Ordinal Data Analysis in R

Measuring Human Perceptions from Surveys

Matteo Ventura

April 16, 2025

Course Description

- Surveys are key tools for measuring human perceptions and latent traits.
- Ordinal and rating data are crucial but require specialized techniques for reliable analysis.
- Ordinal data is frequent in real-world applications (customer satisfaction, psychology, medicine).
- This course provides hands-on training in analyzing ordinal data, emphasizing latent patterns.
- Covers theoretical foundations and practical applications using R.
- Highlights the importance of using specialized methods over treating ordinal data as numerical to avoid loss of statistical power and misleading conclusions.

Course Objectives

- Understand what ordinal data is, how it differs from other types of data, and the challenges involved in its analysis.
- Compute and interpret reliability and validity measures.
- Fit proportional odds models in R and interpret the results.
- Analyse rating data by applying CUB models.

The Role of Measurement in Science

- Measurement is a fundamental activity in science.
- We acquire knowledge by observing and quantifying the world.
- Measurement is essential in a wide range of research contexts.
- Scientists often encounter measurement problems even if not primarily interested in measurement theory.

Examples of Measurement Problems

- **Health Psychology:** A psychologist needs a scale to distinguish "wants" from "expectations" regarding physician visits, as existing scales blur this distinction. Improvised questions may lack reliability and validity.
- **Epidemiology:** An epidemiologist conducting secondary analysis on national health survey data needs to investigate the link between perceived psychological stress and health status, but the survey lacks a validated stress measure. A poorly constructed scale could lead to misleading conclusions.
- **Marketing:** A marketing team needs a way to reliably measure parental aspirations across a broad sample to test a hypothesis about product interest for high-end infant toys. This is not easily provided by focus groups.

The Importance of Careful Measurement

- Using arbitrary or poorly designed measurement tools increases the risk of collecting inaccurate data.
- Developing carefully constructed measurement instruments appears to be the most reliable solution.
- Historically, measurement problems were well-known in natural sciences like physics and astronomy.
- Among social scientists, a debate arose regarding the measurability of psychological variables.
- A primary reason was the difficulty in objectively ordering or summing sensory perceptions.

Measurement Classification - Stevens (1946)

- The American psychologist Stevens (1946) disagreed with the perspective that psychological variables were impossible to measure.
- He contended that the rigid requirement of "strict additivity," as seen in measurements of length or mass, was not essential for measuring sensations.
- He pointed out that individuals could make reasonably consistent ratio judgments regarding the loudness of sounds.
- Stevens further argued that this "ratio" characteristic enabled the data derived from such measurements to be mathematically analyzed.
- He is known for categorizing measurements into nominal, ordinal, interval, and ratio scales.
- Despite criticism, Stevens' classification is the most commonly accepted and used internationally.

Properties of Measurement Scales

- Stevens identified four properties for describing the scales of measurement:
- **Identity**: each value has a unique meaning.
- **Magnitude**: the values of the variable have an ordered relationship to one another.
- **Equal intervals**: the data points along the scale are equally spaced.
- **A minimum value of zero**: the scale has a true zero point.

Nominal Scale of Measurement

- This scale has certain characteristics, but doesn't have any form of numerical meaning.
- The data can be placed into categories but can't be multiplied, divided, added or subtracted from one another.
- It's also not possible to measure the difference between data points.
- It defines only the **identity** property of data.
- **Examples:** Gender, Ethnicity, Eye colour.

Ordinal Scale of Measurement

- It defines data that is placed in a specific order.
- While each value is ranked, there's no information that specifies what differentiates the categories from each other.
- These values can't be added to or subtracted from.
- Defines **identity** and **magnitude**.
- **Examples:** satisfaction data points in a survey, where 'one = happy, two = neutral and three = unhappy'.

Interval Scale of Measurement

- The interval scale contains properties of nominal and ordered data.
- The difference between data points can be quantified.
- This type of data shows both the order of the variables and the exact differences between the variables.
- They can be added to or subtracted from each other, but not multiplied or divided.
- In this scale of measurement the zero is just a convention and not absolute.
- Defines **identity**, **magnitude**, and **equal intervals**.

Ratio Scale of Measurement

- This scale include properties from all four scales of measurement.
- The data is nominal and defined by an identity, can be classified in order, contains intervals and can be broken down into exact value.
- Data in the ratio scale can be added, subtracted, divided and multiplied.
- Ratio scales also differ from interval scales in that the scale has a ‘true zero’.
- The number zero means that the data has no value point.
- **Examples:** Weight, height and distance are all examples of ratio variables.

Scales and Questionnaires Development

- Measurement plays a vital role across scientific disciplines.
- In the behavioral and social sciences, the area devoted to measurement is called psychometrics.
- This subfield concentrates on evaluating psychological and social constructs.
- These are most often assessed using questionnaires.
- The following are some practical guidelines to develop measurement scales and questionnaires.

Step 1: Determine Clearly What You Want to Measure

- Researchers often discover initial ideas are vague, leading to costly changes.
- Key questions include whether the scale should be theory-based or explore new directions, its level of specificity, and which aspects to emphasize.
- **Define the theory**: Basing scale development on relevant substantive theories is essential for clearly defining the construct. A theoretical basis helps establish the construct's boundaries. In absence of theory, create a conceptual framework.
- **Determine the level of specificity**: Consider how general or specific the measurement should be. This decision affects how well the scale works in predicting or relating to other variables.
- **Define which aspects are emphasized**: Scale developers must clearly distinguish the target construct from related ones. Including items outside the intended focus can lead to confusion or inaccurate measurement. Item selection should match the specific research

Step 2: Generate an Item Pool

- Items should be carefully selected or created to match the specific construct you aim to measure.
- That means you need a clear idea of what the scale is supposed to do, and every item should reflect that goal.
- Imagine the construct as something hidden or latent, which can't be observed directly.
- The items are the visible signs or behaviors that reflect this hidden thing.
- So, each item acts like a small "test" of how much of that construct a person has.
- If your items truly measure the construct, then someone with a high level of the trait should tend to score higher on all of them.
- When constructing the item pool, consider the following aspects:

Aspects of Item Pool Generation (Cont.)

- **The latent construct**: A good scale includes multiple items for reliability, but every single item must still be strongly connected to the latent construct. A construct is a single, unified idea that can be thought of as causing how someone responds to related items. A category is just a grouping of different constructs. Items related to the same category don't mean they measure the same underlying construct.
- **Redundancy** is crucial for reliability: multiple items allow common content to summate while idiosyncrasies cancel out. Avoid superficial redundancy. Useful redundancy involves expressing the same core idea differently.
- **The number of items**: Start with more items than planned (e.g., 3-4 times as many). An initial pool 50% larger might suffice if items are hard to generate. Eliminate items based on lack of clarity or relevance.
- **The wording**: Including both positively and negatively worded items is a common strategy to reduce acquiescence bias. However, reversing

Step 3: Determine the Format for Measurement

- Defining the measurement format is a critical step.
- This decision impacts data quality, variability, instrument sensitivity, and research conclusions.
- Most scale items consist of two parts: a stem and a series of response options.
- A key aspect is the number of response options.
- One way to increase variability is to include lots of scale items. Another is to provide numerous response options.
- However, too many options can exceed respondents' ability to meaningfully discriminate, leading to "false precision".
- Researchers must balance the need for variability with cognitive limitations.
- Another issue is whether the number of options should be even or odd.
- This choice depends on the type of question and objectives.
- An odd number usually allows expressing neutrality. An even number forces a choice.

Common Measurement Formats

- **Likert scales**: Widely used to measure attitudes/opinions by assessing degree of agreement/disagreement. Typically 5 or 7 points. Response anchors label points. Scales with odd points often include a neutral midpoint. Applied in employee engagement, customer satisfaction, clinical evaluations.

Caution

Although Likert scale data is often numerically coded, remember their ordinal nature and approach calculation of means with caution. Median and mode are more robust.

Common Measurement Formats (Cont.)

- **Semantic Differential scales**: Measure attitudes toward an object/person/idea through pairs of bipolar adjectives. Present concept followed by rows of opposite adjective pairs. Typically 5-7 intermediate points. Respondents evaluate concept by selecting a point. Commonly used in market research, branding. Explore connotative meaning, revealing emotional/evaluative dimensions.
- **Rankings**: Data where items are ordered according to a specific criterion or preference. Respondents arrange items in a sequence. Indicates relative order but not magnitude of difference.

Common Measurement Formats (Cont.)

- **Visual Analog Scale (VAS)**: Presents a continuous line between two descriptors and the respondents mark a point. Allows continuous scoring. Interpretation can be subjective, comparisons across individuals may be difficult. Highly sensitive for detecting subtle changes within individuals over time. May reduce bias from recalling previous discrete responses.
- **Binary Options**: Offer two choices (e.g., agree/disagree, yes/no). Simple for respondents but yields minimal variability per item. More items are required for adequate scale variance.

Step 4: Experts' Review

- Expert review plays a key role in strengthening content validity.
- Reviewers help ensure that the items meaningfully represent the construct.
- Experts assess how well each item reflects the construct definition.
- They also evaluate the clarity and precision of item wording.
- In addition, experts may highlight important aspects that have been overlooked.
- However, content experts may not be familiar with psychometric principles. They might recommend eliminating seemingly redundant items, not realizing some redundancy is intentional for reliability.
- While expert input is highly valuable, final decisions should rest with the scale developer.

Subsequent Steps in Scales Development

- Following the initial design, prepare for validation and data collection.
- Include additional items for later validation (detecting bias, assessing construct validity).
- Administer the questionnaire to a development sample. Sample should be sufficiently large and representative.
- Thorough evaluation of individual items is undertaken. Examine intercorrelations, handle negatively correlated items (e.g., reverse scoring), assess item-scale correlation. Analyze variance and means to ensure discrimination.
- Factor analysis is employed to confirm dimensionality. Reliability (e.g., Cronbach's alpha) is calculated.
- Finally, the length of the scale is optimized. Balance brevity (less respondent burden) with reliability (generally higher with longer scales). Weak items are considered for removal. Cross-validation (e.g., splitting sample) can ensure stability.

Principles for Visualizing Ordinal Data

- **Most important principle:** Always represent ordinal categories in their natural, ordered sequence in any visual representation.
- In bar charts, bars should be arranged based on the logical order of the ordinal scale (e.g., from "Low" to "High").
- For stacked and divergent bar charts, segments should follow this intrinsic order within each bar.
- The choice of chart should align with the research question and the specific aspect of ordinal data being investigated.
- Not all chart types are equally effective for representing ordered categorical data.

Bar Charts for Ordinal Data

- Represent each ordinal category with a bar, whose height or length corresponds to the frequency or count.
- Fundamentally, the bars must be arranged in the logical order of the ordinal variable (e.g., from lowest to highest category).
- They can be vertical or horizontal; horizontal orientation is often preferred for readability of long category labels.
- Bar charts provide a clear and easily understandable visualization of the distribution of a single ordinal variable.

Bar Charts Example

Description: A bar chart showing the frequency of responses for customer satisfaction levels. The levels are ordered from "Very Dissatisfied" to "Very Satisfied". (Code to generate this plot would be shown in the practical session).

Plot: [Insert Vertical Bar Chart Plot Here]

Horizontal Bar Charts Example (Likert)

Description: A horizontal bar chart illustrating responses to a single Likert scale question. Categories are ordered from "Strongly Disagree" to "Strongly Agree". Horizontal orientation improves label readability. (Code to generate this plot would be shown in the practical session).

Plot: [Insert Horizontal Bar Chart Plot Here]

Stacked Bar Charts for Ordinal Data

- Show multiple ordinal categories within a single bar.
- Useful for comparing the distribution of ordinal data across different groups or conditions.
- They can be displayed as counts or as percentages (where each bar totals 100%).
- Allow comparison of both total amounts within each group and the proportion of each ordinal category within those groups.
- Provide insights into how distributions differ between categories.

Stacked Bar Charts Example (Semantic Differential)

Description: A stacked bar chart showing smartphone evaluations using semantic differential scales. Each bar represents a dimension (e.g., "Ineffective - Effective"), segmented by rating level (from 1/Negative to 5/Positive). Useful for comparing evaluation distributions across different smartphones for each dimension.

(Code to generate this plot would be shown in the practical session).

Plot: [Insert Stacked Bar Chart Plot Here]

This visualization effectively reveals patterns such as which smartphone is perceived as more innovative, which has the most consistent ratings across dimensions, and where the greatest differences exist. The stacked bar format is effective for semantic differential scales because it shows the full distribution of responses, not just averages, allowing you to see whether opinions are polarized or consistent.

Divergent Stacked Bar Charts for Ordinal Data

- Specifically designed to visualize ordinal data with a neutral central category or bipolar responses, such as Likert scales and semantic differentials.
- Segments representing responses on one side of the neutral point extend in one direction, while segments representing responses on the other side extend in the opposite direction from a central baseline.
- They effectively illustrate the balance between positive and negative responses and the distribution of opinions.
- Divergent stacked bar charts are the recommended visualization for Likert-type scales as they clearly show the proportion of responses in each category and the overall tendency of agreement or disagreement.

Divergent Stacked Bar Charts Example (Likert)

Description: A divergent stacked bar chart showing responses to multiple Likert scale survey questions (e.g., product ease of use, customer service helpfulness). Positive responses ("Agree", "Strongly Agree") extend to one side, negative responses ("Disagree", "Strongly Disagree") to the other, with a neutral center. Clearly shows the distribution and tendency of agreement/disagreement for each question.
(Code to generate this plot would be shown in the practical session).

Plot: [Insert Divergent Stacked Bar Chart Plot Here]

Other Possible Visualizations

- Depending on the specific analytical objective, these alternative visualizations can provide valuable perspectives on ordinal data.
- Particularly useful when exploring relationships between variables or tracking changes in rankings.
- **Mosaic plots**: Show the relationship between two or more categorical variables, including ordinal ones. Uses tiled rectangles whose area is proportional to the frequency of each combination of categories.
- **Line charts (Bump charts)**: Visualize the change in rank of different items over time or between categories, emphasizing movement in relative positions.

Mosaic Plot Example (Education vs. Satisfaction)

Description: A mosaic plot visualizing the relationship between two ordinal variables: education level and job satisfaction. The width of each column represents the proportion of respondents with that education level, and the height of each colored section within columns represents the proportion reporting that satisfaction level.
(Code to generate this plot would be shown in the practical session).

Plot: [Insert Mosaic Plot Here]

Bump Chart Example (Product Rankings)

Description: A bump chart showing product rankings over different quarters. Lines connect the ranks of each product across time, illustrating changes in relative position. The y-axis is reversed so rank 1 is at the top.
(Code to generate this plot would be shown in the practical session).

Plot: [Insert Bump Chart Plot Here]