# Statistical significance using *p* -value

**Objectives:**

- Able to understand the concepts of statistical inference and statistical significance.
- Able to apply the concept of statistical significance(p-value) in analyzing the data.
- Able to interpret the concept of statistical significance(p-value) in making valid conclusions.

**Team Members:** Laila Mathkour - Khalid Aleedan - Khalid Alhusainan-mohammed ghandour

**Team Leaders**: Rawan Alwadee & Mohammed ALYousef

**Revised By:** Maha Alghamdi

**Dr. Shaffi Ahamed**

**Resources:**

- 436 Lecture Slides + Notes

Important – Notes – KAPLAN Behavioral Science and Social Sciences

436researchteam@gmail.com

Editing file

Feedback link

## Hypothesis Testing

### One sided two sided test

- two tailed tests (not directional)
- One tailed tests (directional)

### null Hypothesis

There is no association between the predictor and outcome variables in the population

### Alternate Hypothesis

There is no association between the predictor and outcome variables in the population

## Type-I and Type-II Errors

- ❖ α = Probability of rejecting H0 when H0 is true
- ❖ α is called significance level of the test
- ❖ β = Probability of not rejecting H0 when H0 is false
- ❖ 1-β is called statistical power of the test

## Statistically significant and not statistically significant

### significant

Sample value not compatible with Ho

### not significant

Sample value not compatible with Ho

The purpose of inferential statistics is to designate how likely it is that a given finding is simply the result of chance. Inferential statistics would not be necessary if investigators studied all members of a population. However, because we can rarely observe and study entire populations, we try to select samples that are representative of the entire population so that we can generalize the results from the sample to the population.

**Why use inferential statistics at all?**

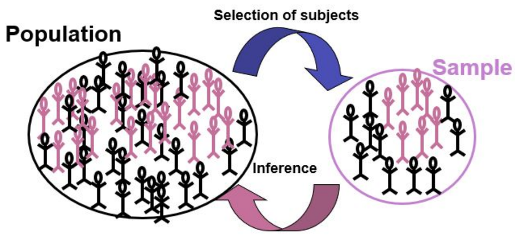
- Average height of <u>all</u> 25-year-old men **(population)** in KSA is a **PARAMETER**.
  Whatever you calculate is called parameter
- The height of the members of a (**sample)** of 100 such men are measured; the average of those 100 numbers is a **STATISTIC.**

❖ Using inferential statistics, we make inferences about population (taken to be unobservable) based on a random sample taken from the population of interest. (we can generate the parameter from the statistic)

| Parameter | Statistic |
|---|---|
| Descriptive measure of a population | Descriptive measure of a sample |
| Not always possible to measure because it needs the actual value in the population | Always possible to measure because it doesn't need the actual value in the population |

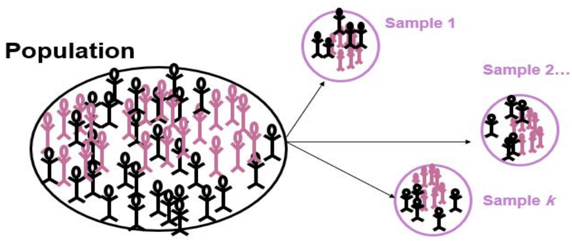
**Is risk factor X associated with disease Y?**

From the sample, we compute an estimate of the effect of X on Y (e.g., risk ratio if cohort study):
- Is the effect real? Did chance play a role?



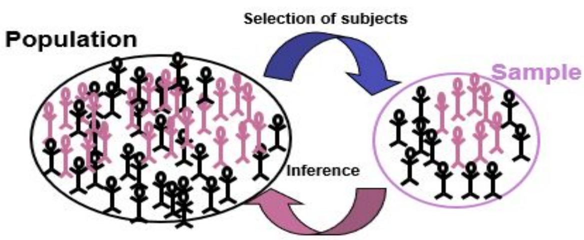

**Why worry about chance?**( because of variability)

Sampling variability…
- you only get to pick one sample!




**Interpreting the results**

Make inferences from data collected using laws of probability and statistics. You have to use these two concepts
- tests of significance (p-value) this lecture
- confidence intervals, next lecture



## Significance testing:(p-value)

- The interest is generally in comparing two groups (e.g., risk of outcome in the treatment and placebo group) like cohort study
- The statistical test depends on the type of data and the study design (we did not use p-value in descriptive data)

Basic steps of statistical inference:
a. Define the research question: what are you trying to show?
b. Define the null hypothesis, generally the opposite of what you hope to show
i. Null hypothesis says that the findings are the result of chance or random factors. If you want to show that a drug works, the null hypothesis will be that the drug does NOT work.
ii. Alternative hypothesis says what is left after defining the null hypothesis. In this example, that the drug does actually work. c. Two types of null hypotheses
i.One-tailed, i.e., directional or "one-sided," such that one group is either greater than, or less than, the other. E.g., Group A is not < than Group B, or Group A is not > Group B
ii. Two-tailed, i.e., nondirectional or "two-sided," such that two groups are not the same. E.g., Group A = Group B

## Hypothesis Testing

- Null Hypothesis (means zero → H0 )(mean there is no different between the two group)
  - There is no association between the predictor (associated factors) and outcome variables in the population
  - Assuming there is no association, statistical tests estimate the probability that the association is due to chance (this is what we will estimate)
- Alternate Hypothesis Ha
  - The proposition that there is an association between the predictor and outcome variable
  - We do not test this directly but accept it by default if the statistical test rejects the null hypothesis

It is important to know that we test null hypothesis not the alternative hypothesis
So if data are giving evidence that there IS a difference we reject null hypothesis, so by default accept alternative hypothesis.

### The Null Hypothesis, $H_o$

- States the assumption (numerical) to be tested
- Begin with the assumption that the null hypothesis is TRUE
- Always contains the '=' sign

### The alternative hypothesis, $H_a$

- Is the opposite of the null hypothesis
- Challenges the status quo
- Never contains just the '=' sign
- Is generally the hypothesis that is believed to be true by the researcher

### One and Two Sided Tests

- Hypothesis tests can be one or two sided (tailed)

❖ **One tailed tests are directional:**لازم تكون متاكد من النتيجة عشان تستخدمه

$H_o$ : μ1- μ2= 0

$H_a$: μ1- μ2 > 0 or HA: μ1- μ2< 0

❖ **Two tailed tests are not directional**:

$H_o$ : μ1- μ2= 0

$H_a$ : μ1- μ2≠ 0

Hypothesis can be one or two sided: (what do we mean by direction?)

- For example: I make a statement there is no difference between male and female grades, I am biased; I make a statement: female's mean grades are higher, this is my alternative hypothesis. I am giving my interest in female grades so I am giving a direction. So you give a direction weather one grope is higher or lower.
- For example, you are doing an interventional study, you are going to reduce mortality. You know it is going to work positively, so you use one side study.

| One sided test | Two sided test |
|---|---|
| A statistical hypothesis test in which alternative hypothesis has only one end. So, it will tell you if there is a relationship between variables in single direction. | A significance test in which alternative hypothesis has two ends. So, if there is a relationship between variables in both direction. |
| The hypothesis directional | The hypothesis non-directional |
| Region of rejection is either left or right | Region of rejection are both left & right |

## Type-I and Type-II Errors

- α = Probability of rejecting $H_o$ when $H_o$ is true
- α is called significance level of the test
- β = Probability of not rejecting $H_o$ when $H_o$ is false
- 1-β is called statistical power of the test

| Examples of type 1 & 2 errors: | |
|---|---|
| α:Type 1 | β : Type 2 |
| In reality there is no deference but you are making error and say there is a difference | In reality there is a deference but you are making error and say there is no difference |

## Diagnosis and statistical reasoning

| Test result | Disease status: Present | Disease status: Absent |
|---|---|---|
| +ve | True +ve (sensitivity) | False +ve |
| -ve | False –ve | True -ve (Specificity) |

| Test result | Significance Difference is Present (Ho *not* true) | Significance Difference is Absent (Ho is true) |
|---|---|---|
| Reject Ho | No error 1-β | Type I err. α |
| Accept Ho | Type II err. β | No error 1–α |

a : significance level
1-b : power

- True positive: When the patient in reality is sick and your test confirms that the patient is sick. (no issue)
- True negative: When the patient in reality isn't sick and your test confirms that the patient isn't sick. (no issue)
- False positive: When the patient in reality isn't sick and your test state that the patient is sick. (type one error)
- False negative: When the patient in reality is sick and your test state that the patient isn't sick. (type two error)

Extra pic.

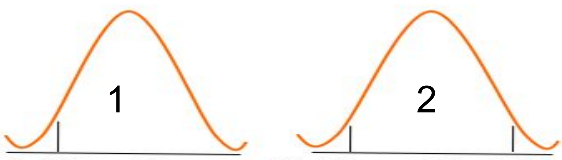
- No test will give 100% sensitivity and 100% specificity. There will be always some number in false negative and false positive.
- You prefix the type one error (α ) as 5% (0.05), why? Recall normal distribution concept, in normal distribution we considered normal as 95%, remining 5% we considered it abnormal.
- In clinical medicine, we accept 95% as normal, and 5% as type one error.

### When To Reject $H_o$ ?

Rejection region: set of all test statistic values for which $H_o$ will be rejected

Level of significance, α (the probability of rejecting the null hypothesis in a statistical test when it is true ):
Specified before an experiment to define rejection region

- One Sided : α = 0.05. it will have direction right or left.
- Two Sided: α/2 = 0.025 here because it two sided, 0.05 will split to two parts.

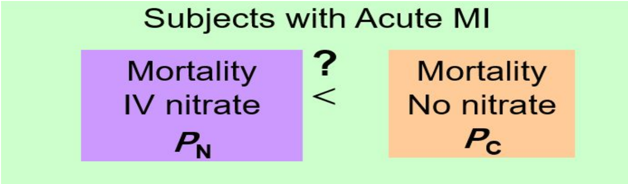


Critical Value = -1.64    Critical Values = -1.96 and +1.96

The z-score needed to reject H0 is called the critical value for significance.

- Pay attention to if it's right or left one sided test
- [Helpful video to understand and calculate critical value.]

## Clinical scenario (1) : Significance testing




Your sample size is people who have acute MI.
You subdivide your subjects who have acute MI into two groups. The first group you are giving them IV nitrate while the second group you are not giving them IV nitrate.
Why you do that? Because you want to see the effect of IV nitrate on mortality and answer your question which is IV nitrate decrease the mortality rates in people who have MI.

- Suppose we do a clinical trial to answer the above question
- Even if IV nitrate has no effect on mortality, due to sampling variation, it is very unlikely that $P_N = P_C$
- Any observed difference b/w groups may be due to treatment or a coincidence (or chance)

### Null Hypothesis(Ho)

- There is no association between the independent and dependent/outcome variables
  - Formal basis for hypothesis testing
- In the example, Ho :"The administration of IV nitrate has no effect on mortality in MI patients" or $P_N - P_C = 0$

### Obtaining P values

| Trial | Number dead / randomized: Intravenous nitrate | Number dead / randomized: Control | Risk Ratio | 95% C.I. | P value |
|---|---|---|---|---|---|
| | | | How do we get this *p*-value? | | |
| Chiche | 3/50 = 0.06 | 8/45 = 0.17 | 0.33 | (0.09,1.13) | 0.08 |
| Bussman | 4/31 | 12/29 | 0.24 | (0.08,0.74) | 0.01 |
| Flaherty | 11/56 | 11/48 | 0.83 | (0.33,2.12) | 0.70 |
| Jaffe | 4/57 | 2/57 | 2.04 | (0.39,10.71) | 0.40 |
| Lis | 5/64 | 10/76 | 0.56 | (0.19,1.65) | 0.29 |
| Jugdutt | 24/154 | 44/156 | 0.48 | (0.28, 0.82) | 0.007 |

*Table adapted from Whitley and Ball. Critical Care; 6(3):222-225, 2002*

In the table, there are the 6 studies in the first column, sample size of iv nitrate patients and control in the second and third column.
So in IV nitrate (in chiche study) 50 patients were randomized, yet 3 have died ( people who died\ total) and we are interested to know how we got the p value and its interpretation?

### Example of significance testing

**In the Chiche trial:**

$p_N = 3/50 = 0.06$; $p_C = 8/45 = 0.178$

- P: proportion of **N**: nitrate IV patients & **C**: control.
- There is a difference between the two proportions, but is it real? Or by chance? We need a statistical evidence.

**Null hypothesis:**

$H_0: p_N - p_C = 0$ or $p_N = p_C$

No difference.

**Statistical test:**

Two-sample proportion

### Test statistic for Two Population Proportions

The test statistic for $p_1 - p_2$ is a Z statistic:

$$Z = \frac{(p_N - p_C) - (P_N - P_C)_0}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_N} + \frac{1}{n_C}\right)}}$$

Observed difference: $(p_N - p_C)$

Null hypothesis: $(P_N - P_C)_0 = 0$

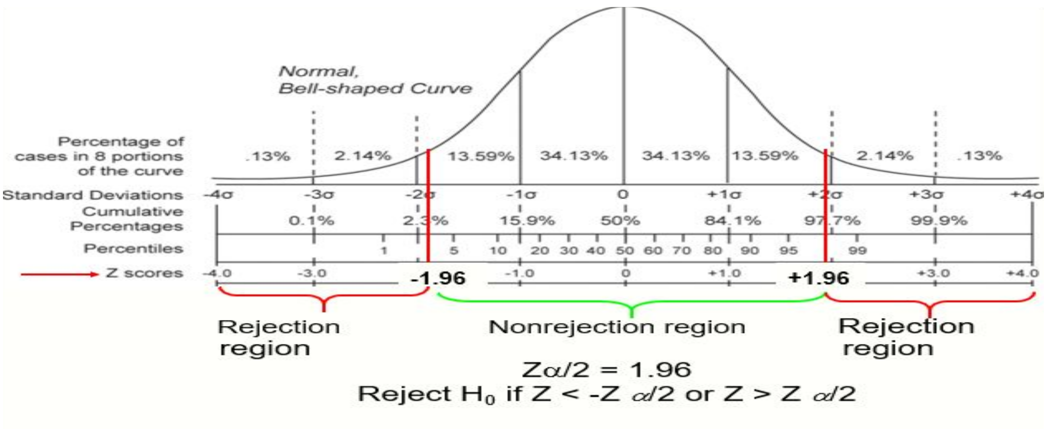
$n_N$: No. of subjects in IV nitrate group

$n_C$: No. of subjects in control group

where $\bar{p} = \frac{X_N + X_C}{n_N + n_C}$ , $p_N = \frac{X_N}{n_N}$ , $p_C = \frac{X_C}{n_C}$

(1) Reference article: Yusuf S, Collins R, MacMahon S, Peto R:Effect of Intravenous nitrates on mortality in acute myocardial infarction: an overview of the randomised trials.Lancet 1988,1:1088-1092

## Testing significance at 0.05 level:




## Two Population Proportions

$$Z = \frac{(0.06-0.178)}{\sqrt{0.116(1-.116)\left(\frac{1}{50}+\frac{1}{45}\right)}} = -1.79$$

This is z score. Where it falls in the normal distribution? According to the normal distribution, So we fail to reject the null hypothesis.
No statistical significance in the proportion of mortality between IV nitrate patients and control

**where** $\bar{p} = \frac{3+8}{45+50} = 0.116$ , $p_N = \frac{3}{45} = 0.06$ , $p_C = \frac{8}{50} = 0.178$

## Statistical test for p1 – p2

What is the corresponding probability value?

Two Population Proportions, Independent Samples

$$Z = \frac{(0.06-0.178)}{\sqrt{0.116(1-.116)\left(\frac{1}{50}+\frac{1}{45}\right)}} = -1.79$$

Two-tail test:

$H_0$: $p_N - p_C = 0$
$H_1$: $p_N - p_C \neq 0$

Since -1.79 is > than -1.96, we fail to reject the null hypothesis.

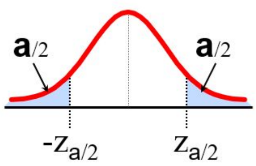
But what is the actual *p*-value?

P (Z<-1.79) + P (Z>1.79)= ?

$Z_{\alpha/2} = 1.96$

Reject $H_0$ if $Z < -Z_{a/2}$

or $Z > Z_{a/2}$



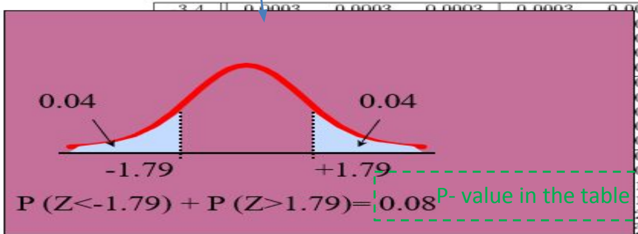




Table 1: Table of the Standard Normal Cumulative Distribution Function Φ(z)

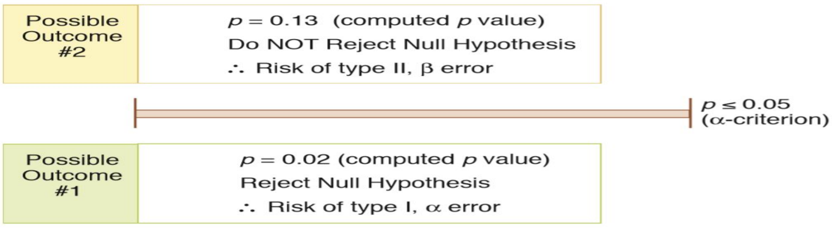
| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| -3.4 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0002 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0003 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0006 | 0.0006 | 0.0005 | 0.0005 | 0.0005 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0008 | 0.0008 | 0.0008 | 0.0007 | 0.0007 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0011 | 0.0011 | 0.0011 | 0.0010 | 0.0010 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0016 | 0.0015 | 0.0015 | 0.0014 | 0.0014 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0022 | 0.0021 | 0.0021 | 0.0020 | 0.0019 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0030 | 0.0029 | 0.0028 | 0.0027 | 0.0026 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0040 | 0.0039 | 0.0038 | 0.0037 | 0.0036 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0054 | 0.0052 | 0.0051 | 0.0049 | 0.0048 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0071 | 0.0069 | 0.0068 | 0.0066 | 0.0064 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0094 | 0.0091 | 0.0089 | 0.0087 | 0.0084 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0122 | 0.0119 | 0.0116 | 0.0113 | 0.0110 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0158 | 0.0154 | 0.0150 | 0.0146 | 0.0143 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0202 | 0.0197 | 0.0192 | 0.0188 | 0.0183 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0256 | 0.0250 | 0.0244 | 0.0239 | 0.0233 |
| -1.8 | 0.0359 | 0.0351 | 0.0344 | 0.0336 | 0.0329 | 0.0322 | 0.0314 | 0.0307 | 0.0301 | 0.0294 |
| -1.7 | 0.0446 | 0.0436 | 0.0427 | 0.0418 | 0.0409 | 0.0401 | 0.0392 | 0.0384 | 0.0375 | 0.0367 |
| -1.6 | 0.0548 | 0.0537 | 0.0526 | 0.0516 | 0.0505 | 0.0495 | 0.0485 | 0.0475 | 0.0465 | 0.0455 |
| -1.5 | 0.0668 | 0.0655 | 0.0643 | 0.0630 | 0.0618 | 0.0606 | 0.0594 | 0.0582 | 0.0571 | 0.0559 |
| -1.4 | 0.0808 | 0.0793 | 0.0778 | 0.0764 | 0.0749 | 0.0735 | 0.0721 | 0.0708 | 0.0694 | 0.0681 |
| -1.3 | 0.0968 | 0.0951 | 0.0934 | 0.0918 | 0.0901 | 0.0885 | 0.0869 | 0.0853 | 0.0838 | 0.0823 |
| -1.2 | 0.1151 | 0.1131 | 0.1112 | 0.1093 | 0.1075 | 0.1056 | 0.1038 | 0.1020 | 0.1003 | 0.0985 |
| -1.1 | 0.1357 | 0.1335 | 0.1314 | 0.1292 | 0.1271 | 0.1251 | 0.1230 | 0.1210 | 0.1190 | 0.1170 |
| -1.0 | 0.1587 | 0.1562 | 0.1539 | 0.1515 | 0.1492 | 0.1469 | 0.1446 | 0.1423 | 0.1401 | 0.1379 |
| -0.9 | 0.1841 | 0.1814 | 0.1788 | 0.1762 | 0.1736 | 0.1711 | 0.1685 | 0.1660 | 0.1635 | 0.1611 |
| -0.8 | 0.2119 | 0.2090 | 0.2061 | 0.2033 | 0.2005 | 0.1977 | 0.1949 | 0.1922 | 0.1894 | 0.1867 |
| -0.7 | 0.2420 | 0.2389 | 0.2358 | 0.2327 | 0.2296 | 0.2266 | 0.2236 | 0.2206 | 0.2177 | 0.2148 |
| -0.6 | 0.2743 | 0.2709 | 0.2676 | 0.2643 | 0.2611 | 0.2578 | 0.2546 | 0.2514 | 0.2483 | 0.2451 |
| -0.5 | 0.3085 | 0.3050 | 0.3015 | 0.2981 | 0.2946 | 0.2912 | 0.2877 | 0.2843 | 0.2810 | 0.2776 |

**Conclusion: The difference in mortality in this study was due to chance, why? Because P-value > 0.05 (0.08)**

p-value: to interpret output from a statistical test, focus on the p-value. The term p-value refers to two things.
-In its first sense, the p-value is a standard against which we compare our results.
-In the second sense, the p-value is a result of computation.

i. The computed p-value is compared with the p-value criterion to test statistical significance. If the computed value is less than the criterion, we have achieved statistical significance. In general, the smaller the p the better.
ii. The p-value criterion is traditionally set at $p \le 0.05$. (Assume that these are the criteria if no other value is explicitly specified.) Using this standard: If $p \le 0.05$, reject the null hypothesis (reached statistical significance) - If $p > 0.05$, do not reject the null hypothesis (has not reached statistical significance).

* We never accept the null hypothesis. We either reject it or fail to reject it. Saying we do not have sufficient evidence to reject it is not the same as being able to affirm that it is true.



KAPLAN Behavioral Science and Social Sciences

Useful videos to watch:

- [Understanding the p-value - Statistics Help]
- [Calculate the P-Value in Statistics - Formula to Find the P-Value in Hypothesis Testing]
- Khan academy: [P-values and significance tests]
- [Level of Significance in Hypothesis Testing]

**p-value**

• After calculating a test statistic we convert this to a p-value by comparing its value to distribution of test statistic's under the null hypothesis

• Measure of how likely the test statistic value is under the null hypothesis

p-value ≤ α ⇒ Reject H0 at level α

p-value > α ⇒ Do not reject H0 at level α

**What is a p- value?**

- 'p' stands for probability
  - Tail area probability based on the observed effect
  - Calculated as the probability of an effect as large as or larger than the observed effect (more extreme in the tails of the distribution), assuming null hypothesis is true

- Measures the strength of the evidence against the null hypothesis
  Smaller p- values indicate stronger evidence against the null hypothesis

**Stating the Conclusions of our Results**

- When the p-value is small, we reject the null hypothesis or, equivalently, we accept the alternative hypothesis.
  - "Small" is defined as a p-value ≤ a, where a = acceptable false (+) rate (usually 0.05).
- When the p-value is not small, we conclude that we cannot reject the null hypothesis or, equivalently, there is not enough evidence to reject the null hypothesis.
  - "Not small" is defined as a p-value > a, where a = acceptable false (+) rate (usually 0.05).

**Statistically significant and not statistically singificant**

| Not statistically significant<br>Do not reject Ho | Statistically significant<br>Reject Ho |
|---|---|
| Sample value compatible with Ho<br><br>Sampling variation is an likely explanation of discrepancy between Ho and sample value | Sample value not compatible with Ho<br><br>Sampling variation is an unlikely explanation of discrepancy between Ho and sample value |

**What is a p- value?**

- p ≤0.05 is an arbitrary cut-point
  - Does it make sense to adopt a therapeutic agent because p-value obtained in a RCT was 0.049, and at the same time ignore results of another therapeutic agent because p-value was 0.051?
- Hence **important** to report the exact p-value and not ≤ 0.05 or >0.05

Meaning of the p-value
i. Provides criterion for making decisions about the null hypothesis
ii. Quantifies the chances that a decision to reject the null hypothesis will be wrong
iii. Tells statistical significance, not clinical significance or likelihood of benefit iv. Limits to the p-value: the p-value does NOT tell us – The chance that an individual patient will benefit – The percentage of patients who will benefit – The degree of benefit expected for a given patient

## P-values

| Trial | Number dead / randomized: Intravenous nitrate | Control | Risk Ratio | 95% C.I. | P value |
|---|---|---|---|---|---|
| Chiche | 3/50 | 8/45 | 0.33 | (0.09,1.13) | 0.08 |
| Some evidence against the null hypothesis | | | | | |
| Flaherty | 11/56 | 11/48 | 0.83 | (0.33,2.12) | 0.70 |
| Very weak evidence against the null hypothesis…very likely a chance finding | | | | | |
| Lis | 5/64 | 10/76 | 0.56 | (0.19,1.65) | 0.29 |
| Jugdutt | 24/154 | 44/156 | 0.48 | (0.28, 0.82) | 0.007 |
| Very strong evidence against the null hypothesis…very unlikely to be a chance finding | | | | | |

Very common

Very rare

## Interpreting P values

1-0.33 = 0.67
Less than 1= protection
More than 1= risk

| Trial | Number dead / randomized: Intravenous nitrate | Control | Risk Ratio | 95% C.I. | P value |
|---|---|---|---|---|---|
| Chiche | 3/50 | 8/45 | 0.33 | (0.09,1.13) | 0.08 |
| …8 out of 100 such trials would show a risk reduction of 67% or more extreme just by chance | | | | | |
| Flaherty | 11/56 | 11/48 | 0.83 | (0.33,2.12) | 0.70 |
| …70 out of 100 such trials would show a risk reduction of 17% or more extreme just by chance…very likely a chance finding | | | | | |
| Lis | 5/64 | 10/76 | 0.56 | (0.19,1.65) | 0.29 |
| Jugdutt | 24/154 | 44/156 | 0.48 | (0.28, 0.82) | 0.007 |
| Very unlikely to be a chance finding | | | | | |

In chiche and fiaherty studies, the sample size is almost the same(45,48) but the p value is different why? because risk ratio (outcome of the study) is different.

In lis and jugdutt studies,risk reduction is alomst the same. However, the p value is different why? because sample size is different

## Interpreting P values
## If the null hypothesis were true…

- Size of the p-value is related to the sample size
- Lis and Jugdutt trials are similar in effect (~ 50% reduction in risk)…but Jugdutt trial has a large sample size

- Size of the p-value is related to the effect size or the observed association or difference
- Chiche and Flaherty trials approximately same size, but observed difference greater in the Chiche trial

| Trial | Intravenous nitrate | Control | Risk ratio | 95% confidence interval | *P* value |
|---|---|---|---|---|---|
| Chiche | 3/50 | 8/45 | 0.33 | (0.09, 1.13) | 0.08 |
| Bussman | 4/31 | 12/29 | 0.24 | (0.08, 0.74) | 0.01 |
| Flaherty | 11/56 | 11/48 | 0.83 | (0.33, 2.12) | 0.7 |
| Jaffe | 4/57 | 2/57 | 2.04 | (0.39, 10.71) | 0.4 |
| Lis | 5/64 | 10/77 | 0.56 | (0.19, 1.65) | 0.29 |
| Jugdutt | 12/77 | 44/157 | 0.48 | (0.28, 0.82) | 0.007 |

**P values**

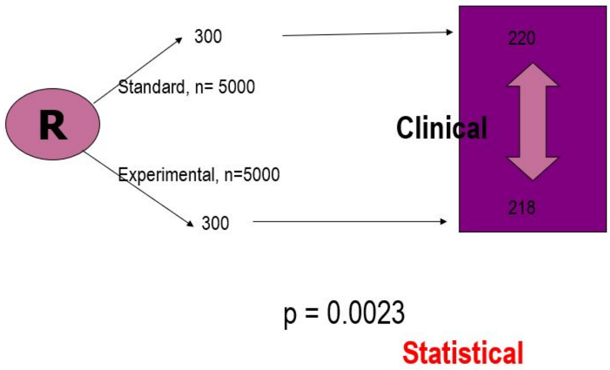
- P values give no indication about the clinical importance of the observed association
- A very large study may result in very small p-value based on a small difference of effect that may not be important when translated into clinical practice
- Therefore, important to look at the effect size and confidence intervals…

Example: If **a new antihypertensive therapy reduced the SBP by 1mmHg as compared to standard therapy we are not interested in swapping to the new therapy.**

**- However, if the decrease was as large as 10 mmHg, then you would be interested in the new therapy.**
**- Thus, it is important to not only consider whether the difference is statistically significant by the possible magnitude of the difference should also be considered.**

## Clinical importance vs. statistical

Cholesterol level, mg\dl




in this example,
you have 5000 sample size in both groups (standard and experimental treatment):
there is only 2 unit difference in clinical statistics, however; p value is highly significant.
so, clinically you will be not be interested .
we got this because of the huge sample size.

| No | Yes | |
|---|---|---|
| 10 | 0 | Standard |
| 7 | 3 | New |

**Clinical**: Absolute risk reduction = 30%
**Statistical**: Fischer exact test: p= 0.211

in this example, you have only 10 cases in each group:
in the standard treatment there is no improvement.
in the new treatment there are 3 cases that improved.
clinically GOOD but statistically NOT GOOD.
why? sample size is small

so, you should know both clinical and statistical significance.

**Reaction of investigator to result of a statistical significance test:**

| Statistical significance | | | |
|---|---|---|---|
| Significant | Not significant | | |
| Annoyed | | Not important | Practical importance of observed effect clinical) significance |
| Elated | Very sad | Important | |