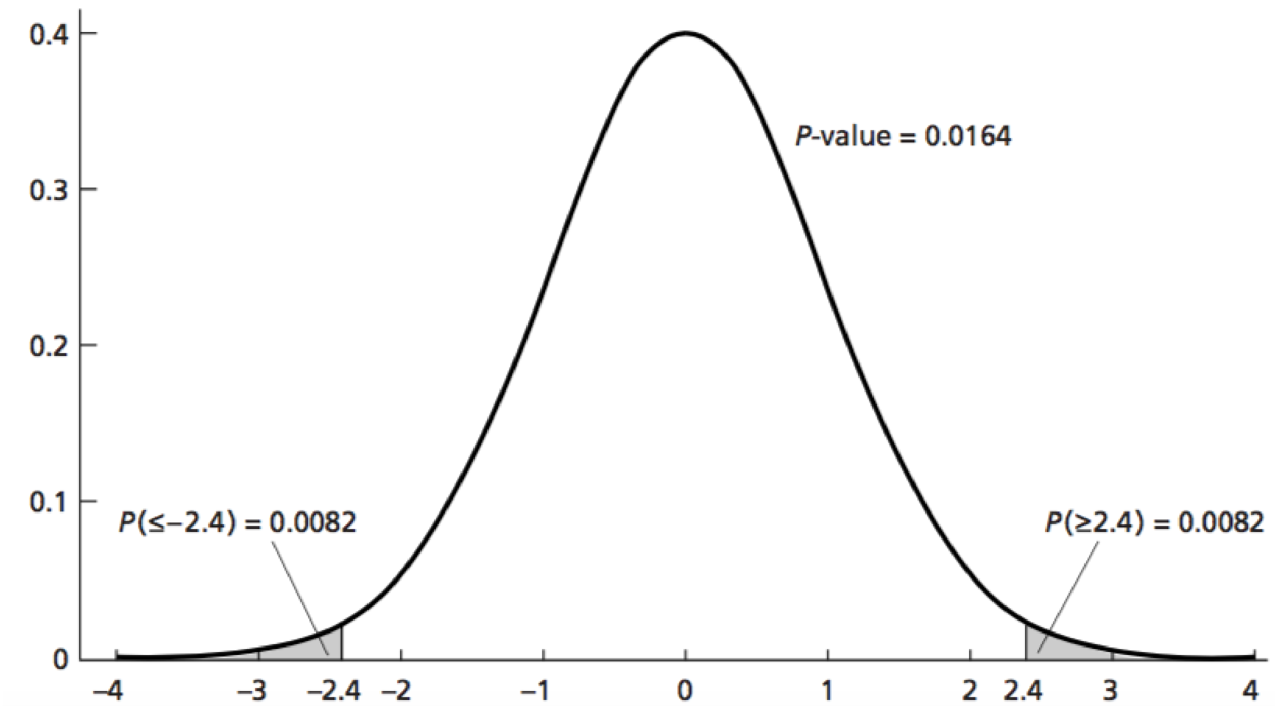
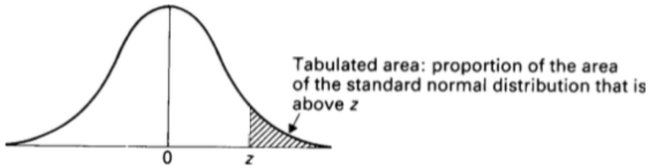


Table A1 Areas in tail of the standard normal distribution.

Adapted from Table 3 of White *et al*. (1979) with permission of the authors and publishers.



	Second decimal place of z									
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
2.2	0.01390	[illegible]	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
2.3	0.01072	[illegible]	[illegible]	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
2.4	0.00820	[illegible]	[illegible]	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
2.5	0.00621	[illegible]	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.00100
3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003

$p = \ 0.0164\ (= 0.0082 + 0.0082)$

$p = 0.01$

Table 8.2 Results of five trials of drugs to lower serum cholesterol, presented as mean difference (drug group minus control group), s.e. of the difference, 95% confidence interval and *P*-value.

Trial	Drug	Cost	No. of patients per group	Difference in mean cholesterol (mg/decilitre)	s.e. of difference	95% CI for difference	*P*-value
1	A	Cheap	30	−40	40	−118.4 to 38.4	0.32
2	A	Cheap	3000	−40	4	−47.8 to −32.2	< 0.001

Note that the estimated effect of drug A was the same (a mean reduction of 40 mg/decilitre) in trials 1 and 2. However because trial 1 was small it provided no evidence against the null hypothesis of no treatment effect. This illustrates an extremely important point: in small studies *a large P-value does not mean that the null hypothesis is true*. This is summed up in the phrase '*Absence of evidence is not evidence of absence*'.

Table 8.2 Results of five trials of drugs to lower serum cholesterol, presented as mean difference (drug group minus control group), s.e. of the difference, 95% confidence interval and *P*-value.

Trial	Drug	Cost	No. of patients per group	Difference in mean cholesterol (mg/decilitre)	s.e. of difference	95% CI for difference	*P*-value
1	A	Cheap	30	−40	40	−118.4 to 38.4	0.32
2	A	Cheap	3000	−40	4	−47.8 to −32.2	< 0.001
3	B	Cheap	40	−20	33	−84.7 to 44.7	0.54
4	B	Cheap	4000	−2	3.3	−8.5 to 4.5	0.54
5	C	Expensive	5000	−5	2	−8.9 to −1.1	0.012

Even when the *P*-value shows strong evidence against the null hypothesis, it is vital to examine the confidence interval to ascertain the range of values for the difference between the groups that is consistent with our data. The *medical importance* of the estimated effect should always be considered, even when there is good statistical evidence against the null hypothesis.

Non-parametric tests for means

- Non-parametric statistics do not follow theoretical distributions of random variables

- It does not require the conditions that parametrics do require.
- They are based solely on the own distribution of the data in the sample
- They are also called free-of-distribution tests
- Non-parametric tests have "less statistical power"

Non-parametric tests for means

Mann-Whitney U Test // Wilcoxon test

- Wilcoxon rank sum test
- Mann-Whitney-Wilcoxon sum rank test

- It is based on the median and ranges of the data
- Not affected by extreme values
- Requires access to the total data (*dataset*)
- Its result is only in terms of p value

Mann-Whitney U test

- A group of ten students were randomized to receive a new pedagogical scheme. At the end, an evaluation exam was carried out on each group. The results are the difference between the post-pre test.

• Intervention (I)	5	0	7	2	19
• Control (C)	6	-5	-6	1	4

  - Only 10 subjects under study
  - It is possible that in 10 subjects the results of the post-pre test are not normally distributed

Mann-Whitney U test

1. Reorganize data

	-6	-5	0	1	2	4	5	5	7	19
Rank	1	2	3	4	5	6	7	8	9	10
Group	C	C	I	C	I	C	I	C	I	I

2. Compare the ranges of each group using statistical techniques and provide a p-value as a result

2. P-value = 0.17

There is no statistically significant difference in the assessment of improvement between the control and intervention group

When t-test and when Mann Whitney U test ?

- If your sample is small (e.g. less than 30)
- If there are suspicions that the distributions of the variables do not follow a normal trend
- Non-parametric tests are more "conservative“

- There is no consensus