Statistical significance using *p*-value

Dr.Shaikh Shaffi Ahamed, PhD
Professor
Department of Family & Community Medicine
College of Medicine, KSU

Learning Objectives

(1) Able to understand the concepts of statistical inference and statistical significance.

(2) Able to apply the concept of statistical significance(p-value) in analyzing the data.

(3) Able to interpret the concept of statistical significance(p-value) in making valid conclusions.

Investigation

Data Collection

Data Presentation

Tabulation
Diagrams
Graphs

Descriptive Statistics

Measures of Location
Measures of Dispersion
Measures of Skewness & Kurtosis

Inferential Statistics

Estimation Hypothesis Testing
Point estimate
Interval estimate

Inferential statistics

Univariate analysis

Multivariate analysis

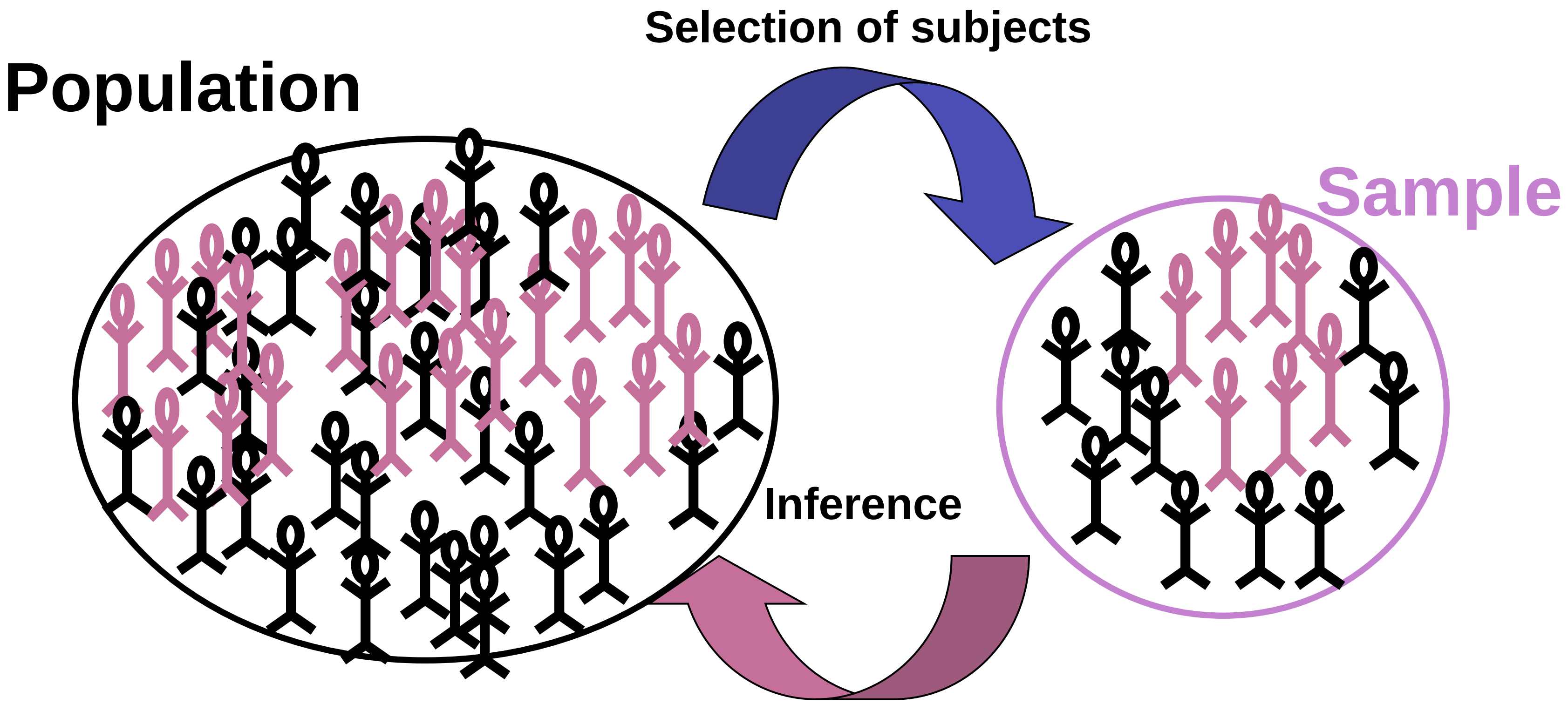
Why use inferential statistics at all?

Average height of all 25-year-old men (population) in KSA is a 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.

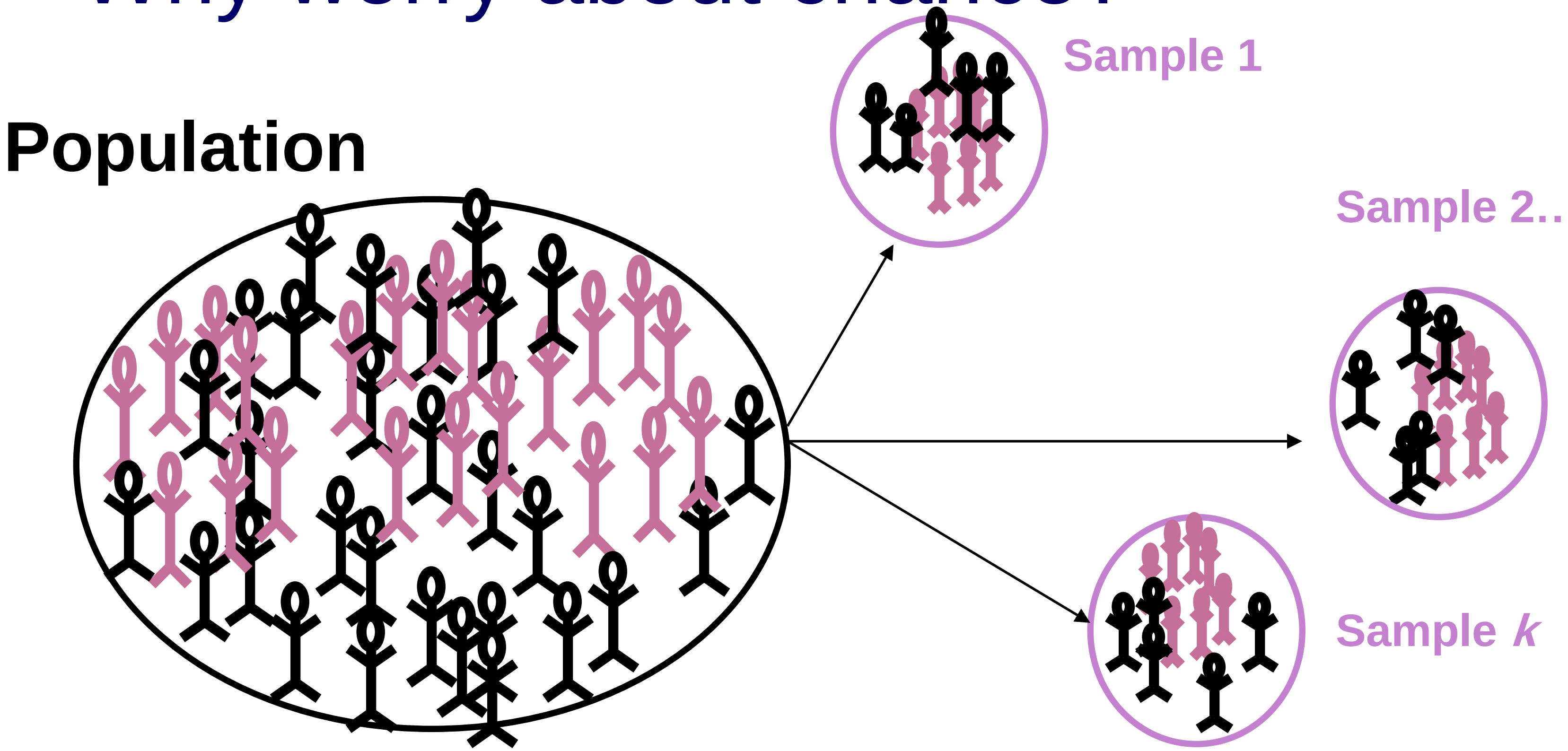
Is risk factor X associated with disease Y?



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?

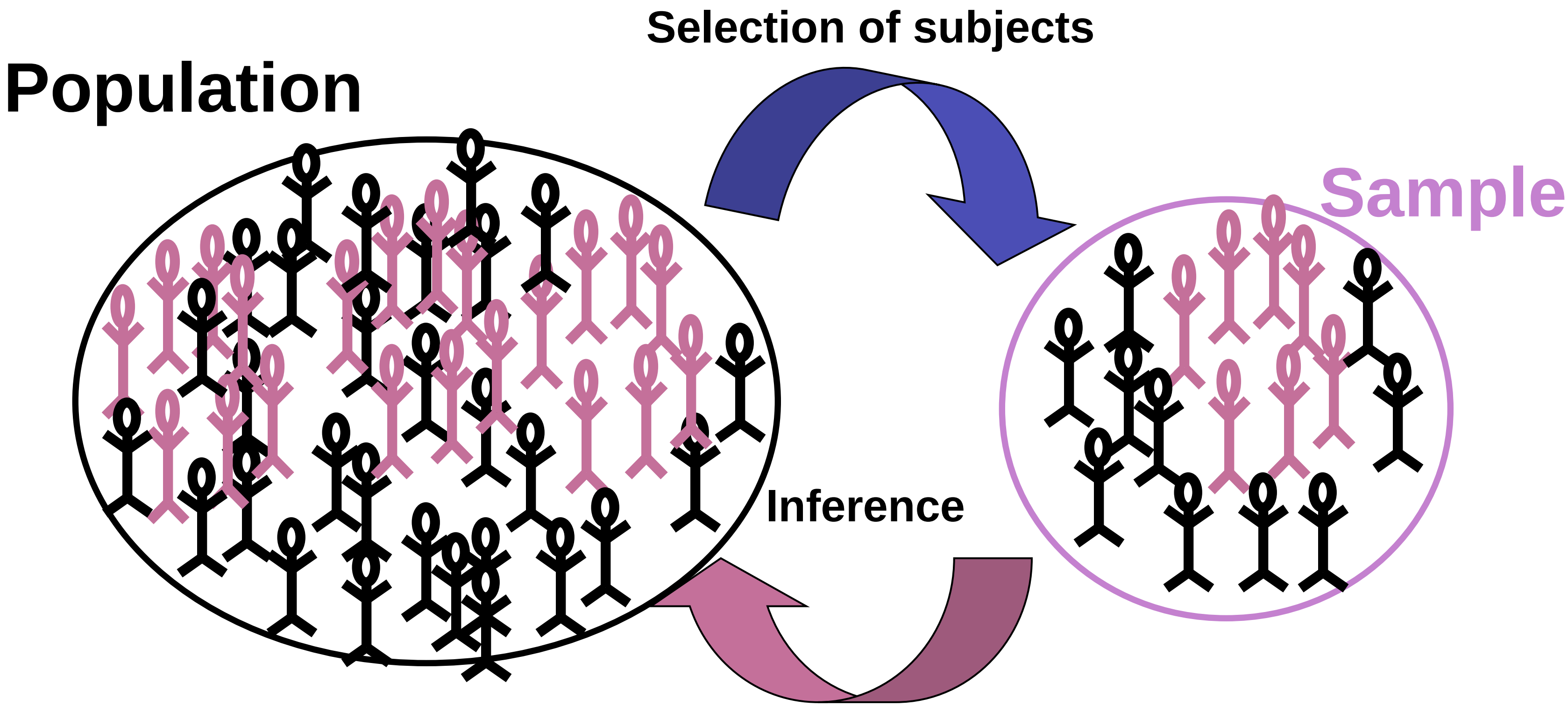
Why worry about chance?



Sampling variability…

- you only get to pick one sample!

Interpreting the results



Make inferences from data collected using laws of probability and statistics

- **tests of significance (p-value)**
- **confidence intervals**

Significance testing

- The interest is generally in comparing two groups (e.g., risk of outcome in the treatment and placebo group)

- The statistical test depends on the type of data and the study design

Hypothesis Testing

- *Null Hypothesis*
  - There is no association between the predictors(associated factors) and outcome variable in the population
  - Assuming there is no association, statistical tests estimate the probability that the association is due to chance
- *Alternate Hypothesis*
  - The proposition that there is an association between the predictors and outcome variable
  - We do not test this directly but accept it by default if the statistical test rejects the null hypothesis

The Null and Alternative Hypothesis

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

The null hypothesis, $H0$

The alternative hypothesis, Ha

:

- 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:

H0: $\mu1 - \mu2 = 0$

HA: $\mu1 - \mu2 > 0$ or HA: $\mu1 - \mu2 < 0$

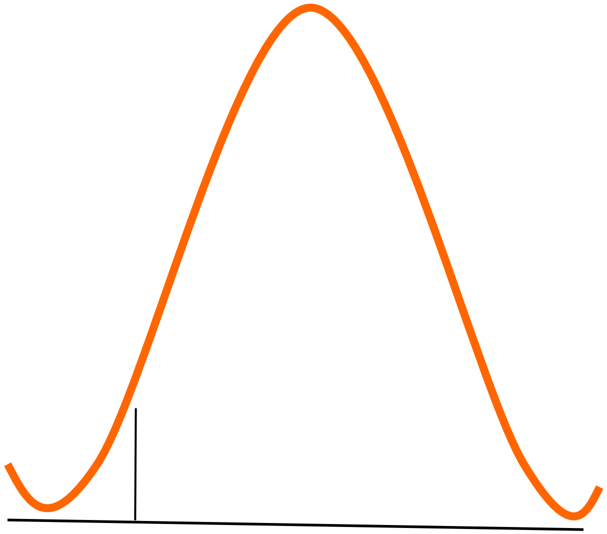
- Two tailed tests are not directional:

H0: $\mu1 - \mu2 = 0$

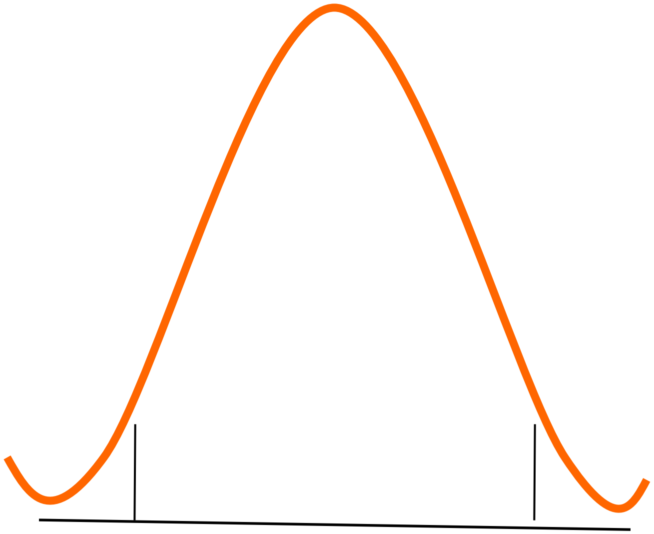
HA: $\mu1 - \mu2 \neq 0$

Level of significance, α: Specified before an experiment to define rejection region

One Sided :   α = 0.05   Two Sided: α/2 = 0.025



Critical Value = -1.64



Critical Values = -1.96 and +1.96

Type-I and Type-II Errors

- α = Probability of rejecting H_0 when H_0 is true
- α is called significance level of the test
- β = Probability of not rejecting H_0 when H_0 is false
- **1-β** is called statistical power of the test

Diagnosis and statistical reasoning

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

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

α : significance level

$1-\beta$: power

Significance testing

Subjects with Acute MI

Mortality IV nitrate P_N	? <	Mortality No nitrate P_C

- 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(H_o)

- There is no association between the independent and dependent/outcome variables
  - Formal basis for hypothesis testing

- In the example, H_o :"The administration of IV nitrate has no effect on mortality in MI patients" <u>or</u> $P_N - P_C = 0$

Obtaining *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
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

How do we get this *p*-value?

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

Example of significance testing

- In the Chiche trial:
  - $p_N = 3/50 = 0.06$; $p_C = 8/45 = 0.178$

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

- Statistical test:
  - Two-sample proportion

Test statistic for Two Population Proportions

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

Observed difference

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

0

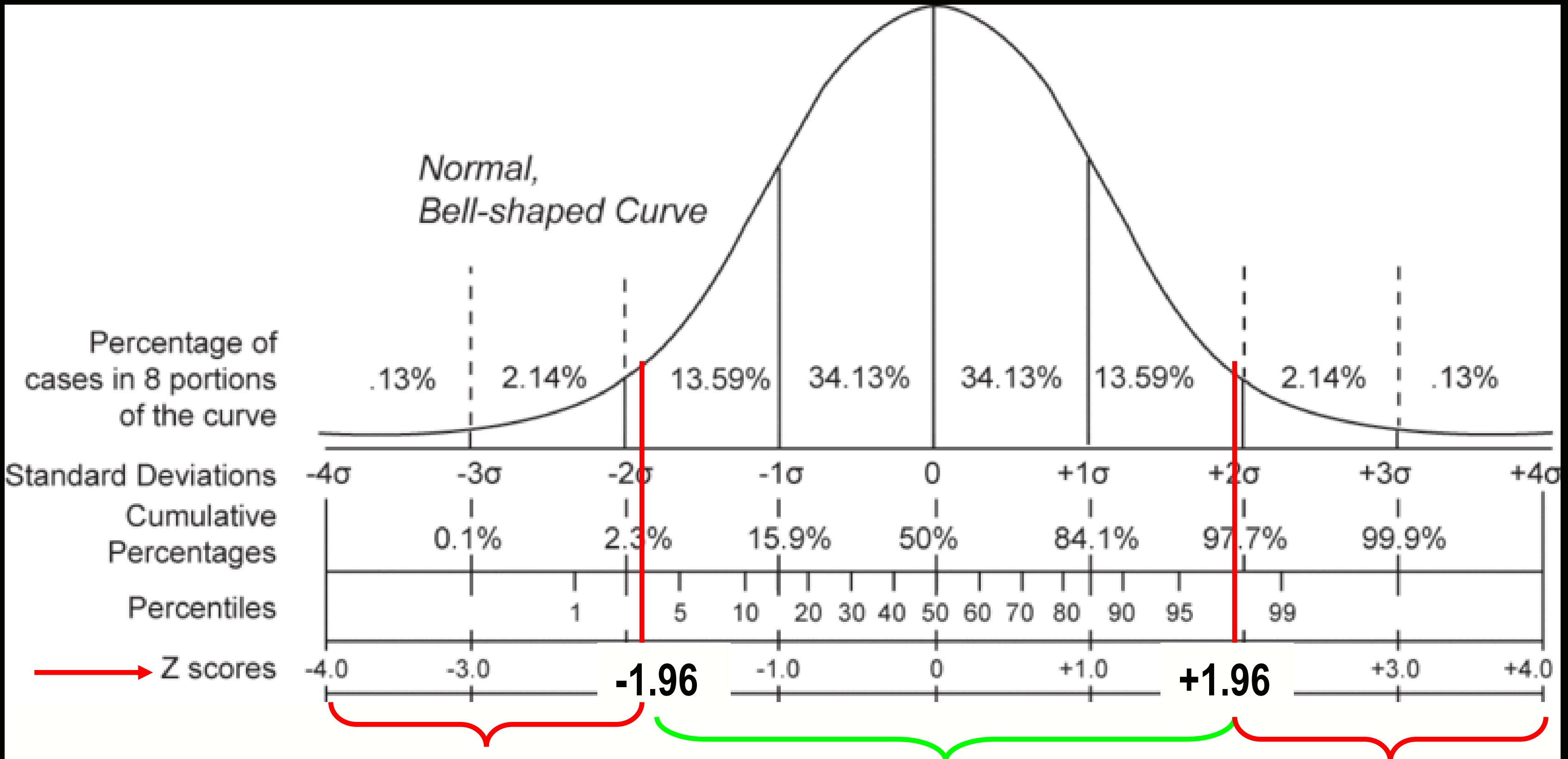
Null hypothesis

No. of subjects in IV nitrate group

No. of subjects in control group

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

Testing significance at 0.05 level



Rejection region

Nonrejection region

Rejection region

$Z\alpha/2 = 1.96$

Reject H_0 if $Z < -Z\ \alpha/2$ or $Z > Z\ \alpha/2$

Two Population Proportions

(continued)

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

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

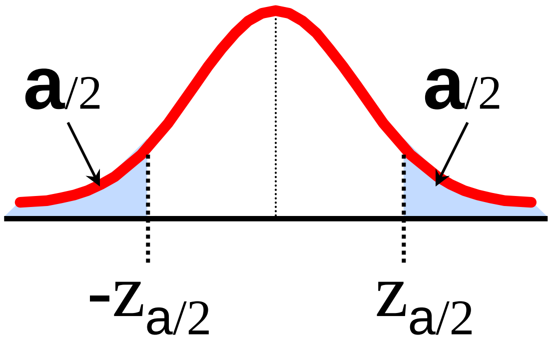
Statistical test for $p_1 - p_2$

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.

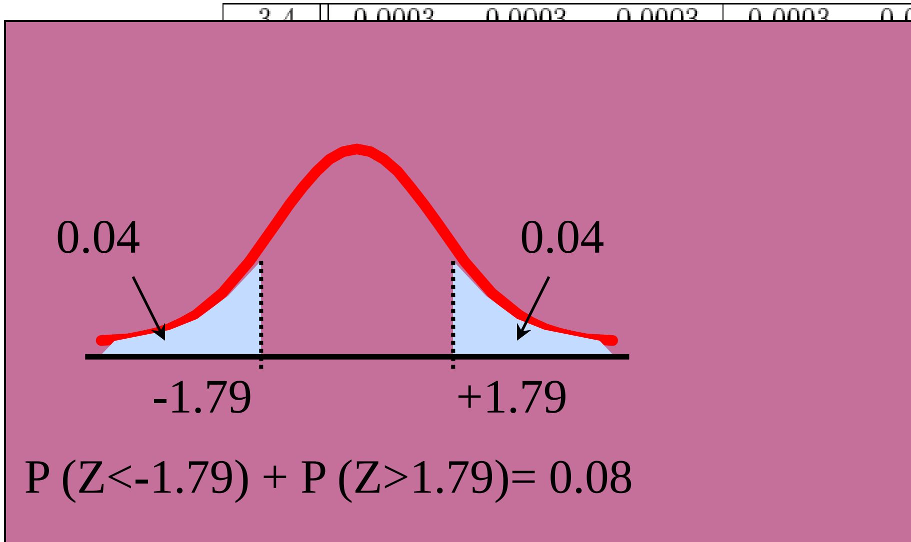
But what is the actual *p*-value?

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 $\Phi(z)$



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
					004	0.0004	0.0004	0.0004	0.0004	0.0003
					006	0.0006	0.0006	0.0005	0.0005	0.0005
					008	0.0008	0.0008	0.0008	0.0007	0.0007
					012	0.0011	0.0011	0.0011	0.0010	0.0010
					016	0.0016	0.0015	0.0015	0.0014	0.0014
					023	0.0022	0.0021	0.0021	0.0020	0.0019
					031	0.0030	0.0029	0.0028	0.0027	0.0026
					041	0.0040	0.0039	0.0038	0.0037	0.0036
					055	0.0054	0.0052	0.0051	0.0049	0.0048
					073	0.0071	0.0069	0.0068	0.0066	0.0064
					096	0.0094	0.0091	0.0089	0.0087	0.0084
					125	0.0122	0.0119	0.0116	0.0113	0.0110
					162	0.0158	0.0154	0.0150	0.0146	0.0143
					207	0.0202	0.0197	0.0192	0.0188	0.0183
					262	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

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 $\leq \alpha \Rightarrow$ Reject H_0 at level α

p-value $> \alpha \Rightarrow$ Do not reject H_0 at level α

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 $\leq \alpha$, where $\alpha =$ 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 > α, where α = acceptable false (+) rate (usually 0.05).

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

STATISTICALLY SIGNIFICANT AND NOT STATISTICALLY SINGIFICANT

- Statistically significant

  Reject Ho

  Sample value not compatible with Ho

  Sampling variation is an unlikely explanation of discrepancy between Ho and sample value

- Not statistically significant

  Do not reject Ho

  Sample value compatible with Ho

  Sampling variation is an likely explanation of discrepancy between Ho and sample value

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
Flaherty	11/56	11/48	0.83	(0.33,2.12)	0.70
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

Some evidence against the null hypothesis

Very weak evidence against the null hypothesis…very likely a chance finding

Very strong evidence against the null hypothesis…very unlikely to be a chance finding

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

Trial	Number dead / randomized: Intravenous nitrate	Number dead / randomized: 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					

Interpreting *P* values

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

- 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

Interpreting *P* values

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

- 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

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

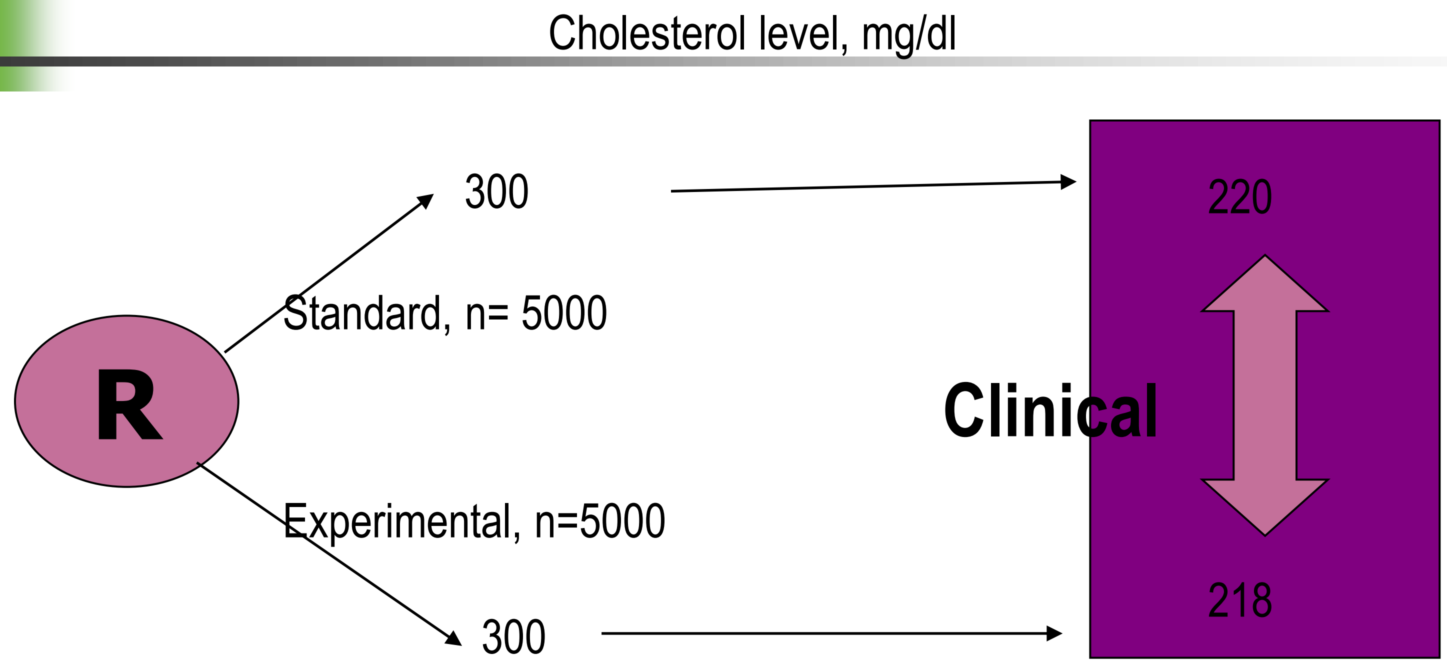
- Therefore, important to look at the effect size and confidence intervals…

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 significance



p = 0.0023

Statistical

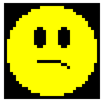
Clinical importance vs. statistical significance

	Yes	No
Standard	0	10
New	3	7

Absolute risk reduction = 30% ← **Clinical**

Fischer exact test: p = 0.211 ← **Statistical**

Reaction of investigator to results of a statistical significance test

		Statistical significance	
		Not significant	**Significant**
Practical importance of observed effect	**Not important**		**Annoyed** 
	Important	**Very sad** 	**Elated**