Module 2

Classical Models for Ordinal Data Analysis

Polytechnic University of Catalonia & University of Barcelona
MESIO Summer School

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Overview of Model Limitations

- When working with ordinal data, applying models designed for continuous or nominal data can lead to significant issues.
- This section will detail the limitations of:
  - Linear Regression
  - Binary Logistic Regression
  - Multinomial Logistic Regression
- Understanding these limitations is crucial for choosing appropriate statistical methods for ordinal outcomes.

Linear Regression: Designed for Continuous Data

- Linear regression is fundamentally designed for dependent variables that are continuous, with values that can take any point within a range (e.g., height, temperature).
- Applying linear regression to ordinal data implicitly treats the ordered categories as if they possess precise numerical scores with equal, quantifiable intervals between them.

Limitation 1: Ignores Non-Interval Nature

- **Core Issue**: Linear regression assumes that the difference between any two adjacent categories is numerically identical.
- **Example**: The "distance" between "Very Dissatisfied" (1) and "Dissatisfied" (2) is assumed to be the same as between "Satisfied" (4) and "Very Satisfied" (5).
- **Reality for Ordinal Data**: This is often false. The psychological or substantive difference between categories may vary.
- **Consequence**: Assigning arbitrary numerical scores (e.g., 1, 2, 3, 4, 5) and applying linear regression imposes an artificial interval structure. This can lead to inaccurate estimates of predictor effects.

Limitation 2: Violation of Assumptions

- Linear regression models rely on key assumptions about the dependent variable and errors:
  1. Dependent variable is continuous.
  2. Errors are normally distributed.
  3. Errors have constant variance (homoscedasticity).
- **For Ordinal Data**: With a limited number of distinct categories, these assumptions are often violated.
- **Prediction Issues**: Linear models can produce predicted values that fall outside the actual range of the ordinal scale (e.g., predicting 0.5 or 5.8 on a 1-5 scale). These predictions are illogical for ordinal outcomes.

Limitation 3: Misleading Interpretation

- **Standard Interpretation**: In linear regression, a coefficient indicates that a one-unit increase in a predictor is associated with a specific change in the *mean score* of the dependent variable.
- **Problem for Ordinal Data**: This interpretation relies on the problematic assumption of equal intervals between categories.
- **Implication**: The interpretation might not accurately reflect the true underlying process that generates the ordinal response.

Binary Logistic Regression: Designed for Two Outcomes

- Binary logistic regression is specifically designed for dependent variables with exactly two possible outcomes (e.g., presence/absence, pass/fail).
- To use it with an ordinal variable that has more than two categories, you are forced to collapse these categories into two binary groups.

Limitation 1: Loss of Information

- **Major Drawback**: Collapsing multiple ordered categories into two results in a significant loss of valuable information.
- **Granularity and Ordering**: The model loses the ability to distinguish between the nuances and the inherent ordering within the original categories.
- **Example**: Reducing a 5-point satisfaction scale to simply "Satisfied" vs. "Not Satisfied" discards whether someone was "Very Dissatisfied" or merely "Dissatisfied."
- A model that can differentiate these levels is inherently more informative.

Limitation 2: Arbitrary Threshold

- The choice of the cut-off point to dichotomize the ordinal scale is often arbitrary.
- Different researchers might choose different thresholds (e.g., ”Neutral” and above vs. ”Dissatisfied” and below, or just ”Satisfied” vs. ”Not Satisfied”).
- **Impact**: This arbitrary decision can significantly alter the results, the estimated effects of predictors, and the overall conclusions drawn from the analysis.

Limitation 3: Reduced Statistical Power

- By reducing the number of outcomes and collapsing categories, the variability present in the dependent variable is potentially reduced.
- **Consequence**: This can lead to a loss of statistical power, making it more difficult to detect significant effects of the predictor variables.
- A model which considers the full range and ordering of the ordinal scale often has greater power.

Multinomial Logistic Regression: Designed for Nominal Data

- Multinomial (or polytomous) logistic regression is appropriate for dependent variables with three or more categories that possess **no natural order**.
- **Examples**: Choice of car color (red, blue, green), religious affiliation, type of occupation.
- While it can handle multiple categories, its underlying structure does not account for any inherent ranking.

Limitation 1: Ignores the Order

- Multinomial logistic regression models the probability of being in each category relative to a chosen baseline category.
- It estimates a **separate set of coefficients** for each category comparison (e.g., Category B vs. A, Category C vs. A, etc.).
- **Key Problem**: It treats the categories as entirely distinct and unordered, completely disregarding the fact that, for ordinal data, Category 3 falls meaningfully between Category 2 and Category 4.

Limitation 2: Difficult Interpretation (in terms of Order)

- Coefficients in a multinomial logit model are interpreted as the change in log-odds of being in a specific category versus the baseline category for a one-unit change in a predictor.
- While technically correct, this interpretation can be **cumbersome** when trying to relate effects back to the overall ordered nature of the dependent variable.
- It does not directly answer intuitive questions like "how does this predictor affect the likelihood of moving up or down the ordinal scale?" Instead, it provides fragmented answers about comparisons to a baseline.

Why Choose Ordinal-Specific Models?

- Statistical models specifically designed for ordinal data respect the ordered nature of the outcome without imposing arbitrary interval assumptions or losing valuable information.
- They provide more accurate parameter estimates and more meaningful interpretations for ordinal variables.

Limitations of Most Commonly Used Models

Modeling Cumulative Probabilities

The Cumulative Logit with Proportional Odds Assumption

Coefficients interpretation

Inference on Parameters

Evaluating Model Fit and Performance

Beyond the Proportional Odds Logit Model

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Core Idea: Respecting Ordinal Order

- The primary methodology for modeling ordinal data focuses on **cumulative probabilities**.
- This approach inherently respects the **ordered nature** of the data.
- It ensures that these probabilities *monotonically increase* as we move up the ordinal scale, which is essential for ordered categories.

Adapting Logistic Regression for Ordinality

- To properly handle ordinality, the cumulative probabilities approach modifies standard logistic regression.
- It applies specific transformations that explicitly consider the order of the categories.
- **Common Transformation**: The **logit transformation** applied to the cumulative probabilities.
  - This enhances the model's ability to capture the ordered nature of the data.
- **Other Transformations**: Probit or Complementary Log-Log (Cloglog) can also be used, depending on data characteristics and theoretical assumptions.

Defining Cumulative Probabilities

Definition: Given an ordinal variable R with m ordered categories, $r_1, r_2, \ldots, r_m$.

- r_1 is the "lowest" category.
- r_m is the "highest" category.
- The categories have a meaningful order: $r_1 \leq r_2 \leq \cdots \leq r_m$.

Defining Cumulative Probabilities

Definition: Given an ordinal variable R with m ordered categories, $r_1, r_2, \ldots, r_m$.

- r_1 is the "lowest" category.
- r_m is the "highest" category.
- The categories have a meaningful order: $r_1 \leq r_2 \leq \cdots \leq r_m$.

A **cumulative probability** for a specific category r_j is the probability that the observed response R falls into category r_j, **or any category below it**.

- Mathematically: $P(R \leq r_j)$.
- We define $m - 1$ such cumulative probabilities, each corresponding to a threshold between categories.

Calculating Cumulative Probabilities: Examples

Let $P(R = r_j)$ be the probability of being in category r_j.

- For the first category r_1: $P(R \leq r_1) = P(R = r_1)$
- For the second category r_2: $P(R \leq r_2) = P(R = r_1) + P(R = r_2)$
- For the third category r_3: $P(R \leq r_3) = P(R = r_1) + P(R = r_2) + P(R = r_3)$
- ...and so on, up to the $(m-1)$-th category $r_{(m-1)}$: $P(R \leq r_{(m-1)}) = P(R = r_1) + \cdots + P(R = r_{(m-1)})$

Note: For the last category r_m, $P(R \leq r_m) = 1$. This probability includes all possible outcomes and therefore provides no information about differences between categories, so it is **not modeled**.

Cumulative Probabilities as Binary Splits

- The use of cumulative probabilities transforms the ordinal modeling problem into a series of binary comparisons while preserving the order.
- Each cumulative probability $P(R \leq r_j)$ inherently creates a binary split at the threshold r_j:
  - **Outcome 1**: The response is in category r_j or lower ($R \leq r_j$).
  - **Outcome 2**: The response is in a category higher than r_j ($R > r_j$).
- By modeling the probability of this binary outcome for each threshold $j = 1, \ldots, m-1$, we effectively capture the transitions between categories along the ordered scale.

Numerical Example: Product Satisfaction

Ordinal Variable: Product Satisfaction, with $m = 4$ ordered categories:

- r_1: Very Dissatisfied (VD)
- r_2: Dissatisfied (D)
- r_3: Satisfied (S)
- r_4: Very Satisfied (VS)

Assumed Specific Category Probabilities:

- $P(R = VD) = 0.10$
- $P(R = D) = 0.20$
- $P(R = S) = 0.40$
- $P(R = VS) = 0.30$

Numerical Example: Calculating Cumulative Probabilities

1. **For r_1 (VD)**: $P(R \leq r_1) = P(R = VD) = 0.10$
   - Implied binary split: $\{VD\}$ vs $\{D, S, VS\}$.

Numerical Example: Calculating Cumulative Probabilities

1. **For r_1 (VD):** $P(R \leq r_1) = P(R = VD) = 0.10$
   - Implied binary split: $\{VD\}$ vs $\{D, S, VS\}$.
2. **For r_2 (D):** $P(R \leq r_2) = P(R = VD) + P(R = D) = 0.10 + 0.20 = 0.30$
   - Implied binary split: $\{VD, D\}$ vs $\{S, VS\}$.

Numerical Example: Calculating Cumulative Probabilities

1. **For r_1 (VD)**: $P(R \leq r_1) = P(R = VD) = 0.10$
   - Implied binary split: $\{VD\}$ vs $\{D, S, VS\}$.
2. **For r_2 (D)**: $P(R \leq r_2) = P(R = VD) + P(R = D) = 0.10 + 0.20 = 0.30$
   - Implied binary split: $\{VD, D\}$ vs $\{S, VS\}$.
3. **For r_3 (S)**: $P(R \leq r_3) = P(R = VD) + P(R = D) + P(R = S) = 0.10 + 0.20 + 0.40 = 0.70$
   - Implied binary split: $\{VD, D, S\}$ vs $\{VS\}$.

Numerical Example: Calculating Cumulative Probabilities

1. **For r_1 (VD)**: $P(R \leq r_1) = P(R = VD) = 0.10$
   - Implied binary split: $\{VD\}$ vs $\{D, S, VS\}$.
2. **For r_2 (D)**: $P(R \leq r_2) = P(R = VD) + P(R = D) = 0.10 + 0.20 = 0.30$
   - Implied binary split: $\{VD, D\}$ vs $\{S, VS\}$.
3. **For r_3 (S)**: $P(R \leq r_3) = P(R = VD) + P(R = D) + P(R = S) = 0.10 + 0.20 + 0.40 = 0.70$
   - Implied binary split: $\{VD, D, S\}$ vs $\{VS\}$.
4. **For r_4 (VS)**: $P(R \leq r_4) = P(R = VD) + P(R = D) + P(R = S) + P(R = VS) = 1.00$
   - **Not modeled**.

For this 4-category variable, the cumulative logit model focuses on the first $m - 1 = 3$ cumulative probabilities.

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

The Cumulative Logit Transformation

- The cumulative logit transformation is used to model the relationship between cumulative probabilities and predictors. For the j-th threshold (from 1 to $m-1$) is the natural logarithm of the cumulative odds:

$$\text{logit}[P(R \leq r_j)] = \log\left(\frac{P(R \leq r_j)}{1 - P(R \leq r_j)}\right)$$

- Since $1 - P(R \leq r_j)$ is the probability that $R > r_j$, this can be rewritten as:

$$\text{logit}[P(R \leq r_j)] = \log\left(\frac{P(R \leq r_j)}{P(R > r_j)}\right)$$

- This transformation maps probabilities (0 to 1) onto the entire real number line $(-\infty, +\infty)$, allowing for linear modeling with predictors.

The Cumulative Logit Proportional Odds Model

- This is the most common statistical model for ordinal data using a cumulative logit link.
- Its basic structure assumes that the cumulative logit for each threshold is a linear function of the predictor variables:

$$\text{logit}[P(R \leq r_j | X_1, \ldots, X_k)] = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- This equation is estimated simultaneously for each of the $m - 1$ cumulative thresholds.

Components of the Proportional Odds Model

$$\text{logit}[P(R \leq r_j | X_1, \ldots, X_k)] = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

Components of the Proportional Odds Model

$$\text{logit}[P(R \leq r_j | X_1, \ldots, X_k)] = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- $P(R \leq r_j | \boldsymbol{X})$: Cumulative probability of being in category r_j or lower, conditional on predictor values $\boldsymbol{X}$.

Components of the Proportional Odds Model

$$\text{logit}[P(R \leq r_j | X_1, \ldots, X_k)] = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- $P(R \leq r_j | \boldsymbol{X})$: Cumulative probability of being in category r_j or lower, conditional on predictor values $\boldsymbol{X}$.
- α_j: **Intercepts** for each of the $m - 1$ cumulative logits.
  - Represent baseline cumulative log-odds when all predictors are zero.
  - Must be ordered: $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_{(m-1)}$ to ensure non-decreasing cumulative probabilities.

Components of the Proportional Odds Model

$$\text{logit}[P(R \leq r_j | X_1, \ldots, X_k)] = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- $P(R \leq r_j | \boldsymbol{X})$: Cumulative probability of being in category r_j or lower, conditional on predictor values $\boldsymbol{X}$.
- α_j: **Intercepts** for each of the $m-1$ cumulative logits.
  - Represent baseline cumulative log-odds when all predictors are zero.
  - Must be ordered: $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_{(m-1)}$ to ensure non-decreasing cumulative probabilities.
- $\beta_1, \beta_2, \ldots, \beta_k$: **Coefficients** for predictor variables $X_1, X_2, \ldots, X_k$.
  - **Proportional Odds Assumption**: There is only one set of β coefficients that applies across all $m-1$ cumulative logit equations.
  - Implies the effect of each predictor on the cumulative log-odds is the **same across all thresholds**.

Demonstrating the Proportional Odds Assumption (1/3)

Consider cumulative odds for two sets of predictors: $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$.

- Odds at threshold r_j for $\boldsymbol{X}^{(1)}$:

$$\text{Odds}(R \leq r_j | \boldsymbol{X}^{(1)}) = \exp\left(\alpha_j + \sum_{i=1}^{k} \beta_i X_i^{(1)}\right)$$

- Odds at threshold r_j for $\boldsymbol{X}^{(2)}$:

$$\text{Odds}(R \leq r_j | \boldsymbol{X}^{(2)}) = \exp\left(\alpha_j + \sum_{i=1}^{k} \beta_i X_i^{(2)}\right)$$

Demonstrating the Proportional Odds Assumption (2/3)

- Odds Ratio (OR) comparing $\boldsymbol{X}^{(2)}$ to $\boldsymbol{X}^{(1)}$ for $R \leq r_j$:

$$\text{OR}_j = \frac{\text{Odds}(R \leq r_j | \boldsymbol{X}^{(2)})}{\text{Odds}(R \leq r_j | \boldsymbol{X}^{(1)})} = \frac{\exp\left(\alpha_j + \sum_{i=1}^{k} \beta_i X_i^{(2)}\right)}{\exp\left(\alpha_j + \sum_{i=1}^{k} \beta_i X_i^{(1)}\right)}$$

Using $e^a / e^b = e^{a-b}$, we simplify:

$$\text{OR}_j = \exp\left(\left(\alpha_j + \sum_{i=1}^{k} \beta_i X_i^{(2)}\right) - \left(\alpha_j + \sum_{i=1}^{k} \beta_i X_i^{(1)}\right)\right)$$

Demonstrating the Proportional Odds Assumption (3/3)

Crucially, the intercept term α_j cancels out:

$$\text{OR}_j = \exp\left(\sum_{i=1}^{k} \beta_i X_i^{(2)} - \sum_{i=1}^{k} \beta_i X_i^{(1)}\right)$$

This further condenses to:

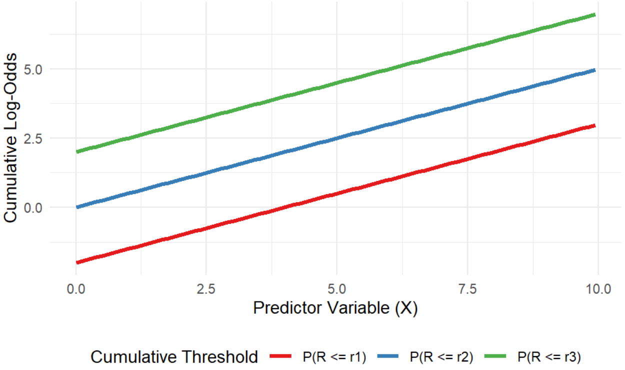
$$\text{OR}_j = \exp\left(\sum_{i=1}^{k} \beta_i (X_i^{(2)} - X_i^{(1)})\right)$$

Conclusion: The Odds Ratio (OR_j) does not contain the subscript j. This means the odds ratio associated with a change in predictor variables is **constant across all** $m-1$ **cumulative thresholds**. This consistency defines the proportional odds assumption.

Graphical Interpretation of Proportional Odds: Parallelism

- On the log-odds scale, the proportional odds assumption translates to **parallelism**.
- If you plot the cumulative log-odds for different levels of a predictor, you would see a set of parallel lines.
- **Example**: For an ordinal outcome ('Low', 'Medium', 'High', 'Very High'), there are three cumulative probabilities: $P(R \leq Low)$, $P(R \leq Medium)$, and $P(R \leq High)$.
- If the PO assumption holds, the log-odds for each cumulative probability would change by the same amount for a one-unit increase in the predictor.

Graphical Interpretation of Proportional Odds: Parallelism



Graphically, three separate curves (one for each cumulative probability) plotted against the predictor would run parallel to each other. They are vertically shifted due to different α_j, but their slopes (β coefficients) are identical.

Advantages of the Proportional Odds Assumption

- **Parsimony and Simplicity**: A major advantage. If the assumption holds, a single set of β coefficients is estimated for each predictor, regardless of the number of categories m.
  - Fewer parameters make the model easier to estimate, interpret, and potentially more stable, especially with smaller sample sizes.
- **Clear and Consistent Interpretation**: A single Odds Ratio per predictor provides a clear, concise summary of its effect across the entire ordinal scale.

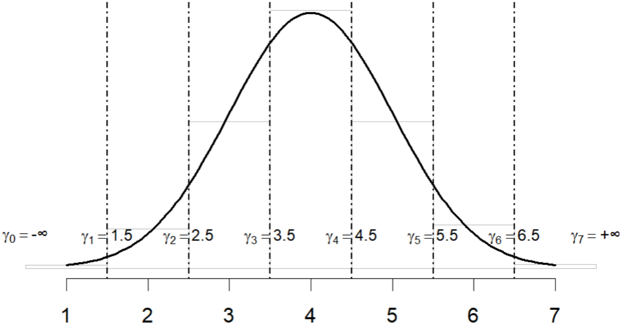
The Latent Variable Concept

- To better understand the Proportional Odds model, we can conceptualize an underlying, unobserved **latent continuous variable**, R^*.
- The observed ordinal response R is assumed to be a categorized version of this continuous R^*.
- We define fixed ordered thresholds, $\gamma_0 < \gamma_1 < \cdots < \gamma_{(m-1)} < \gamma_m$, on the R^* scale.

Mapping Latent Variable to Ordinal Categories

The observed ordinal category R is determined by which interval R^* falls into, defined by these thresholds:

- $R = r_1$ if $R^* \leq \gamma_1$
- $R = r_2$ if $\gamma_1 < R^* \leq \gamma_2$
- $R = r_j$ if $\gamma_{(j-1)} < R^* \leq \gamma_j$
- $R = r_m$ if $R^* > \gamma_{(m-1)}$



From the Latent Variable to the Logit Model

We assume R^* is linearly related to predictors plus some error, similar to linear regression:

$$R^* = \beta_0^* + \beta_1^* X_1 + \cdots + \beta_k^* X_k + \epsilon$$

If the error ϵ follows a standard logistic distribution, this structure directly leads to the cumulative logit model form:

$$\text{logit}[P(R \leq r_j | \boldsymbol{X})] = (\gamma_j - \beta_0^*) - \beta_1^* X_1 - \cdots - \beta_k^* X_k$$

Here, $\alpha_j = \gamma_j - \beta_0^*$ and $\beta_i = -\beta_i^*$.

Proportional Odds in the Latent Variable Framework

In this framework, the Proportional Odds assumption implies two key aspects:

1. The thresholds (γ_j) on the R^* scale are fixed and **do not depend on the predictor variables** X.
2. The effect of each predictor X_i is simply to **shift the entire distribution of the latent variable** R^* along the continuous scale. This shift is of the same magnitude regardless of where the fixed thresholds γ_j are located.

This "parallel shift" of the latent distribution is the fundamental reason why the odds ratios for cumulative probabilities are proportional across all thresholds.

Importance of Testing the PO Assumption

- Violating the Proportional Odds (PO) assumption can lead to **inaccurate conclusions** about the effects of the predictors.
- The most robust and commonly recommended way to test the PO assumption is by using a **Likelihood Ratio Test**.
- This test compares a **constrained model** (the standard PO model) with a more flexible, **unconstrained model** where the PO assumption is relaxed for a specific predictor.

Likelihood Ratio Test for PO Assumption

- Many statistical software packages (e.g., ‘clm()‘ function from the ‘ordinal‘ package in R) allow relaxing the PO assumption for specific predictors using arguments like ‘nominal = predictor‘.
- This creates a ”partial proportional odds” model (also known as a generalized ordinal logit model).
- We then compare this more complex model (where the assumption is relaxed) to our original, simpler proportional odds model using a Likelihood Ratio Test (e.g., ‘anova()‘).

Hypotheses for PO Assumption Test

- **Null Hypothesis (H_0)**: The Proportional Odds assumption holds for the specified predictor.
  - Meaning: Its coefficient is constant across all thresholds.
- **Alternative Hypothesis (H_1)**: The Proportional Odds assumption does not hold for the specified predictor.
  - Meaning: Its coefficient varies across thresholds.

Example: Testing PO Assumption in R

Imagine an ordinal outcome ‘Response‘ (e.g., ’Low’, ’Medium’, ’High’) and a continuous predictor ‘Experience‘.

```
library(ordinal)
# (Data generation and model fitting code omitted for brevity)

# Fit a standard Proportional Odds model
po_model_hypo <- clm(Response ~ Experience, data = data)
summary(po_model_hypo)

# Fit a model where the PO assumption is relaxed for 'Experience'
non_po_model_hypo <- clm(Response ~ Experience, nominal = ~ Experience, da
summary(non_po_model_hypo)

# Compare the models using anova()
anova(po_model_hypo, non_po_model_hypo)
```

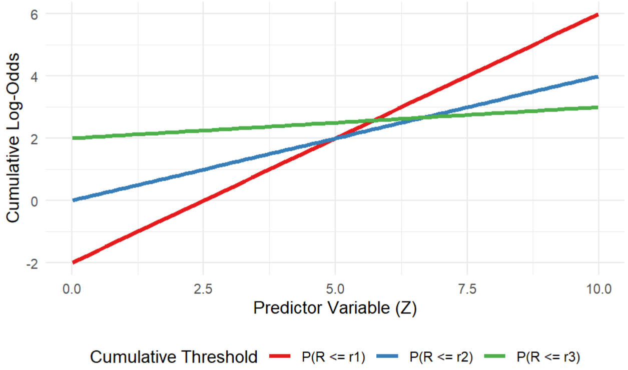
Interpretation of Likelihood Ratio Test Results

- 'po_model_hypo': Standard PO model (effect of 'Experience' constant).
- '$non_po_model_hypo$': Generalized ordinal logit model (effect of 'Experience' allowed to vary).
- **AIC Comparison**: Lower AIC often suggests a better model. If the simpler model has lower AIC, it's often preferred.
- **'LR.stat' (Likelihood Ratio Statistic)**: Test statistic for comparing nested models.
  - Calculated as $2 \cdot (\log L_{\text{unconstrained}} - \log L_{\text{constrained}})$.
  - A small 'LR.stat' indicates the unconstrained model does not offer a significantly better fit.
- **'Pr(¿Chisq)' (p-value)**: Associated with the 'LR.stat'.
  - If $p \geq 0.05$: No statistically significant evidence that the PO assumption is violated. The effect of the predictor does not appear to vary significantly across thresholds.
  - If $p < 0.05$: Statistically significant evidence that the PO assumption is violated.

Graphical Inspection for PO Assumption

- While formal tests are rigorous, they can be overly sensitive, especially in large datasets, flagging numerically small and practically insignificant violations.
- **Graphical inspection** provides a visual assessment of whether coefficients truly remain constant across thresholds.
- This involves fitting a more flexible model (e.g., a non-PO model) and then plotting the estimated coefficients for each predictor across the different thresholds.
- If the PO assumption holds, the estimated coefficients for a given predictor should be very similar across all thresholds, and their confidence intervals should largely overlap.

What a PO Violation Means



Graphically, this means the lines for different cumulative probabilities would not be parallel. Their slopes would differ, indicating that the β coefficients are not the same for each threshold.

Addressing PO Violation: When to Act

- If formal tests and especially graphical inspection indicate a **statistically AND practically significant** violation, proceeding with the standard Proportional Odds model is problematic:
  - Coefficient estimates would be biased.
  - Interpretation would be misleading.
- There are two main alternative models to handle this problem:
  1. Generalized Ordinal Logit Models (Partial Proportional Odds Models)
  2. Multinomial Logit Model

Alternative Model 1: Generalized Ordinal Logit Models

- Also known as Partial Proportional Odds (PPO) Models, are a flexible extension of the PO model.
- They allow coefficients of **specific predictor variables to vary** across the cumulative logit equations (i.e., across thresholds).
- Other predictors (satisfying PO) can still have constant effects.
- Instead of a single β_i for predictor X_i, a PPO model estimates a separate β_{ij} for each cumulative logit r_j.

Generalized Ordinal Logit Models: Advantages

- **Flexibility**: Directly addresses the violation by allowing coefficients to differ where necessary.
- **Parsimony (relative to Multinomial)**: If only a few predictors violate the PO assumption, it estimates fewer parameters than a full multinomial logit model, being more parsimonious and potentially more stable.
- **Maintains Ordinality**: Crucially, it still respects the inherent ordering of the outcome categories. Interpretations are still about "moving up or down" the ordered scale, but the strength of that effect can differ.
- **Richer Interpretation**: Allows for a nuanced understanding, e.g., a predictor might strongly distinguish "Low" from "Medium/High" but have a weaker effect distinguishing "Medium" from "High".

Alternative Model 2: Multinomial Logit Model

- Treats the outcome categories as purely **nominal (unordered)**, even if they are inherently ordinal.
- Fits a separate binary logistic regression model for each category, comparing it to a chosen reference category.
- **Advantages**:
  - **No PO Assumption**: Makes no assumption about constant predictor effects across categories; automatically handles any PO violation.
  - **Maximum Flexibility**: Most flexible approach for categorical outcomes, allowing completely different effects for each category comparison.

Multinomial Logit Model: Disadvantages

- **Loss of Ordinal Information**: Can lead to less precise estimates and interpretations that don't fully reflect the nature of the outcome.
- **Increased Complexity and Reduced Parsimony**:
  - Estimates $(m-1) \cdot k$ coefficients (where k is number of predictors), significantly more than a PO model (k coefficients) or even a PPO model.
  - This increased number of parameters can lead to:
    - Larger standard errors: Less statistical power.
    - Difficulty in interpretation: Cumbersome to interpret multiple sets of coefficients and odds ratios.
    - Overfitting: Especially with smaller sample sizes, models might fit current data well but generalize poorly.

Note on PO Assumption Violation

Statistical vs. Practical Significance

- A statistically significant p-value from a formal test (e.g., Likelihood Ratio Test) does not always signify a **practically meaningful** violation of the PO assumption, especially in large datasets.
- **Graphical inspection is invaluable**:
  - If estimated coefficients across thresholds are very similar and their confidence intervals largely overlap (near-parallelism), the practical implication of the violation might be minimal despite statistical rejection. In this case, retaining the more parsimonious standard PO model is often appropriate.
  - Conversely, if coefficients vary widely with little to no overlap in their confidence intervals, indicating both statistical and practical significance, then adopting a more flexible model (PPO or Multinomial Logit) becomes necessary for accurate data reflection.

Limitations of Most Commonly Used Models

Modeling Cumulative Probabilities

The Cumulative Logit with Proportional Odds Assumption

Coefficients interpretation
- Continuous variables
- Categorical Variables

Inference on Parameters

Evaluating Model Fit and Performance

Beyond the Proportional Odds Logit Model

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Interpreting Predictor Effects (β_i)

- Consider the effect of changing a single quantitative predictor, X_i, by one unit, while holding all other predictors constant.
- The change in the **cumulative log-odds** for a one-unit increase in X_i is simply the coefficient β_i.
- Therefore, β_i is the change in the cumulative log-odds for a one-unit increase in X_i, holding other predictors constant.

Interpreting Predictor Effects (Odds Ratios)

- To obtain the **Odds Ratio (OR)**, we exponentiate the coefficient: $OR_i = \exp(\beta_i)$.
- This OR_i represents the **multiplicative change in the cumulative odds** for a one-unit increase in X_i.
- Specifically, for a one-unit increase in predictor X_i (holding other predictors constant), the odds of being in category r_j or any category below it, versus being in a category above r_j, are multiplied by $\exp(\beta_i)$.

Important Note: Due to the Proportional Odds assumption, this multiplicative effect, $\exp(\beta_i)$, is the **same for all $m - 1$ thresholds**.

Direction of Effect: Positive Coefficient ($\beta_i > 0$)

- If $\beta_i > 0$ (and thus $\exp(\beta_i) > 1$):
  - A one-unit increase in X_i **increases the cumulative log-odds**.
  - This means it **increases the odds of being in category r_j or below**.
  - **Interpretation**: A positive β_i indicates that higher values of X_i are associated with a **greater likelihood of being in the lower (or earlier) categories** of the ordinal variable R.
  - Equivalently, it's associated with a lower likelihood of being in the higher categories.

Direction of Effect: Negative Coefficient ($\beta_i < 0$)

- If $\beta_i < 0$ (and thus $\exp(\beta_i) < 1$):
  - A one-unit increase in X_i **decreases the cumulative log-odds**.
  - This means it **decreases the odds of being in category r_j or below**.
  - **Interpretation**: A negative β_i indicates that higher values of X_i are associated with a **greater likelihood of being in the higher (or later) categories** of the ordinal variable R.
  - Equivalently, it's associated with a lower likelihood of being in the lower categories.

Latent Variable Interpretation

- Recall the latent variable model: $R^* = \beta_0^* + \beta_1^* X_1 + \cdots + \beta_k^* X_k + \epsilon$.
- In this framework, a positive β_i^* means that increasing X_i **increases the value of the latent variable** R^*.
- Since higher values of R^* correspond to higher ordinal categories, a positive β_i^* implies a shift towards **higher categories**.

Reconciling Signs: Latent vs. Cumulative Logit

- As derived, the cumulative logit model coefficient β_i (for $P(R \leq r_j)$) is typically the negative of the latent variable coefficient: $\beta_i = -\beta_i^*$.
- **Example**: If $\beta_i^* > 0$ (meaning X_i increases the latent R^*, shifting towards higher categories), then $\beta_i < 0$ in the cumulative logit model.
- A negative β_i in the cumulative logit model is indeed associated with a greater likelihood of being in higher ordinal categories (as seen previously, it decreases the odds of being in lower categories).
- This confirms the consistency between the two interpretations, although the sign convention can initially be confusing.

Standard Interpretation: Focus on Cumulative Logit Output

- To avoid confusion, it is generally easiest and most standard to interpret the results directly from the estimated coefficients (β_i) and Odds Ratios ($\exp(\beta_i)$) from the cumulative logit model output:
- **If exp**$(\beta_i) > 1$:
  - Higher values of X_i are associated with **increased odds of being in a lower category** (or equivalently, decreased odds of being in a higher category).
  - The shift is towards the beginning of the ordered scale.
- **If exp**$(\beta_i) < 1$:
  - Higher values of X_i are associated with **decreased odds of being in a lower category** (or equivalently, increased odds of being in a higher category).
  - The shift is towards the end of the ordered scale.

Interpreting Dummy Variables in Proportional Odds Models

- When including categorical predictors, they are often represented by **dummy variables**.
- For a binary categorical variable (e.g., Gender: Male, Female), one category is chosen as the **reference category**.
- Example: Let's define a dummy variable for Gender:
  - $X_{\text{male}} = 1$ if Gender = Male
  - $X_{\text{male}} = 0$ if Gender = Female (Reference Category)

Model Structure with Dummy Variable

The proportional odds model incorporating Gender would look like:

$$\text{logit}[P(R \leq r_j Gender, OtherPredictors)] = \alpha_j + \beta_{male} X_{male} + \beta_{OthPred} X_{OthPred}$$

- β_{male}: Coefficient for the Male dummy variable.
- $\beta_{OthPred}$: Coefficients for any other predictor variables.

Interpretation of Dummy Variable Coefficients

- The coefficient for a dummy variable (β_{dummy}) represents the **difference in the cumulative log-odds** between the category represented by the dummy variable and the reference category.
- This difference is observed **holding all other predictors constant**.

Odds Ratio for a Dummy Variable:

- The Odds Ratio (OR) is calculated as $e^{\beta_{\text{dummy}}}$.
- This OR represents the ratio of the odds of being in a **lower outcome category** for the dummy variable group compared to the reference group, holding other predictors constant.

Example: Gender Predictor for Self-Rated Health

Ordinal Outcome: "Self-Rated Health" with categories (1=Poor to 5=Excellent).
Predictor: Gender (Male, Female).
Reference Category: Female.
Assumed Result: Estimated $\beta_{male} = 0.4$.
Calculated Odds Ratio: $e^{0.4} \approx 1.49$.

Interpreting the Example Results

Coefficient ($\beta_{male} = 0.4$):

- Males have, on average, **0.4 units higher cumulative log-odds of being in a lower self-rated health category** (e.g., Poor, Fair, Good, Very Good) compared to females, holding other predictors constant.

Odds Ratio ($OR_{male} = 1.49$):

- The odds of a male reporting "Poor or Fair or Good or Very Good Health" (i.e., a lower category) vs. "Excellent Health" are **1.49 times the odds for a female**, holding all other predictors constant.
- This proportionality holds across all cumulative splits:

Simpler Terms: Males have a higher odds of being in the less healthy (lower) categories of self-rated health compared to females. This effect is assumed to be consistent across the entire range of health categories due to the proportional odds assumption.

Limitations of Most Commonly Used Models

Modeling Cumulative Probabilities

The Cumulative Logit with Proportional Odds Assumption

Coefficients interpretation

Inference on Parameters

Evaluating Model Fit and Performance

Beyond the Proportional Odds Logit Model

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Inference on Parameters: Overview

- After fitting the model using Maximum Likelihood Estimation, it's crucial to assess the statistical significance and precision of the estimated parameters.
- We will cover:
  - Significance Tests for Individual Predictors (Wald Test)
  - Confidence Intervals for Coefficients and Odds Ratios

Significance Tests for Individual Predictors: Wald Test

- **Purpose**: To determine if an individual predictor has a statistically significant effect.
- **How it Works**: Most software provides a Wald Test for each coefficient.
  - Calculates a z-statistic: $Z = \frac{\text{Estimate}}{\text{Std Error}}$.
  - This is squared to get a χ^2 statistic with 1 degree of freedom: $Wald\chi^2 = Z^2$.
  - A p-value is then calculated from this χ^2 statistic.
- **Hypotheses**:
  - H_0: The coefficient for this predictor is zero (no effect).
  - H_1: The coefficient for this predictor is not zero (significant effect).
- **Interpretation**:
  - If $p < 0.05$ (or chosen α): Reject H_0. Predictor has a statistically significant effect.
  - If $p \geq 0.05$: Fail to reject H_0. Insufficient evidence of a significant effect.

Confidence Intervals (CIs) for Coefficients

- While p-values indicate significance, CIs provide a **range of plausible values** for the true population parameter, offering insights into precision.
- **Interpretation for Coefficients (β):**
  - A 95% CI for a coefficient means that if the study were repeated many times, 95% of the calculated CIs would contain the true population coefficient.
  - If a CI for β **does not include 0**, then the coefficient is statistically significant at the corresponding alpha level (e.g., 0.05 for a 95% CI).

Confidence Intervals (CIs) for Odds Ratios (e^{β})

- **Interpretation for Odds Ratios (e^{β}):**
  - A 95% CI for an OR means we are 95% confident that the true population OR lies within this range.
  - If a CI for an OR **does not include 1**, then the effect is statistically significant.
  - If the CI is entirely **above 1**, the predictor significantly **increases the odds of a lower outcome**.
  - If the CI is entirely **below 1**, the predictor significantly **decreases the odds of a lower outcome** (i.e., increases the odds of a higher outcome).

Limitations of Most Commonly Used Models

Modeling Cumulative Probabilities

The Cumulative Logit with Proportional Odds Assumption

Coefficients interpretation

Inference on Parameters

Evaluating Model Fit and Performance

Beyond the Proportional Odds Logit Model

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Evaluating Model Fit: Pseudo R-squared Measures

- In linear regression, R^2 explains variance. For ordinal logistic regression (and other generalized linear models), traditional R^2 is not appropriate due to the non-linear link function.
- **Pseudo R-squared** measures provide an analogous quantification of model fit.
- They compare the log-likelihood of the fitted model (L_{model}) to the log-likelihood of a null (intercept-only) model (L_{null}).
- **Common Types**:
  - **McFadden's** R^2: $1\left(\frac{L_{model}}{L_{null}}\right)$
  - **Cox & Snell's** R^2: $1\left(\frac{L_{model}}{L_{null}}\right)^{\frac{2}{n}}$

Pseudo R-squared: Important Caveat

Caution: Do NOT Compare Directly to Linear Regression R^2

- Pseudo R-squared values are typically much **lower** than R^2 values from linear regression models, even for models that fit the data very well.
- A McFadden's R^2 of 0.20 might be considered **very good** in an ordinal logistic regression context, whereas a linear R^2 of 0.20 would often be considered weak.
- Interpret them cautiously and primarily for **relative comparison** between different ordinal models on the same dataset.

Evaluating Model Fit: Likelihood Ratio Test

- Used for comparing two **nested models**: a simpler (restricted) model versus a more complex (full) model that contains all parameters of the simpler model plus additional ones.
- **Purpose**: To see if the more complex model significantly improves the fit.
- Let $L_{restricted}$ be the maximum likelihood of the simpler model.
- Let L_{full} be the maximum likelihood of the more complex model.
- **Likelihood Ratio Test Statistic (Λ)**:

$$\Lambda = -2 \cdot \log \left(\frac{L_{restricted}}{L_{full}} \right) = -2 \cdot [\log(L_{restricted}) - \log(L_{full})]$$

Likelihood Ratio Test: Hypotheses and Interpretation

- **Null Hypothesis (H_0)**: The additional parameters in the full model do not significantly improve the fit; the simpler model is sufficient.
- **Alternative Hypothesis (H_1)**: The more complex model provides a significantly better fit.
- **Distribution**: Under H_0, Λ asymptotically follows a chi-squared distribution (χ^2) with degrees of freedom equal to the difference in the number of parameters between the two models.
- **Interpretation**:
  - A **large** Λ **value** (and small p-value) indicates that the more complex model fits the data significantly better than the simpler model.
  - This leads to the rejection of H_0, suggesting the additional parameters are meaningful.

Evaluating Model Fit: Information Criteria

- Information criteria balance model fit with model complexity.
- Useful for comparing **non-nested models** or multiple competing models.
- **Lower values generally indicate a better model**.
- Goal: Find a model that explains the data well without being overly complex.
- **AIC (Akaike Information Criterion)**:

$$\text{AIC} = -2\log(L_{model}) + 2k$$

- **BIC (Bayesian Information Criterion)**:

$$\text{BIC} = -2\log(L_{model}) + k\log(n)$$

Information Criteria: Components

- L_{model}: Maximum likelihood of the fitted model.
- k: Number of parameters in the model, including intercepts (cut-points) in ordinal regression.
- n: Sample size (only for BIC).

AIC vs. BIC:

- **AIC** tends to favor more complex models and is generally better for **prediction accuracy**.
- **BIC** tends to favor simpler, more parsimonious models and is often preferred for **model selection** when the goal is to identify the ”true” underlying model.

The Proportional Odds Model: A Starting Point

- The Proportional Odds (PO) Model with a logit link function is the most common default for ordinal data analysis.
- **Reasons for Popularity**:
  - Good interpretability (log-odds and odds ratios).
  - Computational stability.
- However, the PO Model is just one member of a broader family of models for ordinal outcomes.
- The choice of model can significantly influence interpretation and fit to the data.
- This section explores alternative approaches.

General Form of Cumulative Models

- In a cumulative model, the "link function" transforms the cumulative probabilities to a linear scale, where they are then modeled by predictors.
- The general form is:

$$g[P(R \leq r_j)] = \alpha_j - \boldsymbol{\beta}^T \boldsymbol{X}$$

- Where:
  - $P(R \leq r_j)$: Cumulative probability of being in category j or lower.
  - $g(\cdot)$: The chosen link function.
  - α_j: Category-specific intercepts (cut-points).
  - $\boldsymbol{\beta}$: Vector of regression coefficients for the predictors.
  - $\boldsymbol{X}$: Vector of predictor variables.
- The logit function is common, but other link functions exist.

Alternative Link Function: Probit Link

- The probit link uses the inverse of the standard normal cumulative distribution function (Φ^{-1}).

$$\text{probit}[P(R \leq r_j)] = \Phi^{-1}[P(R \leq r_j)] = \alpha_j - \boldsymbol{\beta}^T \boldsymbol{X}$$

- It models cumulative probabilities on a scale that corresponds to the normal distribution.
- **Coefficient Interpretation**: Coefficients are interpreted in terms of standard deviation units of the underlying latent normal variable.
  - A one-unit increase in X_k leads to a β_k standard deviation change in the latent variable.
  - They **do not** have the direct odds ratio interpretation of the logit model.
- **Usage**: Often preferred when there's a theoretical belief that the underlying continuous variable driving the ordinal outcome is normally distributed.

Alternative Link Function: Log-log Link

- The Log-Log link is defined as:

$$\text{loglog}[P(R \leq r_j)] = \log(-\log(P(R \leq r_j))) = \alpha_j - \beta^T \boldsymbol{X}$$

- **Usage Context**: This link is often used when the probability of the lowest category is expected to decrease very quickly, or when the process leading to higher categories accelerates rapidly.
- **Interpretation**:
  - Less straightforward than the logit or probit.
  - It's more sensitive to changes in the upper tail of the probability distribution.
  - Implies that the probability of being in a lower category decreases rapidly.

Beyond Link Functions: Different Probability Explanations

- Besides changing the link function for cumulative probabilities, models can also alter what probabilities they are designed to explain.
- This leads to different model structures that may be more suitable for specific ordinal data patterns.

Alternative Model Structure: Adjacent Categories Logit Model

- Instead of cumulative probabilities, this model focuses on the log-odds of being in category r_j versus the **next adjacent category** r_{j+1}.

$$\log\left(\frac{P(R = r_j)}{P(R = r_{j+1})}\right) = \alpha_j - \boldsymbol{\beta}^T \mathbf{X} \quad \text{for} \quad j = 1, \ldots, m-1$$

- **Key Feature**: The coefficients ($\boldsymbol{\beta}$) are **not constrained** to be the same across all adjacent log-odds comparisons.
  - This means it **does not assume the proportional odds assumption**.
  - Each α_j is a separate intercept, and $\boldsymbol{\beta}$ can be a separate vector of coefficients ($\boldsymbol{\beta}_j$) for each j.
- **Interpretation**: For each pair of adjacent categories j and $j+1$, $\exp(\beta_k)$ is the odds ratio of being in category j versus $j+1$ for a one-unit change in X_k.

Adjacent Categories Logit Model: Pros and Cons

- **Advantages**:
  - **Flexibility**: It does not impose the proportional odds assumption, making it more flexible.
- **Disadvantages**:
  - **Increased Complexity**: It leads to the estimation of $m - 1$ sets of predictors.
  - **More Parameters**: Many more parameters compared to the Proportional Odds Model ($k \times (m - 1)$ coefficients vs. k coefficients).
  - **Larger Standard Errors**: Can lead to larger standard errors and require larger sample sizes for stable estimation.

Alternative Model Structure: Continuation Ratio Logit Model

- This model focuses on the log-odds of being in category j versus being in a **higher category**, given that the outcome is **at least** j.
- It's a **sequential modeling approach**.
- The model is expressed as:

$$\log\left(\frac{P(R = r_j \mid R \geq r_j)}{P(R > r_j \mid R \geq r_j)}\right) = \alpha_j - \beta_j^T \mathbf{X} \quad \text{for } j = 1, \ldots, m-1$$

- Where:
  - $P(R = r_j \mid R \geq r_j)$: Conditional probability of observing category r_j, given the outcome is r_j or higher.
  - $P(R > r_j \mid R \geq r_j)$: Conditional probability of observing a category higher than r_j, given the outcome is r_j or higher.
  - β_j: Coefficients can vary across sequential comparisons.

Continuation Ratio Logit Model: Interpretation and Applications

- **Coefficient Interpretation**: $\exp(\beta_{jk})$ represents the odds ratio of observing category j versus observing a category higher than j, given the outcome is at least j, for a one-unit increase in x_k.
- **Flexibility**: Typically does not assume proportional odds (coefficients β_j can vary). A "proportional odds" variant can be imposed by forcing β to be constant.
- **Suitability**: Especially suited for situations where ordinal categories represent a natural progression or a series of choices.
- **Common Applications**:
  - **Educational Attainment**: Odds of graduating high school vs. continuing to college, then college vs. graduate studies.
  - **Disease Progression**: Odds of staying at current disease stage vs. progressing to next, more severe stage.
  - **Consumer Behavior**: Odds of making a basic purchase vs. upgrading to a premium version.