How to Interpret Categorical Data in Health Research?

Heru Santoso Wahito Nugroho[1] (Corresponding Author), Sanglar Polnok[2], Joel Rey U. Acob[3], Abdullah Al Mamun[4], Rafif Naufi Waskitha Hapsari[4], Rozan Asyrofi Rakyan Risang Aji[4], Bahtiar[5], Arief Budiono[6], Winarko[1], Handoyo[1], Hery Koesmantoro[1], Sunarto[1], Teta Puji Rahayu[1], Astin Nur Hanifah[1], Nuryani[1], Suparji[1], N. Surtinah[1], Beny Suyanto[1], Resta Betaliani Wirata[7], Nimsi Melati[7], Retno Vivi Intening[7], Tri Wahyuni Ismoyowati[7], Tanko Titus Auta[8]

[1]Health Department, Poltekkes Kemenkes Surabaya, Indonesia; heruswn@gmail.com
[2]Faculty of Nursing, naresuan University, Thailand
[3]Faculty of Nursing, Visayas State University, Philippines
[4]Knowledge Dynamics, Rangpur, Bangladesh
[5]Nursing Department, Poltekkes Kemenkes Makassar, Indonesia
[6]Faculty of Law, Universitas Muhammadiyah Surakarta, Indonesia
[7]Nursing Department, STIKes Bethesda Yakkum Yogyakarta, Indonesia
[8]Department of Nursing, Bingham University Karu, Nasarawa State, Nigeria

ABSTRACT

In health research, many categorical variables are found that must be analyzed using descriptive statistical methods, then interpreted. Currently, there are still many health researchers who interpret categorical variable data in the health sector in the same way as social research in general, namely focusing on the categories with the largest proportions. In fact, for health research, researchers should focus on the expression of unexpected categories. Therefore, this should be a concern for the health research community, especially in the campus environment as the first place for students and health researchers to study and apply various types of health research. This is an urgency, so that the quality of health research can be immediately improved, especially in terms of data interpretation. It was concluded that there were many errors in interpreting categorical variable data in the health sector, so this had to be anticipated from the time of education on campus.

Keywords: health research; categorical variables; data interpretation; unexpected category

INTRODUCTION

The problem that is still frequently encountered today is that there are still many researchers who assume that categorical variables in the health sector are no different from categorical variables in other fields, so they are viewed from the same angle and treated in the same way. When dealing with categorical variables, social researchers will usually focus on the largest proportion.[1] For example, one study stated that the largest type of work among national health insurance participants in Sario District, Indonesia was wage earners (60.4%);[2] or the majority of students study at the science and technology faculty with a proportion of 65%.[3] It should be noted that social researchers use the largest proportion to interpret an expression of a categorical variable, because each category has the same level. For example, there are no different levels for jobs as farmers, entrepreneurs, wage earners (employees), doctors, lawyers and so on. So in the first example above, social researchers interpret that the type of work that dominates is wage-earning workers. For the second example, social researchers also consider that various types of majors in universities are equally good. Thus, science and technology faculties, humanities faculties or others are at the same level. So in the second example, social researchers interpret that the place of study that is most popular with students is the science and technology faculty.

CURRENT PROBLEMS IN THE INTERPRETATION OF CATEGORICAL DATA

In general (not all), categorical variables in the health sector require another approach to interpreting them because often one category has different levels. Normatively, there are categories that are expected and there are also categories that are not expected.[4] For example, the variable for the incidence of stunting among toddlers in Jember Regency, Indonesia in 2022 is distributed into 2 categories, namely stunting = 34.9% and normal = 65, 1%.[5] Of course the expected category is normal. However, many researchers, especially novice researchers or research students, are trapped by the usual interpretations of social researchers, so they write: The majority of toddlers have normal nutritional status (65.1%). This gives the impression that the researchers are satisfied that

the majority of toddlers have good nutritional status or are in a healthy condition in terms of nutrition. In fact, the prevalence of stunting at 34.9% is actually a very big problem for society (Figure 1). So this stunting category should be the focus of interpretation.
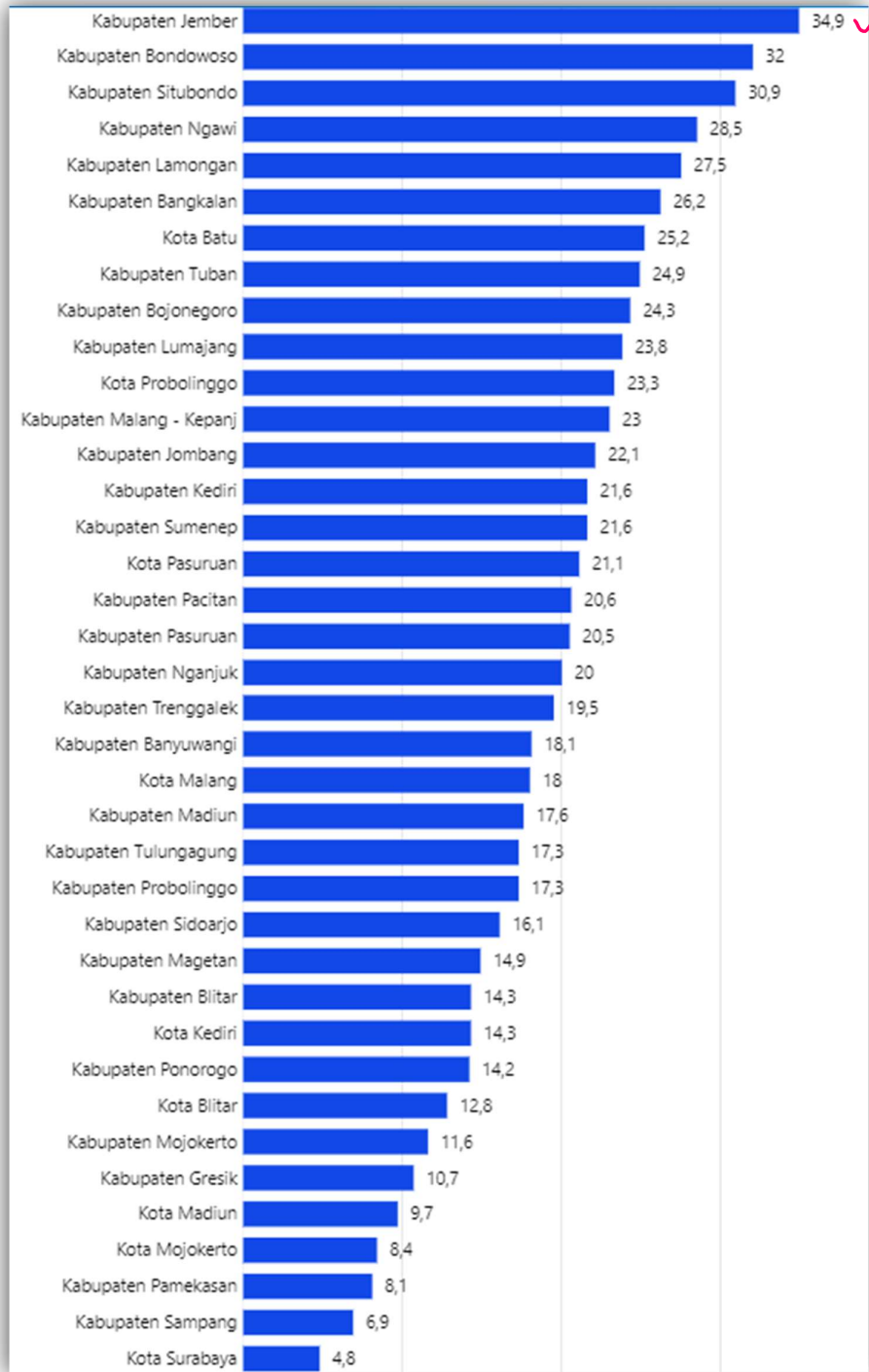

Figure 1. Jember Regency is the district with the highest prevalence of stunting in East Java, Indonesia in 2022

Based on the explanation above, it is best to interpret the expression of a categorical variable for a health researcher to focus on the unexpected category, no matter what the proportion number. To make it easier, researchers can compare it with targets set by local relevant stakeholders or compare it with the previous year's prevalence. Returning to the example of the case in Jember Regency above, stakeholders target that the prevalence of stunting in Jember Regency is less than 10%. So when it was found that the prevalence of stunting among toddlers in Jember Regency was 34.9%, the correct interpretation was that the prevalence of stunting among toddlers in Jember Regency was very high or very far from the target that had been set. Thus, it will give the impression that researchers are concerned because the prevalence of stunting is still a local problem.

The above errors should be a concern, especially for educators in the campus environment, because it is on campus that students and novice researchers learn to do research.[6] It is on campus that this problem can be overcome from an early age, because campus is the earliest place for many parties. to learn about health research. If this error is straightened out from an early age during their education, then when they graduate, health students will be familiar with the correct interpretation procedures for categorical variable data in the health sector, resulting in higher quality research results.[7-16]

CONCLUSION

Many errors have been found in interpreting categorical variable data in the health sector, and this must be anticipated from the time of university education.

REFERENCES

1. Miola AC, Miot HA. Comparing categorical variables in clinical and experimental studies. J Vasc Bras. 2022 Apr 1;21:e20210225. doi: 10.1590/1677-5449.20210225.
2. Umboh AG, Kolibu FK, Korompis GEC. Pekerjaan, pendapatan, pemanfaatan pelayanan kesehatan peserta jaminan kesehatan nasional (Non-PBI) di Kecamatan Sario. Jurnal Kebijakan Kesehatan Indonesia. 2023;12(3);160-167.
3. Nugroho MP, Hartady T, Lesmana R. Hubungan karakteristik individu terhadap tingkat pengetahuan dan penerapan kesejahteraan hewan pada mahasiswa Universitas Padjadjaran. Acta Veterinaria Indonesiana. 2022;10(2):133–141.
4. Wong EC, Maher AR, Motala A, Ross R, Akinniranye O, Larkin J, Hempel S. Methods for identifying health research gaps, needs, and priorities: a scoping review. J Gen Intern Med. 2022 Jan;37(1):198-205. doi: 10.1007/s11606-021-07064-1.
5. Annur CM. Ini rincian angka balita stunting di wilayah Jawa Timur pada 2022, Kabupaten Jember terbesar [Internet]. Katadata. 2022 [cited 2023 Aug 2]. Available from: https://databoks.katadata.co.id/datapublish/2023/02/14/ini-rincian-angka-balita-stunting-di-wilayah-jawa-timur-pada-2022-kabupaten-jember-terbesar#:~:text=Kabupaten%20Jember%20merupakan%20wilayah%20dengan%20prevalensi%20balita%20stunting,balita%20stunting%20daerah%20tersebut%20pada%202021%20sebesar%2023%2C9%25.
6. Bocarro JN, Casper JM, Bush KA, Steptoe A, DuPree S, Blake V, Kanters MA. An exploratory study of a health and wellness intervention on STEM college students during COVID-19. Recreational Sports Journal. 2023 Apr;47(1):35–48. doi: 10.1177/15588661231154490.
7. Preisser JS, Koch GG. Categorical data analysis in public health. Annu Rev Public Health. 1997;18:51-82. doi: 10.1146/annurev.publhealth.18.1.51.
8. Bensken WP, Pieracci FM, Ho VP. Basic introduction to statistics in medicine, part 1: describing data. Surg Infect (Larchmt). 2021 Aug;22(6):590-596. doi: 10.1089/sur.2020.429.
9. Henley SS, Golden RM, Kashner TM. Statistical modeling methods: challenges and strategies. Biostatistics & Epidemiology. 2020;4(1):105-139. doi: 10.1080/24709360.2019.1618653
10. Kader GD, Perry M. Variability for categorical variables. Journal of Statistics Education. 2007;15(2). doi: 10.1080/10691898.2007.11889465
11. Kulkarni DK. Interpretation and display of research results. Indian J Anaesth. 2016 Sep;60(9):657-661. doi: 10.4103/0019-5049.190622.

12. Badawy M, Ramadan N, Hefny HA. Healthcare predictive analytics using machine learning and deep learning techniques: a survey. Journal of Electrical Systems and Inf Technol. 2023;10(40). doi: 10.1186/s43067-023-00108-y

13. Donaires OS, Cezarino LO, Liboni LB, et al. Multivariate data analysis of categorical data: taking advantage of the rhetorical power of numbers in qualitative research. Qual Quant. 2023;57:5283–5312. doi: 10.1007/s11135-022-01589-1

14. Ross PT, Hart-Johnson T, Santen SA, Zaidi NLB. Considerations for using race and ethnicity as quantitative variables in medical education research. Perspect Med Educ. 2020 Oct;9(5):318-323. doi: 10.1007/s40037-020-00602-3.

15. Adane K, Gizachew M, Kendie S. The role of medical data in efficient patient care delivery: a review. Risk Manag Healthc Policy. 2019 Apr 24;12:67-73. doi: 10.2147/RMHP.S179259.

16. Xu B, Feng X, Burdine RD. Categorical data analysis in experimental biology. Dev Biol. 2010 Dec 1;348(1):3-11. doi: 10.1016/j.ydbio.2010.08.018.