

Comparing categorical variables in clinical and experimental studies

Comparação entre variáveis categóricas em estudos clínicos e experimentais

Anna Carolina Miola¹, Hélio Amante Miot¹ 

How to cite: Miola AC, Miot HA. Comparing categorical variables in clinical and experimental studies. *J Vasc Bras.* 2022;21:e20210225. <https://doi.org/10.1590/1677-5449.20210225>

Many studies of a quantitative nature use qualitative variables, within both the biomedical sciences and the social sciences. These variables are also known as categorical and their magnitude is expressed in terms of the frequency with which each of their categories occurs. Qualitative variables can be subdivided into dichotomous (for example, sex, death, cure), ordinal (for example, cancer staging, pulse amplitude, functional class, phototype, anesthetic risk) or polytomous/multinomial (for example, sexual orientation, ABO type, marital status, religion, race, aneurysm type, type of chronic ulcer).¹⁻³

When qualitative variables are employed, the phenomenon measured can be represented as the percentage of occurrence in each category, and subgroups should be compared in terms of the proportion of the sample that is attributed to each class.³ There is an extensive literature on techniques for statistical analysis of qualitative variables;⁴⁻⁶ whereas this text will deal with comparison of proportions between categorical variables. Comparative analysis of proportions between subgroups employs different concepts from parametric statistics, providing lower statistical power (larger type II error) in analogous situations, such as when a quantitative variable (for example, age) is categorized (for example, < 30 years, 30–59 years, ≥ 60 years).^{7,8}

According to frequentist statistics,⁹ the probability of a proportion of events selected at random, without replacement of cases, can be generalized from the chi-square distribution, while Pearson's chi-square test is based on the difference between the frequencies observed and the frequencies ideally expected for each category and can be used to compare how well a sample fits a known distribution (for example, for

comparison with the literature) or independence between different samples.¹⁰ Despite the popularity of Pearson's chi-square test, other methods such as the G test (likelihood ratio) and the Goodman test (contrasts between proportions) are also used to compare proportions. However, absolute superiority between them has not yet been systematically defined.¹¹⁻¹⁴

An observed proportion's fit can be compared to a description from the literature or a theoretical prediction (for example, expression of a phenotype according to segregation of a gene).¹⁵ For example, Tamega et al.¹⁶ studied ABO and Rh blood typing of 69 patients with lupus erythematosus, comparing them against the expected frequencies of these categories among blood donors at the institution. Pearson's chi-square test (of fit) returned a *p*-value 0.081 for ABO types and a *p*-value of 0.721 for Rh types, accepting the hypothesis that these blood type classes did not differ from what was expected in the local population.¹⁶

In clinical-epidemiological research, it is highly usual to present an initial descriptive table containing demographic data on subgroups, in order to demonstrate their homogeneity. For example, Amiri et al.¹⁷ included 110 cases and 110 controls in a cross-sectional study to test associations between anthropometric indices and type 2 diabetes mellitus. Of these diabetic patients, 75 (51%) were female, whereas there were only 72 women (49%) in the control group. In this sample, the difference in proportion between the groups (2%) was not considered significant (*p*-value = 0.668) for this dichotomous variable according to Pearson's chi-square test (of independence).

While versatile, Pearson's chi-square test has inadequate performance (larger type I error) in smaller samples ($n \leq 40$), especially in which > 20% of

¹Universidade Estadual Paulista "Júlio de Mesquita Filho" – UNESP, Faculdade de Ciências Médicas e Biológicas de Botucatu, Botucatu, SP, Brasil.

Financial support: None.

Conflicts of interest: No conflicts of interest declared concerning the publication of this article.

Submitted: December 11, 2021. Accepted: January 20, 2022.

The study was carried out at Departamento de Dermatologia, Faculdade de Ciências Médicas e Biológicas de Botucatu, Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP), Botucatu, SP, Brazil.



Copyright© 2022 The authors. This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

expected values are ≤ 5 , which is relatively common in biomedical research scenarios. Several procedures are recommended in this situation, ranging from combining categories to increase the predicted value (for example, dichotomizing skin colors as white vs. not white, combining less common B blood groups with AB) or using other statistical tests.

There is an intense academic debate about which analytical strategies should be adopted for situations in which Pearson's chi-square test is contraindicated, while different tests for categorical variables can behave differently depending on the manner in which the data are collected (randomized or not), since a large proportion of studies do not employ a completely randomized sampling structure.¹⁸⁻²⁰ The Barnard and Boschloo exact tests are two examples that correct for these limitations for 2×2 contingency tables.^{21,22} In turn, the G test (with Williams' correction) can be used for multinomial comparisons in situations in which Pearson's chi-square test is contraindicated.^{21,23} Estimates of the (exact) p -value using resampling (bootstrapping) or Monte Carlo simulation are also effective in cases with modest samples or subgroups with a low predicted rate of occurrence.^{19,24}

Fisher's exact test is cited in many texts as a solution for cases in which Pearson's chi-square test is not indicated, but it inflates the type II error, in addition to being based on a conditional probability model, which contrasts with what is usually proposed in biomedical research (variable marginal totals).^{25,26} Along the same lines, correcting Pearson's chi-square test with Yates' procedure is excessively conservative in 2×2 tables. Use and interpretation of these tests should be parsimonious when they return p -values close to the significance level.^{22,24}

For more complex designs, involving interaction between more than two categorical variables or multivariate adjustments in which the dependent variable is categorical, other methods of analysis can be used, such as Poisson regression (log-linear), logistic regression, and multinomial regression, which, as with Pearson's chi-square test, are penalized in cases

with low frequencies in subgroups. On the other hand, multivariate methods, such as multiple correspondence analysis, are unaffected by the contingencies of tests of hypotheses and can support exploratory analyses of categorical data.^{4,27,28} Meanwhile, the problems linked to analysis of ordinal data and calculation of sample sizes for studies involving proportions have been covered previously.^{2,29-32}

When the results of comparisons of multinomial variables are significant, it remains to be determined which of the internal proportions diverge from the expected, since the test result (for Pearson's chi-square test, for example) refers to the overall behavior of the proportions, so it is necessary to proceed to a *post hoc* analysis of the subcategories. Analysis of the residuals of the contingency table (standardized and adjusted) is a widely-used strategy that returns a Z statistic (Z_{res}) for each proportion found, enabling multiple comparison between them to identify which specific variables most contribute to the result observed in the global test.³³ By analyzing the residuals shown in Table 1, it can be concluded that cancer patients referred from clinics exhibited more incidental tomographic diagnoses of pulmonary thromboembolism than those admitted via the emergency room, however, no differences were found in the proportions from inpatients and those from ICU.³⁴

Another option for analysis of subgroups is Goodman and Kruskal's λ test, which is a measure of the proportional reduction in error in the contingency table analysis for multinomial data, indicating the point to which modal categories and frequencies for each value of the independent variable differ from the values of the independent variable.³⁵ In the same manner, the table can be partitioned into 2×2 subtables. However, the multiple comparisons must be adjusted to control inflation of the type I error, using the Bonferroni procedure, for example.²⁰

Epidemiological research often employs dichotomous outcomes (for example, cure, death, sickness) to compare two or more groups (for example, placebo vs. treatment). Characteristics intrinsic to the designs of studies have led to a growing tendency for

Table 1. Analysis of residuals in data from Carneiro et al.³⁴ on the origin of cancer patients with pulmonary thromboembolism (PTE) on computed tomography of the thorax, when the finding was incidental or there was a prior suspicion.

Origin	No suspicion of PTE		PTE suspected previously	
	(n = 48)	Zres (p-value)	(n = 60)	Zres (p-value)
Clinic	28 (59%)	+5.1 (<0.001)	7 (12%)	-5.1 (<0.001)
Wards	16 (33%)	-0.2 (0.856)	21 (35%)	+0.2 (0.856)
Intensive care unit	2 (4%)	-1.4 (0.161)	7 (12%)	+1.4 (0.161)
Emergency room	2 (4%)	-4.5 (<0.001)	25 (43%)	+4.5 (<0.001)

p -value (global) < 0.001; Pearson's chi-square.

comparisons of these proportions to be estimated from their epidemiological measures of effect, such as odds ratios, relative risk, or prevalence ratios, rather than merely according to the results of statistical tests of proportion.^{36,37} Both the *p*-value and the confidence interval of such associations can be calculated directly for these estimates using logistic, ordinal, multinomial, or Poisson regression models.³⁸

The need to adjust results for covariates that are of importance in the causal model (for example, age, sex, smoking) has demanded wider adoption of these regression techniques for analysis of categorical data. Contingencies in the presence of modest samples or rarity of events in one of the categories can be overcome using bootstrapping techniques, resampling data more than 1,000 times. However, since these methods consider the relationships between subcategories, they do not deal adequately with cases in which one category is zero, in contrast with exact statistical techniques (Barnard's test, for example).

Table 2 shows examples of methods for analysis of comparisons between two hypothetical treatments (surgical vs. conventional) analyzed with tests of comparison of proportions and regression models, according to sample characteristics. In the special case of estimation of the magnitude of the effect of a study (for example, relative risk and odds ratios) in which there were zero occurrences of one of the categorical variables, it is possible to resort to the (artificial) addition of 0.5 units to the outcome of each group.^{5,39,40}

Comparison of proportions between groups can also be evaluated unidirectionally or bidirectionally (one/two-tailed), since many analyses are by their nature one-directional, such as comparison of mortality rates from a disease among vaccinated and unvaccinated people or tests of non-inferiority between two treatments.⁴¹ In such cases, the study hypothesis does not contemplate the possibility that the result could be analyzed bidirectionally, since there is only interest in the effect in one direction. One-tailed analyses of proportions do not enjoy consensus among epidemiologists because, although they have greater statistical power and require smaller samples, they increase the likelihood of type I error.²⁴ One-tailed analyses are widely used in studies of viability (pilot studies) and in proof-of-concept studies, which are conducted before traditional clinical trials.⁴²⁻⁴⁴

Situations that involve dependent data should be assessed with the McNemar test (2×2 tables), Cochran's Q test (several groups, dichotomous response), or generalized estimating equations. These analyses, in common with use of resampling techniques, one-tailed estimates, regressions and analyses of variables that demand multivariate adjustment, should be supervised by an experienced statistician.

Finally, comparison of categorical variables is a common need in biomedical studies and inferential conclusions can differ depending on the analytical method employed, especially when the frequencies in subgroups are low. The choice of analytical technique requires theoretical grounding and its description must be justified in the methodology in terms of the parameters for its use.

Table 2. Hypothetical examples of (two-tailed) comparisons of incidence of death from a disease treated with a surgical procedure or a conventional treatment.

Examples	Statistical test	Statistic/effect	<i>p</i> -value
2 deaths in 100 surgeries (2%) vs. 16 deaths in 100 conventional treatments (16%)	Pearson's chi-square	$\chi^2 = 11.97$; Df = 1	<0.001
1 death in 50 surgeries (2%) vs. 8 deaths in 50 conventional treatments (16%)	Barnard	Score = 2.45	0.016
Zero deaths in 50 surgeries (0%) vs. 8 deaths in 50 conventional treatments (16%)	Poisson (robust) regression (robust; 1000 resampling)	RR = 0.13 95%CI = 0.03 to 0.53	0.005
	Barnard	Score = 2.95	0.004
		RR = 0.06 ^a 95%CI = 0.01 to 0.45	0.034

Df = degrees of freedom; RR = relative risk; 95%CI = 95% confidence interval. ^aRelative risk calculated after addition of 0.5 units to the outcome in each group: 0.5 deaths among surgeries, 8.5 deaths among conventional treatments.

■ REFERENCES

- Greenhalgh T. How to read a paper: statistics for the non-statistician. I: Different types of data need different statistical tests. *BMJ*. 1997;315(7104):364-6. <http://dx.doi.org/10.1136/bmj.315.7104.364>. PMID:9270463.
- Miot HA. Analysis of ordinal data in clinical and experimental studies. *J Vasc Bras*. 2020;19:e20200185. <http://dx.doi.org/10.1590/1677-5449.200185>. PMID:34211532.
- Perkins SM. Statistical inference on categorical variables. *Methods Mol Biol*. 2007;404:73-88. http://dx.doi.org/10.1007/978-1-59745-530-5_5. PMID:18450046.
- Pereira JCR. Análise de dados qualitativos: estratégias metodológicas para as ciências da saúde humanas e sociais. São Paulo: EdUSP; 1999.
- Agresti A. An introduction to categorical data analysis. 2nd ed. New Jersey: John Wiley & Sons; 2020.
- Quinn GP, Keough MJ. Experimental design and data analysis for biologists. Cambridge: Cambridge University Press; 2002. <http://dx.doi.org/10.1017/CBO9780511806384>.
- Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*. 2006;25(1):127-41. <http://dx.doi.org/10.1002/sim.2331>. PMID:16217841.
- Naggara O, Raymond J, Guilbert F, Roy D, Weill A, Altman DG. Analysis by categorizing or dichotomizing continuous variables is inadvisable: an example from the natural history of unruptured aneurysms. *AJNR Am J Neuroradiol*. 2011;32(3):437-40. <http://dx.doi.org/10.3174/ajnr.A2425>. PMID:21330400.
- Zaslavsky BG. Bayesian versus frequentist hypotheses testing in clinical trials with dichotomous and countable outcomes. *J Biopharm Stat*. 2010;20(5):985-97. <http://dx.doi.org/10.1080/10543401003619023>. PMID:20721786.
- Turner N. Chi-squared test. *J Clin Nurs*. 2000;9(1):93. PMID:11041649.
- Goodman LA. On the multivariate analysis of three dichotomous variables. *Ajs*. 1965;71(3):290-301. <http://dx.doi.org/10.1086/224088>. PMID:5897475.
- Eberhardt KR, Fligner MA. A comparison of two tests for equality of two proportions. *Am Stat*. 1977;31:151-5.
- Haber M. A comparison of some conditional and unconditional exact tests for 2x2 contingency tables: a comparison of some conditional and unconditional exact tests. *Commun Stat Simul Comput*. 1987;16(4):999-1013. <http://dx.doi.org/10.1080/03610918708812633>.
- Martín Andrés A, Mato AS, Herranz TI. A critical review of asymptotic methods for comparing two proportions by means of independent samples. *Commun Stat Simul Comput*. 1992;21(2):551-86. <http://dx.doi.org/10.1080/03610919208813035>.
- Holmo NF, Ramos GB, Salomao H, et al. Complex segregation analysis of facial melasma in Brazil: evidence for a genetic susceptibility with a dominant pattern of segregation. *Arch Dermatol Res*. 2018;310(10):827-31. <http://dx.doi.org/10.1007/s00403-018-1861-5>. PMID:30167816.
- Tamega AA, Bezerra LVGP, Pereira FP, Miot HA. Blood groups and discoid lupus erythematosus. *An Bras Dermatol*. 2009;84(5):477-81. <http://dx.doi.org/10.1590/S0365-05962009000500005>.
- Amiri P, Javid AZ, Moradi L, et al. Associations between new and old anthropometric indices with type 2 diabetes mellitus and risk of metabolic complications: a cross-sectional analytical study. *J Vasc Bras*. 2021;20:e20200236. <http://dx.doi.org/10.1590/1677-5449.200236>. PMID:34630540.
- Ludbrook J. Analysis of 2 x 2 tables of frequencies: matching test to experimental design. *Int J Epidemiol*. 2008;37(6):1430-5. <http://dx.doi.org/10.1093/ije/dyn162>. PMID:18710887.
- Oliveira NL, Pereira CAB, Diniz MA, Polpo A. A discussion on significance indices for contingency tables under small sample sizes. *PLoS One*. 2018;13(8):e0199102. <http://dx.doi.org/10.1371/journal.pone.0199102>. PMID:30071022.
- Lloyd CJ. A new exact and more powerful unconditional test of no treatment effect from binary matched pairs. *Biometrics*. 2008;64(3):716-23. <http://dx.doi.org/10.1111/j.1541-0420.2007.00936.x>. PMID:18047530.
- Barnard GA. Significance tests for 2 x 2 tables. *Biometrika*. 1947;34(1-2):123-38. <http://dx.doi.org/10.1093/biomet/34.1-2.123>. PMID:20287826.
- Lydersen S, Fagerland MW, Laake P. Recommended tests for association in 2 x 2 tables. *Stat Med*. 2009;28(7):1159-75. <http://dx.doi.org/10.1002/sim.3531>. PMID:19170020.
- Goodman LA. On methods for comparing contingency tables. *J Roy Stat Soc: Series A (General)*. 1963;126(1):94-108. <http://dx.doi.org/10.2307/2982447>.
- Amiri S, Modarres R. Comparison of tests of contingency tables. *J Biopharm Stat*. 2017;27(5):784-96. <http://dx.doi.org/10.1080/10543406.2016.1269786>. PMID:27936354.
- Ludbrook J. Analysing 2 x 2 contingency tables: which test is best? *Clin Exp Pharmacol Physiol*. 2013;40(3):177-80. <http://dx.doi.org/10.1111/1440-1681.12052>. PMID:23294254.
- Choi L, Blume JD, Dupont WD. Elucidating the foundations of statistical inference with 2 x 2 tables. *PLoS One*. 2015;10(4):e0121263. <http://dx.doi.org/10.1371/journal.pone.0121263>. PMID:25849515.
- Sourial N, Wolfson C, Zhu B, et al. Correspondence analysis is a useful tool to uncover the relationships among categorical variables. *J Clin Epidemiol*. 2010;63(6):638-46. <http://dx.doi.org/10.1016/j.jclinepi.2009.08.008>. PMID:19896800.
- Watts DD. Correspondence analysis: a graphical technique for examining categorical data. *Nurs Res*. 1997;46(4):235-9. <http://dx.doi.org/10.1097/00006199-199707000-00009>. PMID:9261298.
- Knapp TR. Treating ordinal scales as ordinal scales. *Nurs Res*. 1993;42(3):184-6. <http://dx.doi.org/10.1097/00006199-199305000-00011>. PMID:8506169.
- Miot HA. Sample size in clinical and experimental studies. *J Vasc Bras*. 2011;10(4):275-8. <http://dx.doi.org/10.1590/S1677-54492011000400001>.
- van Smeden M, Moons KG, de Groot JA, et al. Sample size for binary logistic prediction models: Beyond events per variable criteria. *Stat Methods Med Res*. 2019;28(8):2455-74. <http://dx.doi.org/10.1177/0962280218784726>. PMID:29966490.
- Campbell MJ, Julious SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. *BMJ*. 1995;311(7013):1145-8. <http://dx.doi.org/10.1136/bmj.311.7013.1145>. PMID:7580713.
- Sharpe D. Chi-square test is statistically significant: now what? *Pract Assess, Res Eval*. 2015;20:8.
- Carneiro RM, van Bellen B, Santana PRP, Gomes ACP. Prevalence of incidental pulmonary thromboembolism in cancer patients: retrospective analysis at a large center. *J Vasc Bras*. 2017;16(3):232-8. <http://dx.doi.org/10.1590/1677-5449.002117>. PMID:29930652.
- Goodman LA, Kruskal WH. Measures of association for cross classifications. *J Am Stat Assoc*. 1954;49:732-64.
- Parshall MB. Unpacking the 2 x 2 table. *Heart Lung*. 2013;42(3):221-6. <http://dx.doi.org/10.1016/j.hrtlng.2013.01.006>. PMID:23490241.

37. Miola AC, Miot HA. P-value and effect-size in clinical and experimental studies. *J Vasc Bras*. 2021;20:e20210038. <http://dx.doi.org/10.1590/1677-5449.210038>. PMID:34267792.
38. Katz MH. *Multivariable analysis: a practical guide for clinicians and public health researchers*. Cambridge: Cambridge University Press; 2011. <http://dx.doi.org/10.1017/CBO9780511974175>.
39. Valenzuela C. 2 solutions for estimating odds ratios with zeros. *Rev Med Chil*. 1993;121(12):1441-4. PMID:8085071.
40. Lawson R. Small sample confidence intervals for the odds ratio. *Commun Stat Simul Comput*. 2004;33(4):1095-113. <http://dx.doi.org/10.1081/SAC-200040691>.
41. Pinto VF. Estudos clínicos de não-inferioridade: fundamentos e controvérsias. *J Vasc Bras*. 2010;9(3):145-51. <http://dx.doi.org/10.1590/S1677-54492010000300009>.
42. Mellor K, Eddy S, Peckham N, et al. Progression from external pilot to definitive randomised controlled trial: a methodological review of progression criteria reporting. *BMJ Open*. 2021;11(6):e048178. <http://dx.doi.org/10.1136/bmjopen-2020-048178>. PMID:34183348.
43. Willan AR, Thabane L. Bayesian methods for pilot studies. *Clin Trials*. 2020;17(4):414-9. <http://dx.doi.org/10.1177/1740774520914306>. PMID:32297539.
44. Thabane L, Lancaster G. A guide to the reporting of protocols of pilot and feasibility trials. *Pilot Feasibility Stud*. 2019;5(1):37. <http://dx.doi.org/10.1186/s40814-019-0423-8>. PMID:30858987.

Correspondence

Hélio Amante Miot
 Universidade Estadual Paulista "Júlio de Mesquita Filho" – UNESP,
 Faculdade de Ciências Médicas e Biológicas de Botucatu,
 Departamento de Infectologia, Dermatologia, Diagnóstico por
 Imagem e Radioterapia
 Av. Prof. Mário Rubens Guimarães Montenegro, SN, Campus
 Universitário de Rubião Jr
 CEP 18618-000 - Botucatu (SP), Brasil
 Tel.: +55 (14) 3811-6015
 E-mail: heliomiot@gmail.com

Author information

ACM - Dermatologist; MSc; PhD, Faculdade de Ciências Médicas e
 Biológicas de Botucatu (UNESP).
 HAM – Dermatologist; PhD, Faculdade de Medicina, Universidade
 de São Paulo (FMUSP); Tenured professor, Faculdade de Ciências
 Médicas e Biológicas de Botucatu (UNESP).