How to summarise ordinal data

STIAN LYDERSEN

E-mail: stian.lydersen@ntnu.no
Stian Lydersen, dr.ing. and professor of medical statistics at the Regional Centre for Child and Youth Mental Health and Child Welfare (RKBU Central Norway) at the Department of Mental Health, Norwegian University of Science and Technology.
The author has completed the ICMJE form and declares no conflicts of interest.

Many studies include ordinal data, such as the integers from 1 to 4. Can the mean or median be a relevant summary measure for such data?

Ordinal scales, often called Likert scales, are frequently used in medical research and many other fields. One example is the following question in the Nord Trøndelag Health Study (HUNT): 'How is your health at the moment?'. The response alternatives are 'Poor', 'Not so good', 'Good', and 'Very good', which we will number from 1 to 4. The categories are ordinal, since higher categories reflect better self-rated health. But the 'distances' between the categories need not be equally large. A meaningful quantitative measure of distance between the categories need not even exist.

Table 1 shows the response distribution for this question in Young-HUNT 1 (1), for girls and boys separately. The girls reported poorer health than the boys. Statistically the difference is highly significant: The Wilcoxon-Mann-Whitney test gives p=0.001. The difference is most pronounced for the category 'Very good', which is reported by 32.9 % of the boys and 24.2 % of the girls. In many contexts, such a difference in percentages would be regarded as clinically relevant.

Table 1

Self-rated health for adolescents between 12 and 20 years, from the Nord Trøndelag Health Study in the period 1995–97.

Category	Number (%), from (1)		Number in or above this category (%)		
	Boys	Girls	Boys	Girls	Difference in %
Poor (=1)	3 (0.4)	3 (0.4)			
Not so good (=2)	57 (8.5)	77 (9.2)	671 (99.6)	837 (99.6)	-0.09
Good (=3)	392 (58.2)	557 (66.3)	614 (91.1)	760 (90.5)	0.62
Very good (=4)	222 (32.9)	203 (24.2)	222 (32.9)	203 (24.2)	8.77
Total	674 (100)	840 (100)			9.30

Median of ordinal data

In this example, the median equals category 3 ('Good') both for boys and girls. How is this possible while there is a highly significant difference between the sexes? It is clear that the median is not a good measure of central tendency for ordinal data, particularly when there are few categories. Nevertheless, many researchers report the median as a measure of central tendency for ordinal data, possibly because some maintain that the median and not the mean is relevant if data are not normally distributed (2). But then, how can the Wilcoxon-Mann-Whitney test show a highly significant difference? The answer is that this test is not limited to testing whether the median differs, but generally tests whether the values in one group are higher than in the other. In our example, this is the case, although the median is equal.

Mean of ordinal data

The mean score for boys and girls is 3.236 and 3.143, respectively, and the difference is 0.093. Can this be an appropriate summary measure for the difference between the groups? It is not intuitive how to interpret this for ordinal data. But the difference between mean scores actually has a practical interpretation for a Likert scale (3): If we merge categories 2, 3, and 4, and compare with category 1, the proportion with higher self-rated health is 671/674=0.9955 for boys and 837/840 = 0.9964 for girls, with a difference or excess probability of −0.0009 or −0.09 %. Corresponding numbers for the other possible dichotomisations are shown in Table 1. The sum of excess probability for boys compared to girls is 9.3 %, i.e. 0.093, identical to the difference between the mean scores. In other words, the scale need not have equal distance between the categories for the difference between the mean scores to have a meaningful interpretation.

What should be reported?

Which summary measures are appropriate for ordinal data? In any case, the actual number in each category should be reported, such as in the first two columns in Table 1. Median (and quartiles) are not suited for ordinal data, at least not when there are few categories. The mean has an interpretation in terms of excess probability, and may be a relevant measure in some contexts.

LITERATURE

1. Vie TL, Hufthammer KO, Holmen TL et al. Is self-rated health in adolescence a predictor of prescribed medication in adulthood? Findings from the Nord Trøndelag Health Study and the Norwegian Prescription Database. SSM Popul Health 2017; 4: 144–52, 2. [PubMed][CrossRef]

2. Lydersen S. Mean and standard deviation or median and quartiles? Tidsskr Nor Legeforen 2020; 140. doi: 10.4045/tidsskr.20.0032. [PubMed] [CrossRef]

3. Karlsson P. Är det OK att använda parametriska metoder när man analyserar likertskalor? Quartilen Svenska statistikersamfundets tidsskriftet 2004; 4: 11–2.