Universidad del Cauca

ERASMUS UNIVERSITEIT ROTTERDAM

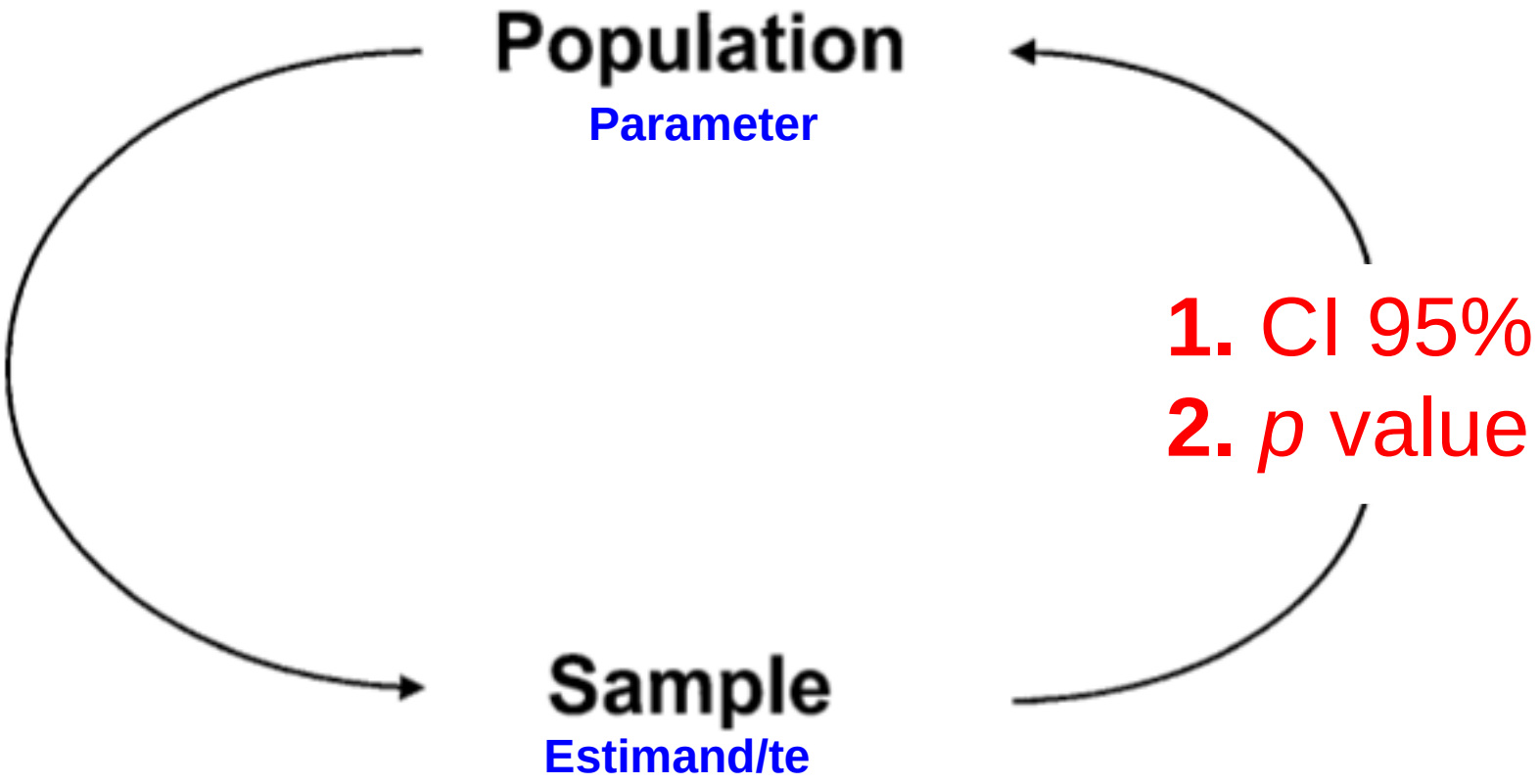
ANALYSIS OF CATEGORICAL OUTCOMES

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

CI95% for a proportion

- A binary variable X of two categories: 1 / 0
- For the sample the proportion is p
- Useful if n is large

$$\sigma_{(p)} = \sqrt{\pi(1-\pi)}$$

$$DE_{(p)} = \sqrt{p(1-p)} \qquad se_{(p)} = \sqrt{\frac{p(1-p)}{n}}$$

$$CI95\% = p \pm 1.96\sqrt{\frac{p(1-p)}{n}}$$

Example,

- To estimate the prevalence of obesity in faculty students
  - n = 200
  - p = 30% = 0,3

$$DE(p) = \sqrt{p(1-p)} = \sqrt{0{,}3(1-0{,}3)} = 0{,}458$$

$$SE(p) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0{,}3(0{,}7)}{200}} = 0{,}032$$

$$CI95\% = p \pm 1.96\left(SE(p)\right) = 0{,}3 \pm 1{,}96(0{,}032)$$

$$CI95\% = 0{,}3\ [0{,}237 - 0{,}362] = [23{,}7\% - 36{,}2\%]$$

Comparing proportions

- Pediatric AIDS Clinical Trial Group (ACTG)
  - “We conducted a randomized, double-blinded, placebo- controlled trial of the efficacy and safety of zidovudine (AZT) in reducing the risk of maternal-infant HIV transmission”
  - 363 pregnant women with HIV were randomized to AZT or placebo (18 postnatal months)

    - 180 AZT HIV+ 13 children
    - 183 Placebo HIV+ 40 children

Conner, E., et al. (1994). Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment, *New England Journal of Medicine* 331: 18.

Risk of transmission (CI 18 months)

	HIV +	HIV -	Total
AZT	13	167	**180**
Placebo	40	143	**183**
Total	**53**	**310**	**363**

CI_{AZT} = 13/180 = 7%

CI_{PLAC} = 40/183 = 22%

Cumulative incidence and CI 95%

- CI_{AZT} = 13/180 = 7% CI95% [3.9 - 12]
- CI_{PLAC} = 40/183 = 22% CI95% [16 - 28]

- 95% confidence intervals are calculated following the binomial distribution (if n is large it approaches normal)

The question is: Is this difference significant or can it be explained by chance (random error)?

Chi^2 test

	HIV +	HIV -	Total
AZT	13	167	**180**
Placebo	40	143	**183**
Total	**53**	**310**	**363**

What would they be then?
the "expected" values
if the H0 were true?

- Test that can be extended to two or more proportions
- It is based on,
  - "Expected" values and "observed" values
  - Expected: If H0 were true
  - Is there a relationship between columns and rows?

$$\chi^2 = \sum_{4\ cells} \frac{(0 - E)^2}{E}$$

"Observed" vs "expected" if H0 were true

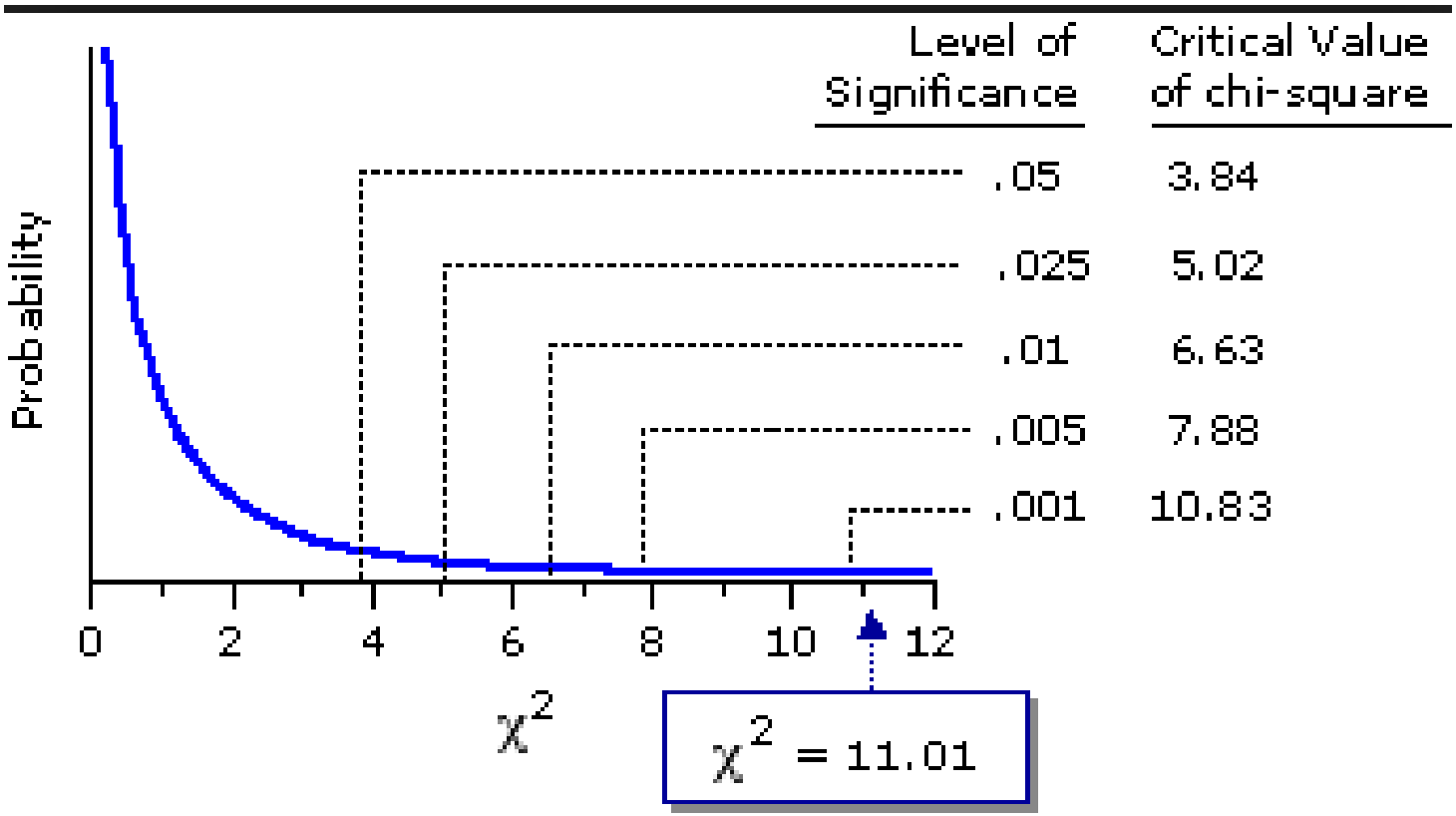
	HIV +	HIV -	Total
AZT	13	167	**180**
Placebo	40	143	**183**
Total	**53**	**310**	**363**

$$\chi^2 = \sum_{4\ cells} \frac{(0 - E)^2}{E}$$

"expected"
if the H0 were true

	HIV +	HIV -	Total
AZT	**26**	**154**	180
Placebo	**27**	**156**	183
Total	53	310	363

P value for Chi^2 test



- CI_{AZT} = 13/180 = 7% CI95% [3.9 - 12]
- CI_{PLAC} = 40/183 = 22% CI95% [16 - 28]

- P value for Chi^2 test = 0.0001

- There is sufficient evidence to reject H0 (equality)
- I accept H1 : There are “statistically significant” differences between AZT and Placebo

Example 2,

- Food poisoning occurred at a certain site. Some believe that the origin was chicken sandwiches

		Food poisoning		
Exposure		**Yes**	**No**	**Total**
	Sandwich	a	b	**a+b**
	Other food	c	d	**c+d**
	Total	**a+c**	**b+d**	**a+b+c+d**

$H_0: \quad P_{sand} = P_{otro}$

$H_1: \quad P_{sand} \neq P_{otro}$

Food poisoning

Exposure		Si	No	Total
	Sandwich	109	116	**225**
	Otro alimento	4	34	**38**
	Total	**113**	**150**	**263**

$p_{sand} = 109/225 = 48\%$

$p_{otro} = 4/38 = 10\%$

P value for Chi^2 test = 0.001

Fisher's exact test

- The chi2 test is based on a normal approximation
- Works well if the samples are large

- Solution,
  - Fisher's exact test
  - Difficult manual calculation
  - It has no approximations and delivers an "exact" p-value
  - Useful in small samples

Fisher's exact test

- 65 pregnant patients – classified as high risk for hypertension in pregnancy – are randomized in the third trimester to Aspirin 100 mg/day versus placebo to assess the effect on the occurrence of hypertension

	HT +	Non HT	Total
ASA 100	4	30	**34**
Placebo	11	20	**31**
Total	**15**	**50**	**65**

CI_{ASA} = 4/34 = 12%

CI_{PLAC} = 11/31 = 35%

Schiff, E. et al. The use of aspirin to prevent pregnancy-induced hypertension and lower the ratio of thromboxane A2 to prostacyclin in relatively high risk pregnancies. *New England Journal of Medicine,* 321, 6.

Fisher's exact test

	HT +	Non HT	Total
ASA 100	4	30	**34**
Placebo	11	20	**31**
Total	**15**	**50**	**65**

CI_{ASA} = 4/34 = 12%

CI_{PLAC} = 11/31 = 35%

- Fisher *p value* = 0.0378
- Chi^2 *p value* = 0.0234

Both tests reject H0

Test for variables with more than two categories

	Stratum 1	Stratum 2	Stratum 3	Total
Male	51	95	38	**184**
Female	17	101	75	**193**
Total	**68**	**196**	**113**	**377**

- The Chi2 test can be extended to variables with more than two categories
  - H0: Variables are independent
  - H1: Variables are related (or depend in some category on one another)

Test for variables with more than two categories

	Stratum 1	Stratum 2	Stratum 3	Total
Male	51	95	38	**184**
Female	17	101	75	**193**
Total	**68**	**196**	**113**	**377**

- Chi^2 p value = 0.000

- I have enough evidence to reject the H0 of independence
- They are significantly related