Biostatistics

Th3

Inferential Statistics
Qualitative Data

Carina Silva
(*carina.silva@estesl.ipl.pt*)
Mestrado em Farmácia

Contents

Introduction

- Godness-of-fit. (one variable)
- Independence (or association) (two variables).
- Homogeneity.

Introduction

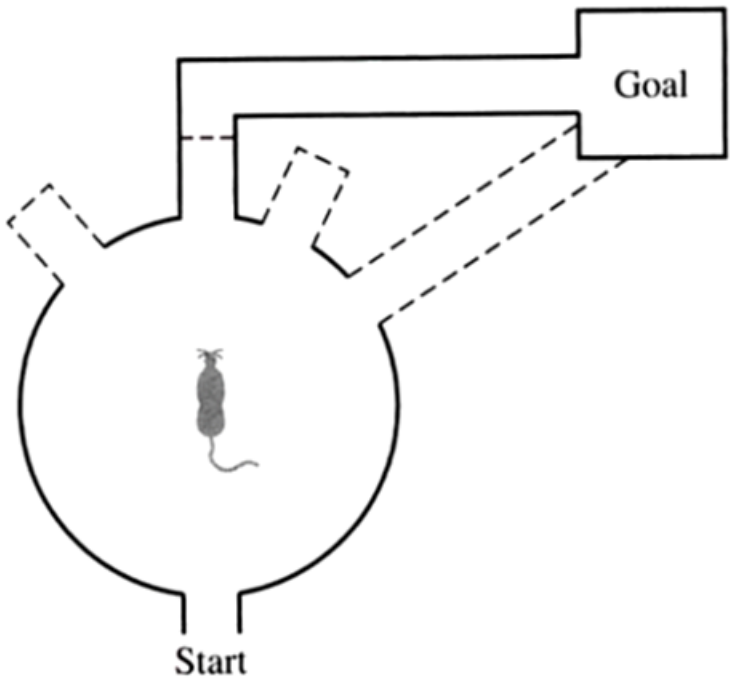
- Non-parametric test.
- Qualitative data.
- Evaluates whether observed frequencies for a qualitative variable (or variables) are adequately described by hypothesized or expected frequencies.

Goodness-of-fit

- Asks whether the relative frequencies observed in the categories of a sample frequency distribution are in agreement with the relative frequencies hypothesized to be true in the population.

Goodness-of-fit

- **Observed Frequency**

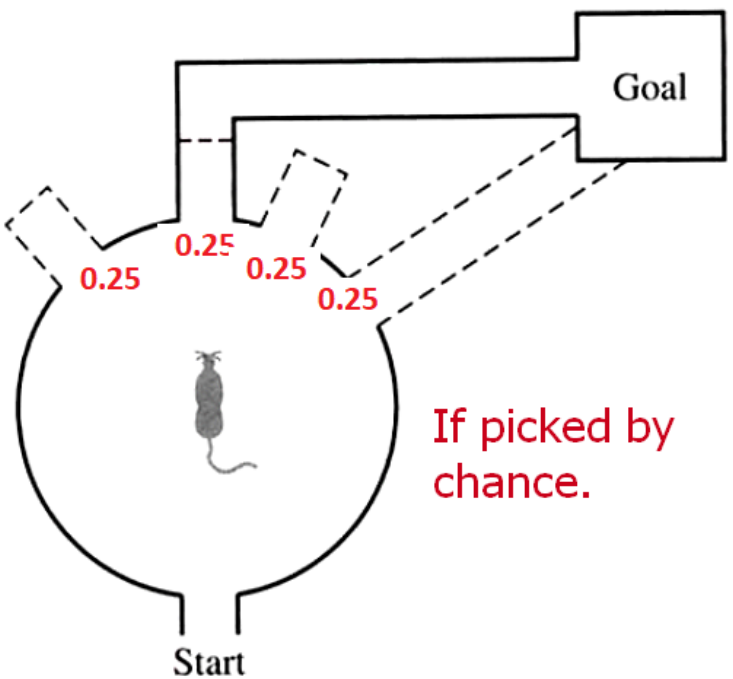


	A	B	C	D
Observed	4	5	8	15

- 1- State Hypotheses
- 2- Calculate a statistic, based on your sample data.
- 3- Create a distribution of this statistic, as it would be observed if the null hypothesis were true.
- 4- Measure how extreme your test statistic from (2) is, as compared to the distribution generated in (3).

Hypothesis

- Let denote p_i the proportion in the i^{th} category.
- H_0 : All p_i are the same.
- At least one p_i differs from the others.



Expected frequency

- The hypothesized frequency for each distribution, given the null hypothesis is true.
- The expected counts are the expected counts if the null hypothesis were true.
- For each cell, the expected count is the sample size (n) times the null proportion, p_i.

	Alley Chosen			
	A	B	C	D
Observed	4	5	8	15
Expected	8	8	8	8

Chi-square statistics

- The chi-square statistic is a test statistic for categorical variables:

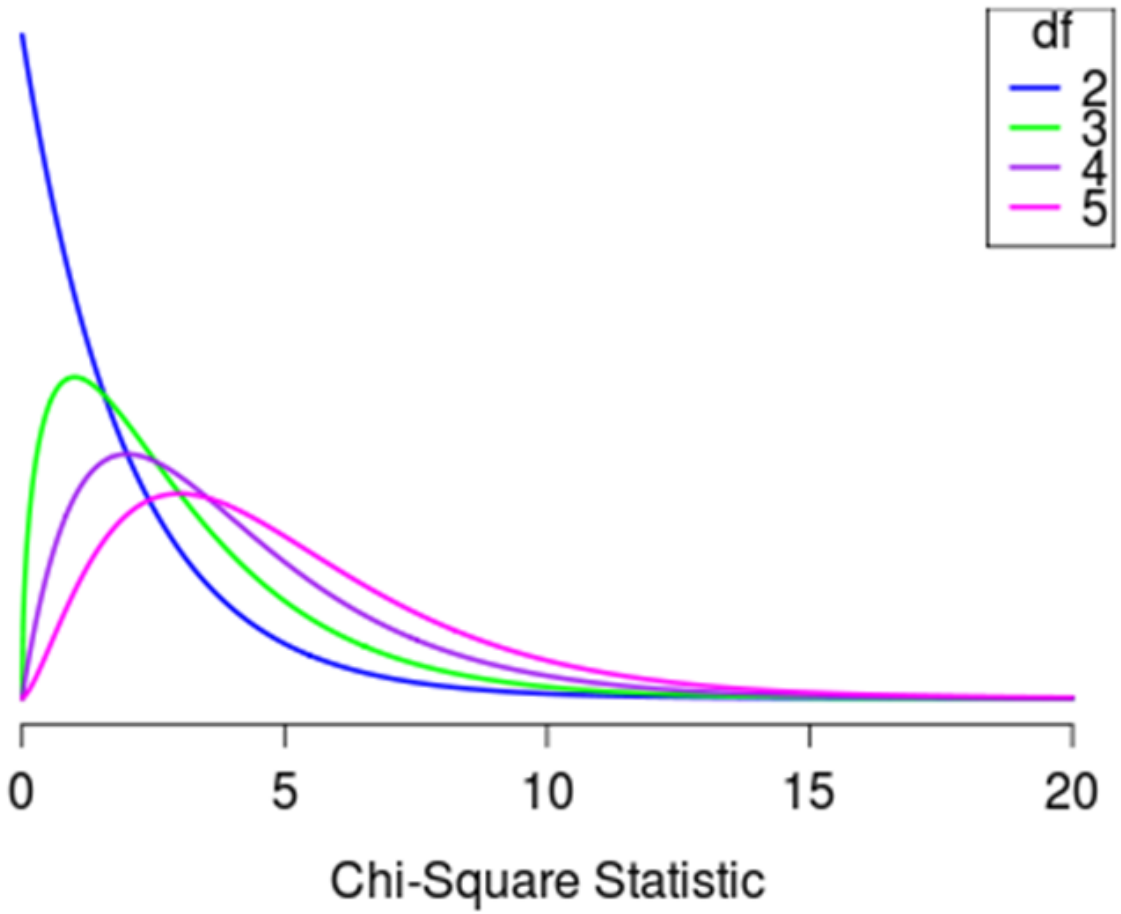
$$\chi_0^2 = \sum \frac{(Observed - expected)^2}{expected}$$

- We have a test statistic. What else do we need to perform the hypothesis test?
- **A distribution of the test statistic assuming H_0 is true.**
- To calculate the p-value for a chi-square test, we always look in the upper tail.
  - Values of the χ^2 are always positive.
  - The higher the χ_0^2 statistic value is, the farther the observed counts are from the expected counts, and the stronger the evidence against the null.

Chi-Square Distribution

- If each of the expected counts are at least 5, AND if the null hypothesis is true, then the χ^2 statistic follows a χ^2 distribution, with degrees of freedom equal to **df = number of categories-1**.
- The null hypothesized proportions for each category do not have to be the same.

Chi-Square Distribution



Decision Rule

- If $\chi_0^2 \geq \chi^2_{(df;1-\alpha)}$ it would reject H_0.

TABELA IV

Distribuição do Qui-Quadrado - χ^2_n

Os valores tabelados correspondem aos pontos x tais que: $P(\chi^2_n \leq x)$

	$P(\chi^2_n \leq x)$													
n	0,005	0,01	0,025	0,05	0,1	0,25	0,5	0,75	0,9	0,95	0,975	0,99	0,995	
1	3,93E-05	0,000157	0,000982	0,003932	0,016	0,102	0,455	1,323	2,706	3,841	5,024	6,635	7,879	1
2	0,010	0,020	0,051	0,103	0,211	0,575	1,386	2,773	4,605	5,991	7,378	9,210	10,597	2
3	0,072	0,115	0,216	0,352	0,584	1,213	2,366	4,108	6,251	7,815	9,348	11,345	12,838	3
4	0,207	0,297	0,484	0,711	1,064	1,923	3,357	5,385	7,779	9,488	11,143	13,277	14,860	4
5	0,412	0,554	0,831	1,145	1,610	2,675	4,351	6,626	9,236	11,070	12,832	15,086	16,750	5
6	0,676	0,872	1,237	1,635	2,204	3,455	5,348	7,841	10,645	12,592	14,449	16,812	18,548	6
7	0,989	1,239	1,690	2,167	2,833	4,255	6,346	9,037	12,017	14,067	16,013	18,475	20,278	7
8	1,344	1,647	2,180	2,733	3,490	5,071	7,344	10,219	13,362	15,507	17,535	20,090	21,955	8

Chi-square statistic value

	Alley Chosen			
	A	B	C	D
Observed	4	5	8	15
Expected	8	8	8	8

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$
$$= \frac{(4-8)^2}{8} + \frac{(5-8)^2}{8} + \frac{(8-8)^2}{8} + \frac{(15-8)^2}{8}$$
$$= 9.25$$

Decision:

Conclusion:

Go to ▸ Link and do exercises 1 and 2 from folder qualitative data.

Independence

- Asks whether observed frequencies reflect the independence of two qualitative variables.
- Compares the actual observed frequencies of some phenomenon (in our sample) with the frequencies we would expect if there were no relationship at all between the two variables in the larger (sampled) population.
- Two variables are independent if knowledge of the value of one variable provides no information about the value of another variable.

Example: Recent studies have found that most teens are knowledgeable about AIDS, yet many continue to practice high-risk sexual behaviors. King and Anderson (1993) asked young people the following question: "If you could have sexual relations with any and all partners of your choosing, as often as you wished, for the next 2 (or 10) years, but at the end of that time period you would die of AIDS, would you make this choice?" A five-point Likert scale was used to assess the subjects' responses. For the following data, the responses "probably no," "unsure" "probably yes", and "definitely ye" were pooled into the category "other." Using the .05 level of significance, test for independence.

	Definitely No	Other
Males	451	165
Females	509	118

Hypothesis

- H_0 : The response to the question and gender are independent.
- H_1 : The response to the question and gender are not independent.

Chi-Square Test for Independence

- Calculate the expected counts for each cell:

$$\text{expected count} = \frac{\text{row total} \times \text{column total}}{n}$$

- Calculate the qui-square statistic value:

$$\chi_0^2 = \sum \frac{(observed - expected)^2}{expected}$$

- Compute the p-value as the area in the tail above the χ^2 statistic using either a randomization distribution, or a χ^2 distribution with $df = (r-1) \times (c-1)$ if all expected counts >5.
- Interpret the p-value in context.

Go to ▸ Link and do exercises 3 and 4 from folder qualitative data.

Fisher's exact test

The tests for association described previously all assume that the samples are sufficiently large so that the estimators have sampling distributions that are approximately normal. However, in many instances studies are based on small samples. This may arise due to cost or ethical reasons. A test due to R.A. Fisher, Fisher's exact test, was developed for this particular situation.

Go to ▸ Link and do exercise 5 from folder qualitative data.