# Ordinal Data Analysis in R
## Measuring Human Perceptions from Surveys

Matteo Ventura, PhD
University of Brescia

July 7 - 11, 2025
Universitat Politècnica de Catalunya - BarcelonaTech (UPC)
MESIO Summer School

# Course Objectives

- Understand what ordinal data is, how it differs from other types of data, and the challenges involved in its analysis.
- Understand, fit proportional odds models in R and interpret the results.
- Analyse rating data with CUB models.
- Introduce Multivariate Models for Ordinal Data

# R Packages

Essential R packages for the course:

- `ggplot2`: For data visualization and high-quality plots.
- `dplyr`: For data manipulation and transformation.
- `tidyr`: For reshaping data into a tidy format.
- `HH`: For Divergent Stacked Bar Charts plots.
- `ggmosaic`: For mosaic plots.
- `CUB`: For fitting CUB models.
- `Ordinal`: For fitting Proportional Odds Models.

# Evaluation

The final evaluation consists of a short report (4–6 pages) in which one or more of the methods introduced in the course are applied. The report should include:

- a description of the dataset,
- a clear definition of the research objectives,
- details on data processing (e.g., data cleaning and analysis),
- a discussion of the results.

You may use your own dataset or choose one from open sources. Suggested sources for datasets with ordinal data:

- Eurobarometer
- World Values Survey (WVS)
- European Social Survey (ESS)
- Data in Brief
- R packages: `ordinal`, `CUB`, `psych`, `likert`, `carData`

## Module 1

# Introduction to Ordinal Data and Survey Design

Universitat Politècnica de Catalunya - BarcelonaTech (UPC)
MESIO Summer School

The Role of Measurement in Science

Scales and Questionnaires development

Visualizing Ordinal Data

# The Role of Measurement in Science

Measurement is a fundamental activity in science, indeed we acquire knowledge about the world around us by observing it, and we usually quantify to give a sense to what we observe. Therefore, measurement is essential in a wide range of research contexts.

- Health Psychology
- Epidemiology
- Marketing

These research fields share a common understanding: using **arbitrary** or **poorly** designed **measurement** tools **increases the risk of collecting inaccurate data**. As a result, developing their own carefully constructed measurement instruments appears to be the most reliable solution.

# The Measurement of Psychological Variables

- 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**. While physical attributes like mass and length seem to possess an intrinsic mathematical structure similar to positive real numbers, the measurement of psychological variables **was considered impossible.**
- A primary reason was the difficulty in objectively ordering or summing sensory perceptions.
  *How can one establish that a sensation of "a little warm" plus another similar sensation equals "twice as warm"?*

# Measurement Classification - Stevens (1946)

- The American psychologist Stevens 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. So, this ratio characteristic enabled the data derived from such measurements to be mathematically analyzed.

# 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.

He proposed the categorization of measurements into **nominal**, **ordinal**, **interval**, and **ratio scales**.

# Nominal Scale of Measurement

Data is assigned to distinct categories.

- These categories serve as labels with no inherent numerical value or order.
- Arithmetic operations (like adding or subtracting categories) are not meaningful.
- Defines only the **identity** property of the data.

**Examples:** Gender, Ethnicity, Eye Color, Political Party.

# Ordinal Scale of Measurement

This scale ranks data in a specific order.

- Data points have a meaningful sequence or position.
- The order is clear, but the exact difference or distance between categories is not known or equal.
- Defines both the **identity** and **magnitude** (or order) properties.
- Arithmetic operations like addition or subtraction are generally not valid or meaningful.

**Examples:**

- Ranking of runners in a race (1st, 2nd, 3rd)
- Satisfaction levels (e.g., Very Satisfied, Satisfied, Neutral, Dissatisfied, Very Dissatisfied)
- Educational levels (e.g., High School Diploma, Bachelor's Degree, Master's Degree, PhD)

# Interval Scale of Measurement

This scale builds upon the nominal and ordinal scales by adding the characteristic of equal intervals between data points.

- Data possesses identity and order, just like lower scales.
- Crucially, the difference between any two consecutive points on the scale is consistent and meaningful.
- It defines the **identity**, **magnitude** (order), and **equal intervals** properties.

With interval data, you can perform addition and subtraction to compare differences. However, multiplication and division are not valid because the zero point is arbitrary and does not represent a true absence of the measured property.

# Interval Scale of Measurement - Examples

- **Temperature in Celsius or Fahrenheit** The difference between 20°C and 30°C is the same as the difference between 30°C and 40°C. But 40°C is not ’twice as hot’ as 20°C because 0°C does not mean ’no heat’.
- **Calendar dates** The year 0 AD is based on a historical event; it doesn’t represent the *absolute beginning of time*. Time existed before that point (measured in BC/BCE years). Because the zero is arbitrary, you cannot meaningfully say that the year 2000 AD is ”twice as old” or ”twice as much time elapsed” as the year 1000 AD.

# Ratio Scale of Measurement

This is the most comprehensive level of measurement, possessing all the characteristics of the nominal, ordinal, and interval scales.

- It has a **true zero** point, which genuinely signifies the complete absence of the quantity being measured.
- Because of this true zero, all arithmetic operations (addition, subtraction, multiplication, and division) are valid and meaningful. You can make ratio comparisons (e.g., ”twice as heavy”).

**Examples:** Height, Weight, Distance, Age, Income, Time Duration.

# Scales of Measurement Summary

| Scale | Property | Operators | Advanced Operations | Central Tendency | Variability |
|---|---|---|---|---|---|
| **Nominal** | Classification, membership | $=, \neq$ | Grouping | Mode | Qualitative variation |
| **Ordinal** | Comparison, level | $>, <$ | Sorting | Median | Range, interquartile range |
| **Interval** | Difference, affinity | $+, -$ | Comparison to a standard | Arithmetic mean | Deviation |
| **Ratio** | Magnitude, amount | $\times, /$ | Ratio | Geometric mean, harmonic mean | Coefficient of variation, studentized range |

**Note**: Each scale incorporates the properties of those above it.

The Role of Measurement in Science

## Scales and Questionnaires development

Visualizing Ordinal Data

# Scales and Questionnaires Development

**Psychometrics** is the area of behavioural and social sciences which is devoted to measurement, and it focuses on evaluating psychological and social constructs.
In this research field, the **questionnaires** are the most often tools which are used to assess these constructs.

The following are some practical guidelines to develop measurement scales and questionnaires.

# Step 1: Clearly Define What You Want to Measure

# Step 1: Clearly Define What You Want to Measure

- **Define the Theory or Conceptual Framework:**
  - Base development on relevant theories or a clear framework.
  - Essential for defining the construct and its boundaries.

# Step 1: Clearly Define What You Want to Measure

- **Define the Theory or Conceptual Framework:**
  - Base development on relevant theories or a clear framework.
  - Essential for defining the construct and its boundaries.
- **Determine the Level of Specificity:**
  - Decide how general or specific the measurement should be.
  - Impacts predictive power relative to other variables.

  **Example:** *Measuring general anxiety vs. test anxiety.*

# Step 1: Clearly Define What You Want to Measure

- **Define the Theory or Conceptual Framework:**
  - Base development on relevant theories or a clear framework.
  - Essential for defining the construct and its boundaries.
- **Determine the Level of Specificity:**
  - Decide how general or specific the measurement should be.
  - Impacts predictive power relative to other variables.

  **Example:** *Measuring general anxiety vs. test anxiety.*
- **Clearly Define Aspects to be Emphasized:**
  - Precisely distinguish the target construct from related ones.
  - Avoid including items measuring unrelated phenomena.

  **Example:** *When measuring depression, ensure items focus on psychological symptoms and don't confuse them with physical symptoms from illness.*

# Step 2: Generate the Item Pool

This phase is about creating or selecting the specific items (questions/statements) for the scale.

- Items must **directly match** the construct defined in Step 1.
- They serve as **observable indicators** of a latent (hidden) trait.
- Well-designed items should show that responses reflect the level of the underlying construct.

# Step 2: Generate the Item Pool

This phase is about creating or selecting the specific items (questions/statements) for the scale.

- Items must **directly match** the construct defined in Step 1.
- They serve as **observable indicators** of a latent (hidden) trait.
- Well-designed items should show that responses reflect the level of the underlying construct.

**Example:** Measuring **Anxiety**

- *Latent Construct:* Anxiety (not directly visible).
- *Items (Visible Signs):* "I feel nervous", "My heart races when I'm stressed", "I worry a lot about things"
- Someone with high anxiety would likely agree strongly with these statements.
- These observable responses indicate the level of the hidden anxiety trait.

# Aspects of Item Pool Generation

When constructing the item pool, consider these important aspects:

# Aspects of Item Pool Generation

When constructing the item pool, consider these important aspects:

- **The Latent Construct:**
  - Items must be **strongly connected** to the *single, unified* construct you defined.
  - Think broadly to cover how the construct is expressed, but avoid straying into other areas.
  - **Construct vs. Category:**
    - A construct is a unified idea (e.g., Fear of symptoms).
    - A category is a grouping (e.g., Barriers to compliance).
    - Items in the same category might measure different underlying constructs.
  - Aim for **unidimensionality** (the scale measures primarily one construct).

# Aspects of Item Pool Generation

When constructing the item pool, consider these important aspects:

- **The Latent Construct:**
  - Items must be **strongly connected** to the *single, unified* construct you defined.
  - Think broadly to cover how the construct is expressed, but avoid straying into other areas.
  - **Construct vs. Category:**
    - A construct is a unified idea (e.g., Fear of symptoms).
    - A category is a grouping (e.g., Barriers to compliance).
    - Items in the same category might measure different underlying constructs.
  - Aim for **unidimensionality** (the scale measures primarily one construct).
- **Redundancy:**
  - **Crucial for reliability**: Multiple items measuring the same thing allow common content to summate.
  - Avoid **superficial** redundancy (minor wording changes).
  - Aim for **useful** redundancy (expressing the same core idea differently).
  - Avoid overly specific items in a broad scale that could undermine unidimensionality.

# Aspects of Item Pool Generation (Cont.)

# Aspects of Item Pool Generation (Cont.)

- **The Number of Items:**
  - Start with a **large pool** (e.g., 3-4 times more than planned for the final scale).
  - Allows for careful selection and helps ensure good internal consistency.
  - Eliminate items based on criteria like lack of clarity or relevance later.

# Aspects of Item Pool Generation (Cont.)

- **The Number of Items:**
  - Start with a **large pool** (e.g., 3-4 times more than planned for the final scale).
  - Allows for careful selection and helps ensure good internal consistency.
  - Eliminate items based on criteria like lack of clarity or relevance later.
- **The Wording:**
  - Consider using both **positively and negatively worded items**.
  - *Benefit:* Can help reduce acquiescence bias (tendency to agree).
  - *Risk:* Reversing wording can **confuse** respondents, potentially reducing reliability, especially in general populations or with complex items.
  - Confusion lowers reliability by leading to inconsistent responses.
  - **Example:**
    - Positive: "I feel energetic."
    - Negative: "I feel tired." (requires reverse scoring)

# Step 3: Determine the Format for Measurement

- The definition of the measurement format impacts data quality, variability, instrument sensitivity, and research conclusions. Most scale items consist of two parts: a stem and a series 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 question type and research objectives:
  - An **odd** number typically includes a midpoint, allowing for a neutral response.
  - An **even** number forces respondents to lean towards one side, preventing a neutral stance.

# 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.
  - *Example Item*: ”I am satisfied with the clarity of the instructions.”
  - *Response Options:*
    1. Strongly Disagree
    2. Disagree
    3. Neither Agree nor Disagree
    4. Agree
    5. Strongly Agree

# 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.
  Commonly used in market research, branding. Explore connotative meaning, revealing emotional/evaluative dimensions.

  *Example: How do you perceive your recent consultation?*

  | | | | | | | | |
  |---|---|---|---|---|---|---|---|
  | Bad | 1 | 2 | 3 | 4 | 5 | 6 | Good |
  | Confusing | 1 | 2 | 3 | 4 | 5 | 6 | Clear |
  | Unpleasant | 1 | 2 | 3 | 4 | 5 | 6 | Pleasant |

  (Select a number on each line that best describes your feeling)

# Common Measurement Formats (Cont.)

- **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.
  - *Example*: Please rank the following features from 1 (most important) to 3 (least important):
    - Price (_______)
    - Quality (_______)
    - Design (_______)

# Common Measurement Formats (Cont.)

- **Visual Analog Scale (VAS)**: Presents a continuous line between two descriptors and the respondents mark a point.
  Interpretation can be subjective and comparisons across individuals may be difficult. However, it is highly sensitive for detecting subtle changes within individuals over time.
  **Example:** Please rate your feeling of fatigue:

  No Fatigue ______________________________ Extreme Fatigue

- **Binary Options**: Offer two choices (e.g., agree/disagree, yes/no).
  It is simple for respondents but yields minimal variability per item. More items are required for adequate scale variance.

# Step 4: Experts' Review

Expert review strengthens *content validity* by ensuring items are relevant and representative of the construct.

**What experts contribute:**

- Assess how well each item matches the construct definition.
- Evaluate item clarity, relevance, and wording.
- Identify missing or overlooked content areas.

**Limitations to consider:**

- Experts may lack psychometric expertise.
- They might suggest removing redundant items, unaware that redundancy can enhance reliability.

Expert feedback is valuable, but the final judgment should remain with the scale developer.

# Subsequent Steps in Scale Development

1. **Preparation for validation:**
   - Refine the questionnaire and include extra items to assess bias and construct validity.
   - Administer the questionnaire to a representative development sample.
2. **Item-level analysis:**
   - Evaluate item intercorrelations and item-total correlations.
   - Identify and address negatively correlated items (e.g., reverse scoring).
   - Examine item means and variances to check for adequate discrimination.
3. **Scale-level analysis:**
   - Conduct factor analysis to verify dimensionality.
   - Assess internal consistency (e.g., Cronbach's alpha).
4. **Scale optimization:**
   - Remove weak or redundant items to shorten the scale while maintaining reliability.
   - Consider cross-validation (e.g., split-sample) to test stability and replicability.

The Role of Measurement in Science

Scales and Questionnaires development

Visualizing Ordinal Data

# Principles for Visualizing Ordinal Data

**Most important principle:** Always represent ordinal categories in their natural, ordered sequence in any visual representation.

The effectiveness of a visualization also depends on your analytical goal:

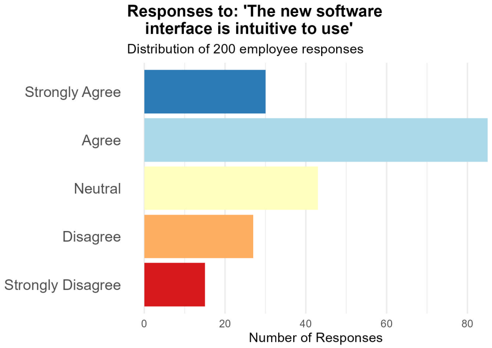
- Choose chart types that reflect the structure and intent of your analysis (e.g., detecting shifts in perception, comparing distributions).
- Be cautious — not all chart types handle ordinal information well. Some may imply equal spacing or distort the sense of order.

# 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.

# 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.

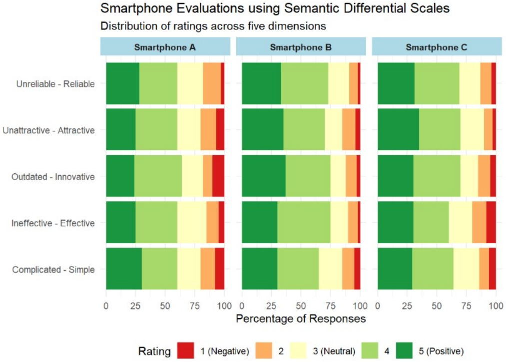


# Stacked Bar Charts for Ordinal Data

- 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.



# 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.

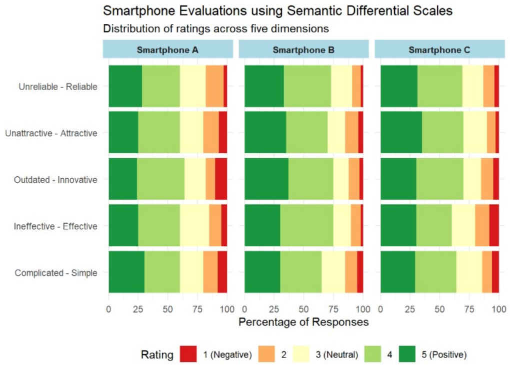



Figure: This barchart is wrong

# Stacked Bar Charts Example (Semantic Differential)

The stacked bar format is effective for semantic differential scales because it shows the full distribution of responses, allowing you to see whether opinions are polarized or consistent.

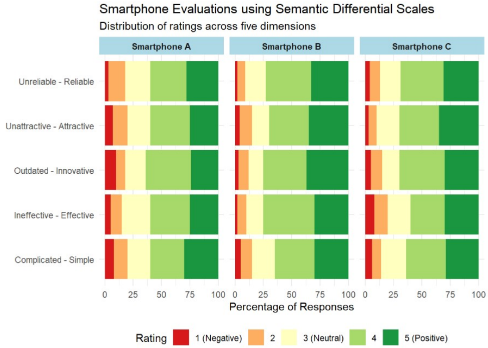


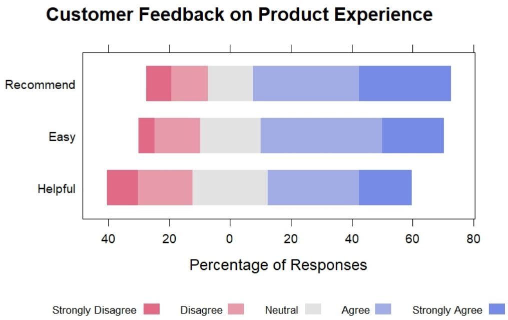

Figure: This barchart is correct

# 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.

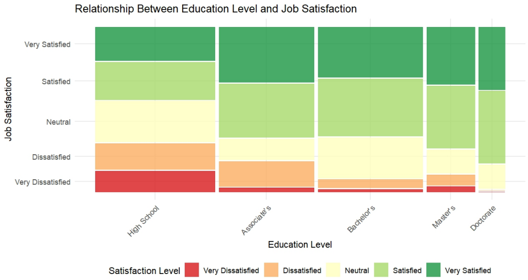


# Other Possible Visualizations

- 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.



# 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.