

## **Abstract**

The use of product-moment correlations to assess relationships between ordered-categorical variables is widespread and generally tolerated in spite of the non-metric character of such variables. Potential problems of this practice have been identified in the literature. This article focusses on one of these, namely that correlation coefficients are poorly comparable across pairs of ordered-categorical variables because their ranges are constrained in non-uniform ways. This not only affects the validity of correlation coefficients for descriptive purposes, but also as bases for subsequent analyses, as in, e.g., principal component analysis, factor analysis, and structural equation modelling. This article focusses on this problem of 'constrained' correlations. It explains the problem and explores conditions under which it is most severe. It provides illustrations based on both fabricated and actual empirical data, the latter from a mass-survey that ubiquitously contains many ordered-categorical variables that are commonly analysed with product-moment correlations. The article is accompanied by a software tool in R for analysing the attainable upper- and lower bounds of correlations in actual data.

## Introduction

In this article we elaborate why correlations involving ordered-categorical variables are generally ‘constrained’, which means that their attainable maximum and minimum values are not +1 and -1, as is often thought, but between values that have to be calculated for each pair of variables. We assess the magnitude of these constraints in fictitious as well as empirical data and analyse under which conditions they are most severe. The article is accompanied by a software tool in R that allows applied researchers to easily assess the extent of the problem in their own data.

The use of product-moment correlations (also known as Pearson correlations) to assess relationships between ordered-categorical variables – henceforth referred to as *ordcats* – is widespread and widely condoned in empirical studies, in spite of such variables not being metric while the product-moment correlation assumes that they are.

We do not focus on this problem from an obsessive desire for methodological purity, were it only because many statistical procedures are quite robust for violations of their assumptions [fn: examples?]. Instead, we draw attention to a problematic consequence of this violation of assumptions that, as far as we know, has been rarely recognised in the applied literature although it is well-known in the theoretical statistical literature. This is the problem that correlation coefficients involving *ordcats* are constrained in their range which may make them mutually incomparable and may lead to further problems downstream the analytical process.

This article proceeds as follows. We start by briefly elaborating the character of ordinal categorical variables (*ordcats*), why they are generally not regarded as being metric, and why, in spite of that, they are nevertheless very frequently used in statistical procedures that assume metric level data. We then proceed to show why correlations based on *ordcats* tend to be constrained, which implies that their range is smaller than from -1 to +1, and why this is problematic in applied research. This raises the question how to determine the actual range of correlations between *ordcats*, which we address in the subsequent section. In section 5 we assess the magnitude of these constraining effects on 7503 correlations between 123 *ordcats* from a mass survey that typically gives rise ordered categorical variables (*ordcats*), the British Election Study. We use the same data to explore conditions that influence the magnitude of the constraining effect. [Other sections in between?] In the concluding section we discuss implications of our findings, suggestions for dealing with this problem for applied researchers, and directions for further research.

### On ordinal-categorical variables (*ordcats*), their level of measurement and their common use in quantitative analyses

*Ordcats* are ordinal variables with a relatively small number of categories (or ranks) compared to the number of cases, thus resulting in partial orderings of cases. Although they can be found in all field of quantitative inquiry they are ubiquitous in data from standardised social science

**Commented [CV1]:** Scott, if you do not like this leave it as unabbreviated terms, or think of a abbreviation (OCV ?)

**Commented [CV2]:** Or ‘applied’ ??

surveys. Likert items and rating scales [describe in fn Likert items and rating scales, add that there are yet other formats such as semantic-differential scales that also yield ordcats] are among the most favoured ways to elicit responses in such surveys. These survey-question formats have most often between 3 and 11 ordinal categories [fn (or not?) about ANES ‘feeling thermometer’ which has 101 categories but in view of the clustering of responses around only some values can perhaps better be regarded as a 13-point ordcat].

The numerical values used to distinguish the categories of ordcats are most often consecutive integers, starting with either 0 or 1, and in applied survey data analyses it is common to interpret these values as scores. A frequently issued objection against this so-called equal-distance scoring practice [fn: add references] is that it implies that the difference in meaning between, e.g., categories 1 and 2 is equal to that between categories 3 and 4, etc., without clear justification. In other words, it implies an interval-level interpretation by fiat. [add fn: In principle it is quite possible to assign values to survey-based ordcats that have a clear empirical justification, for example by calibrating against psychophysical measurements (cf., Lodge; Tillie; others), or against the results of psychometric/linguistic analysis of verbal response labels (cf Sönning 2024). However, such procedures are rarely integrated in applied survey research because of their high costs.] The consequence of this objection is that ordcats should be considered ordinal variables and that they should not be analysed with statistical procedures requiring metric data. In spite of this objection we find that ordcats are very commonly analysed with product-moment correlations or with procedures that build upon such correlations. We find this clearly exhibited in published articles in high-ranking academic journals [fn: explain how we did this] and in highly regarded textbooks where examples are given of the use of correlations between ordcats as input for procedures such as PCA, FA or SEM [fn: examples: Kline; Byrne; xxx]. This practice is so established that it is occasionally rationalised by referring to ordcats as being ‘quasi-interval level’ measures [fn: reference]. One can, however, also find some justifications in the literature for this practice, which emphasises the robustness of correlation coefficients for violations of variables involved. Labovitz (1967, 1970) in particular has compared correlations between ordcats using the equal-scoring distance with a monotonic random scoring (a procedure that only assumes ordinality between categories). He claims that the differences are negligible, which suggests that even if the theoretical objection against unfounded interval level interpretations of ordcats would be justified, that objection would nevertheless have such minor practical consequences as to make it irrelevant in most situations. Not surprisingly, Labovitz’ work has, in turn, been criticised itself on various grounds, most particularly of (inadvertently) mainly covering situations where this robustness holds, and not very well other situations (cf O’Brien 1979; xxx).

At this place we will not focus on discussions about the equal-distance scoring practice. Our main reason for referring to it is to demonstrate the existence of lively debates in the applied literature about potential issues when using ordcats in procedures requiring metric data. The existence of these debates makes it remarkable that another important issue has –to the best of our knowledge– remained mainly overlooked in the applied literature while it has been well-known in the theoretical statistical literature. This is the problem that correlations between ordcats are generally ‘constrained’, which means that they do not range between -1 and +1, but between minima and maxima that have to be calculated separately for each pair. The nature of this issue is the subject of the next section.

### Ordinal categorical variables (ordcats) and ‘constrained’ correlations

- Simple example of 4x4 crosstab
  - Conclusion: for every given distribution of a variable, there are only two distributions of the other variable which do not constrain r: the same, and the inverse. Every other distribution does constrain r
  - In case of constraintment, the relationship becomes partially nonlinear, by necessity (does this means that the problem of constraintment does not exist for eta?) Is there a measure for nonlinearity? (polynomial multiple correlation?)

Commented [CV3]: Does this need proof?

### Determining the minimum and maximum attainable values of r for given marginal distributions

- Explain different strategies which will all yield the same answers: separate sorting and pairing / Frechet-Hoeffding bounds/ etc
  - Proofs in appendix?

### How severely are correlations constrained – or: how long is a piece of string?

- Explain strategy of using BES data and resulting dataset of 7503 item-pairs
- Descriptive analyses:
  - Define a measure for extent of constraintment
  - Describe univariate distribution of this measure; asymmetry etc.
  - Probability of reaching the r-max and r-min (and implications: does this mean that the constraintment is even more severe?)

### Factors influencing the degree to which correlations are constrained

- Analyses of the same dataset of 7503 item pairs
- Other characteristics of these include

- Characteristics of each of the two vars separately (their variance / skew / kurtosis / mean etc), as well as dyadic characteristics (difference of means / skew/ etc)
- Hypotheses/expectations:
  - Effect of difference of means and of difference of skew
  - Perhaps effect of number of categories (constraint less severe the larger the number of categories)

### **Consequences of constrained correlations in applied research**

- Descriptive interpretation of correlations (Cohen's criteria) is undermined
- It is also undermined by every correlation being expressed on its own unique scale/unique calibration
- Need a nice example of 2 correlations with  $r(a,b) < r(b,c)$  and where  $r(a,b)_{\max} < r(b,c)$ : is it smaller because the linear relationship is weaker, or because the constraint is stronger?

### **Conclusions**

### **References**

### **Appendices**