Case Studies and Statistical Consulting a.y. 2024-2025

Psychometric Validation and Food Insecurity Prevalence Estimation in Albania: A Comprehensive Analysis of the FIES Scale using the Rasch Model

Paola Maria Lepore
Federica Sfeir

2025-07-17

Introduction

This project aims to analyze food insecurity in Albania in the year 2017 using microdata from FAO. The core focus is the application of the Rasch model to the Food Insecurity Experience Scale (FIES), allowing for the measurement of latent food insecurity severity and the estimation of prevalence across population subgroups, assessing its reliability. Particular attention will be paid to comparing gender, geographic area, educational level, and income in order to explore potential systematic inequalities in the perception or testing of food insecurity.

The choice of Albania as a study context is particularly significant for several compelling reasons. Like many post-communist nations, Albania has undergone and continues to navigate substantial economic and social transformations. These transitions can uniquely influence household food security, underscoring the critical need for accurate and nuanced monitoring of the situation. Furthermore, marked socio-economic disparities are often prevalent within Albania, particularly between urban and rural areas and across different socio-economic strata. Differential Item Functioning (DIF) analysis, as conducted for variables such as urban residence and potentially gender, is therefore fundamental to ascertaining whether the FIES scale functions equitably across these diverse subgroups, thereby providing more granular data crucial for policy formulation. Crucially, reliable and validated data on food insecurity prevalence are indispensable for both the Albanian government and international organizations operating in the country. Such data are vital for monitoring progress towards the Sustainable Development Goals, identifying the most vulnerable populations, and subsequently designing and implementing targeted social protection programs and effective food interventions.

Dataset Overview

The dataset used in this analysis consists of 1,000 observations collected in Albania during 2017. It comprises a total of 29 variables, covering a wide range of demographic, socioeconomic, and food security dimensions.

Each observation includes a sampling weight (weights), which varies from approximately 0.22 to 2.16, with a mean centered around 1. These weights are essential for producing population-level estimates that reflect the survey design.

```
## [1] 1000   29
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.2163  0.5298  0.8717  1.0000  1.3420  2.1634
```

The dataset includes several key variables grouped into three main categories:

1. Demographic characteristics
   - Gender: A binary variable coded as 1 for male and 2 for female.
   - Education: An ordinal variable indicating educational attainment, where 1 corresponds to a low level of education, 2 to medium, and 3 to high.

- Urban: A categorical variable capturing area of residence, with four levels: 1 = rural, 2 = town, 3 = city, and 4 = capital.
- HH_size: Reports household size, with an average of approximately 3.4 members per household.
- Employment: A binary variable indicating whether the respondent is employed, although it contains substantial missing data.
- hours_work: Indicates hours worked per week.

2. Income
   - income_int_dol: Measures household income in USD, with a minimum value of 21.5 USD and a maximum of over 15,000 USD. The mean income is approximately 2,672 USD, indicating substantial variability across households.
   - income_quint: A categorical variable representing the income quintile, ranging from 1 (poorest 20%) to 5 (richest 20%).

3. Food insecurity indicators (FIES): the dataset includes the standard 8-item Food Insecurity Experience Scale (FIES), developed by the Food and Agriculture Organization (FAO). These binary items assess whether the respondent experienced difficulties in accessing adequate food over the past 12 months. The items are formulated as experience-based questions with yes/no responses (coded as 1 and 0 respectively and I don't know (DK) or refuse to answer (RF) coded as DK/RF=NA), and are ordered approximately by severity. The items are:
   - 1. WORRIED - You were worried you would not have enough food to eat?
   - 2. HEALTHY - You were unable to eat healthy and nutritious food?
   - 3. FEWFOOD - You ate only a few kinds of foods?
   - 4. SKIPPED - You had to skip a meal?
   - 5. ATELESS - You ate less than you thought you should?
   - 6. RUNOUT - Your household ran out of food?
   - 7. HUNGRY - You were hungry but did not eat?
   - 8. WHLDAY - You went without eating for a whole day?

These items are widely used in global monitoring to assess mild to severe food insecurity, with higher affirmative counts indicating higher severity.

In this dataset, the FIES variables are well-completed, with very low levels of missing data, allowing for robust application of the Rasch model. A small number of missing values (NA) were detected across the FIES items, ranging from 0 to 5 per item. Given their limited presence and the large sample size, we opted to ignore these missing values for the purpose of the analysis, assuming they are missing completely at random (MCAR) and thus unlikely to bias the results.

```
## WORRIED HEALTHY FEWFOOD SKIPPED ATELESS  RUNOUT  HUNGRY  WHLDAY
##       0       1       5       2       0       1       5       2
```

```
## [1] 985   8
```

Descriptive Analysis

The following figure provides a visual overview of three key sociodemographic characteristics of the surveyed population: gender, education level, and area of residence.
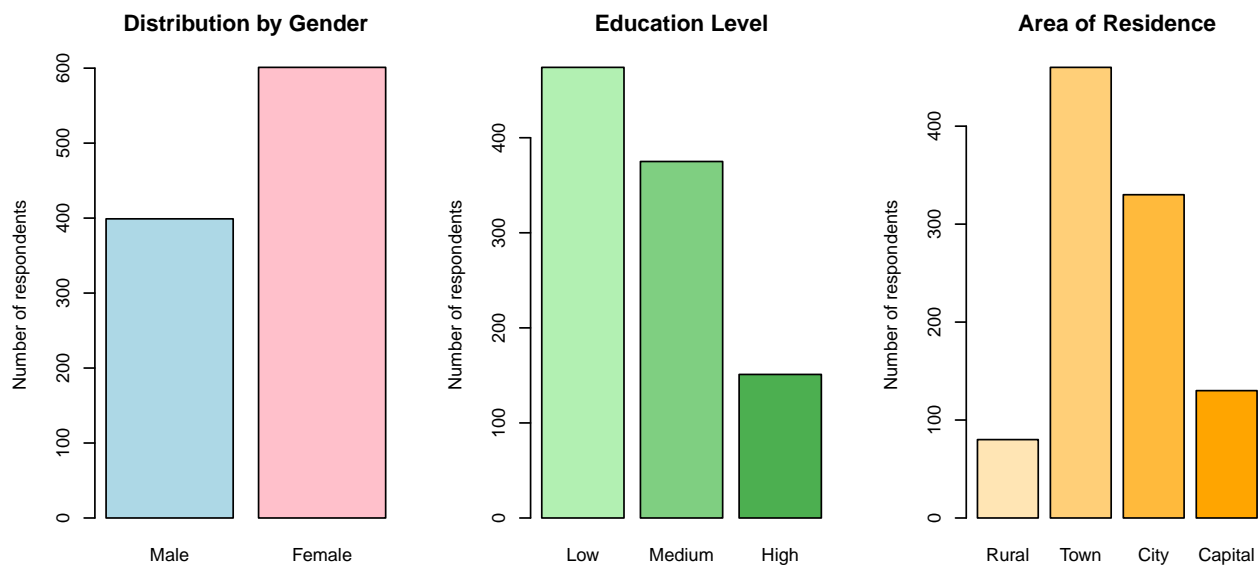
Gender distribution: the sample appears to be relatively balanced in terms of gender. The bar chart shows a slightly higher number of female respondents compared to males, although the difference is modest. This balance supports gender-disaggregated analyses of food insecurity.

Education level: education is categorized into three levels low, medium, and high. The most frequent category is low education, followed by low education, with high education being the least represented. This distribution suggests a population where access to higher education may still be limited, reflecting national trends in educational attainment.

Area: the population is distributed across four types of areas rural, towns, cities and the capital. The largest group resides in towns, followed by cities and rural areas, while the capital accounts for the smallest proportion of the sample. This indicates that the majority of respondents live in semi-urban or urban environments, which has important implications for interpreting food insecurity dynamics across different settlement types.

The two figures present two complementary visualizations of household income in the Albanian 2017 sample.
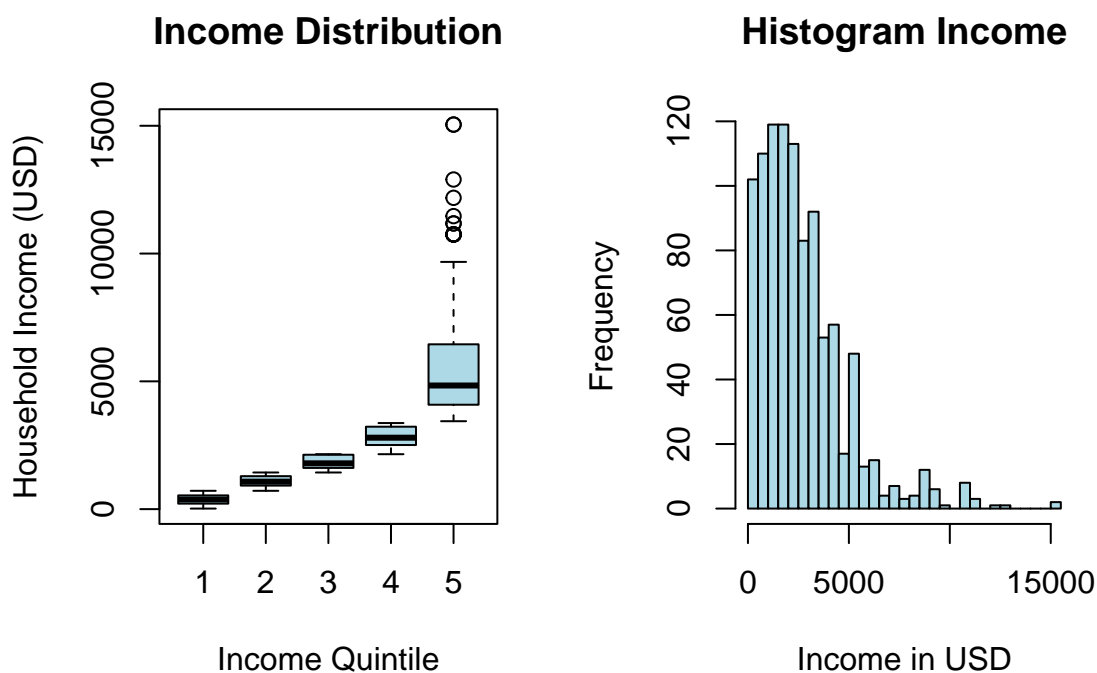

The boxplot on the left illustrates the distribution of household income across quintiles. The progression

from the 1st to the 5th quintile shows a clear and steep income gradient. While the lower four quintiles are relatively compact and homogenous in distribution, the 5th quintile exhibits both a significantly higher median and a much wider interquartile range. It also presents a notable number of outliers, some exceeding 15,000 USD, indicating strong income concentration among a small group of high earners.

This visual evidence points to intra-group inequality within the richest segment, which may not be captured when quintiles are treated as uniform strata. The histogram on the right reinforces this finding by showing a highly skewed distribution, with the majority of households concentrated in the lower income brackets. The long right tail reflects the influence of high-income outliers, confirming that the overall distribution is far from symmetric.

The figure presents the distribution of responses (Yes/No) to each of the eight items in the Food Insecurity Experience Scale (FIES).
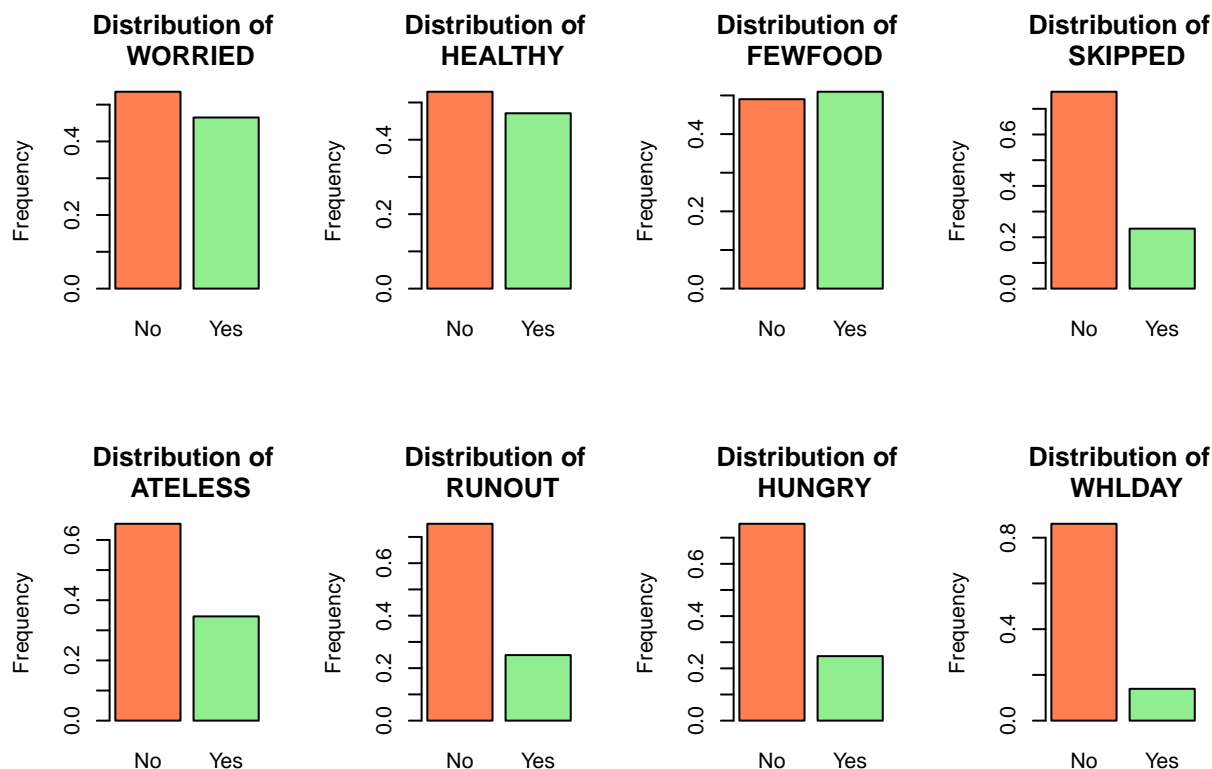

Each bar chart shows the relative frequency of respondents affirming or denying a specific experience related to food insecurity within the past 12 months. The items appear to follow a logical progression in severity, starting from more common and mild experiences such as worrying about food (WORRIED) and limited food variety (FEWFOOD), to more severe conditions such as going hungry (HUNGRY) or not eating for a whole day (WHLDAY). Items such as WORRIED and HEALTHY show relatively balanced distributions, with a significant proportion of respondents answering "Yes".

As severity increases (e.g., ATELESS, RUNOUT, HUNGRY, WHLDAY), the share of "Yes" responses decreases, reflecting the expected structure of the scale: more severe conditions are less frequent. Notably, RUNOUT and HUNGRY still show non-negligible proportions of affirmative responses, indicating that a considerable subset of the population experienced moderate to severe food insecurity. The final item, WHLDAY, is the rarest, with a small proportion of "Yes" responses consistent with its role as the most severe endpoint of the FIES scale.

These distributions support the use of the Rasch model: the items appear to represent increasing levels of severity along a unidimensional construct. Additionally, the varying frequencies suggest that the items will have different difficulty (severity) parameters, which the Rasch model can estimate.

The following histogram shows the distribution of Raw Scores, which represent the sum of affirmative responses given by a respondent across all items in a scale. For the FIES, each score ranges from 0 to 8, with higher values indicating a higher number of food insecurity experiences reported by an individual or an household.

Distribution of Raw Scores

Raw Score (number of positive responses)
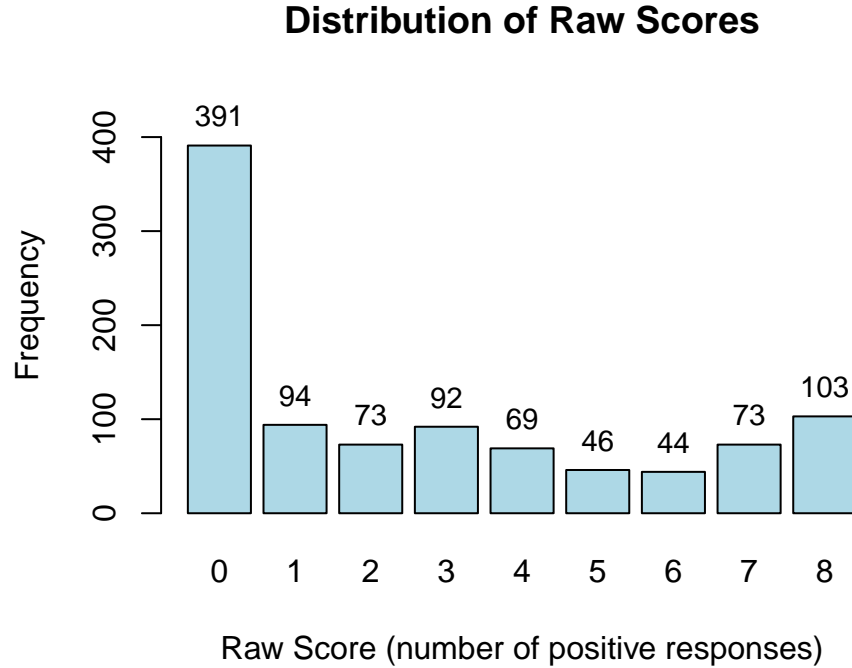
Although raw scores are ordinal in nature, the Rasch model treats them as sufficient statistics for estimating a respondent's latent trait level. This means that, under the Rasch model's assumptions, two individuals with the same raw score are assumed to have the same level of the latent trait, regardless of which specific items they endorsed.

From the histogram, it is immediately evident that the distribution is highly skewed to the right. A large majority of respondents, more than 400 individuals, reported a Raw Score of 0. This means that they did not experience any of the listed food insecurity conditions during the reference period, suggesting that most of the sample is food secure.

As we move to the right along the x-axis, the frequency of responses gradually decreases, indicating fewer individuals experiencing moderate levels of food insecurity (scores between 1 and 5). Interestingly, the distribution rises again slightly at the upper end of the scale (scores 6–8), particularly at the maximum score of 8. This local increase suggests the presence of a smaller but significant subgroup of respondents experiencing severe food insecurity, who affirmed nearly all items in the scale.

Rasch Model Formulation

The Rasch Model is a probabilistic Item Response Theory (IRT) model that relates the probability of a correct response to an item to the person's ability and the item's difficulty. The most common formulation for dichotomous items is done throguh the logit model:

$$P(X_{ij} = 1|\theta_i, \beta_j) = logit^{-1}(\theta_i - \beta_j) = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}}$$

It computes the probability that a person will respond positively to item i, given the person's ability and the item difficulty: it increases as the latent trait increases and decreases as the severity increases. The parameter θ_i is the latent trait for individual i and the parameter β_j is the severity for item j.

Model's Assumptions

The Rasch's model is based on some fundamental assumptions:

- The probability that a respondent answers positively to an item is a logistic function of the latent trait (food insecurity).

- The eight items measure different shades of the same single latent trait, ensuring unidimensionality.

- Items are conditionally independent, that is they are correlated exclusively due to their association with the latent trait.

- The probability of responding positively to an item should always increase as the level of the person's latent trait increases, ensuring monotonicity.

- All the items have the same discrimination power, that is they are all associated with the latent trait with the same intensity.

Preliminary Data Analysis: Ensuring Rasch Model Suitability

It is crucial to conduct preliminary data analyses to ensure the validity, reliability, and fundamental assumptions of the model are met. This fundamental step is essential for obtaining robust and interpretable results.

Our preliminary assessment focuses on key psychometric properties of the scale items:

- Internal Consistency (Reliability)

- Item Interrelationships

- Unidimensionality

Internal Consistency

The Reliability of the model is evaluated using Cronbach's Alpha to determine the overall coherence and precision of the items. It indicates how closely related a set of items are as a group, reflecting their coherence in measuring a single construct.

$$\alpha = \frac{k}{k-1} \left[1 - \frac{\sum_{j=1}^{k} Var(y_{.j})}{Var(y_0)} \right]$$

Where $Var(y_{.j})$ is the single item's variance and $Var(y_0)$ is the variance of the total scale score.

[1] 0.9181552

The value computed with our dataset is $\alpha = 0.918$, which indicates that the FIES items demonstrate excellent internal consistency, supporting their use as a coherent measure of food insecurity.

Item Interrelationships

The item interrelationships are explored through the Item Correlation Matrix to understand the patterns and strength of relationships among the items.

Given that the items in our scale are binary (e.g., yes/no responses), we used tetrachoric correlations to estimate the relationships between items. This method assumes an underlying continuous latent trait and is more appropriate than Pearson correlations for binary variables, providing a more accurate measure of item associations in the context of latent constructs.

corrplot 0.95 loaded
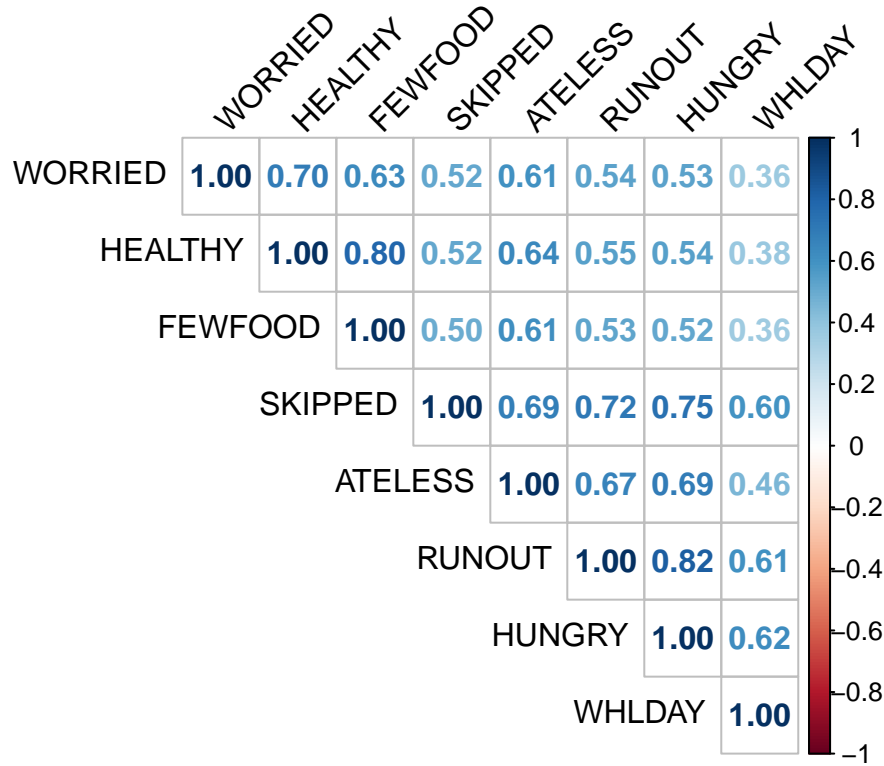
Generally we observe positive and moderate-to-strong correlations among all items. The correlations between HEALTHY and FEWFOOD (0.80) and between RUNOUT and HUNGRY (0.82) are very high. This is common in scales where items representing very close levels of severity may be highly correlated because they are nearly equivalent in their degree of difficulty.

WHLDAY, the most severe level of food insecurity, shows the lowest correlations with the other items, particularly with the less severe items (e.g., WORRIED: 0.36, HEALTHY: 0.38). This pattern is consistent with the hierarchical nature of the FIES scale: people who experience milder forms of food insecurity, such as WORRIED or HEALTHY, are much less likely to experience the more severe form (WHLDAY), leading to lower correlations between the two extremes of the severity scale.

Conversely, those who experience WHLDAY have almost certainly also experienced SKIPPED, RUNOUT and HUNGRY, which explains the higher correlations between these higher severity items.

Unidimensionality

The unidimensionality is a fundamental assumption of the Rasch model: the scale must measure one single underlying latent construct, that is Food Insecurity Severity. To assess the unidimensionality of the FIES scale, we conducted a factor analysis. Although the items are binary, factor analysis is still appropriate when using tetrachoric correlations, which estimate the relationships between dichotomous items under the assumption of an underlying continuous latent trait, allowing for a valid evaluation of the scale's dimensional structure.

The Bartlett's Test returns a p-value equal to 0, indicating that our data is feasible for factor analysis. Then, as shown in the screeplot, the analysis clearly suggests one dominant factor.

Parallel Analysis Scree Plots
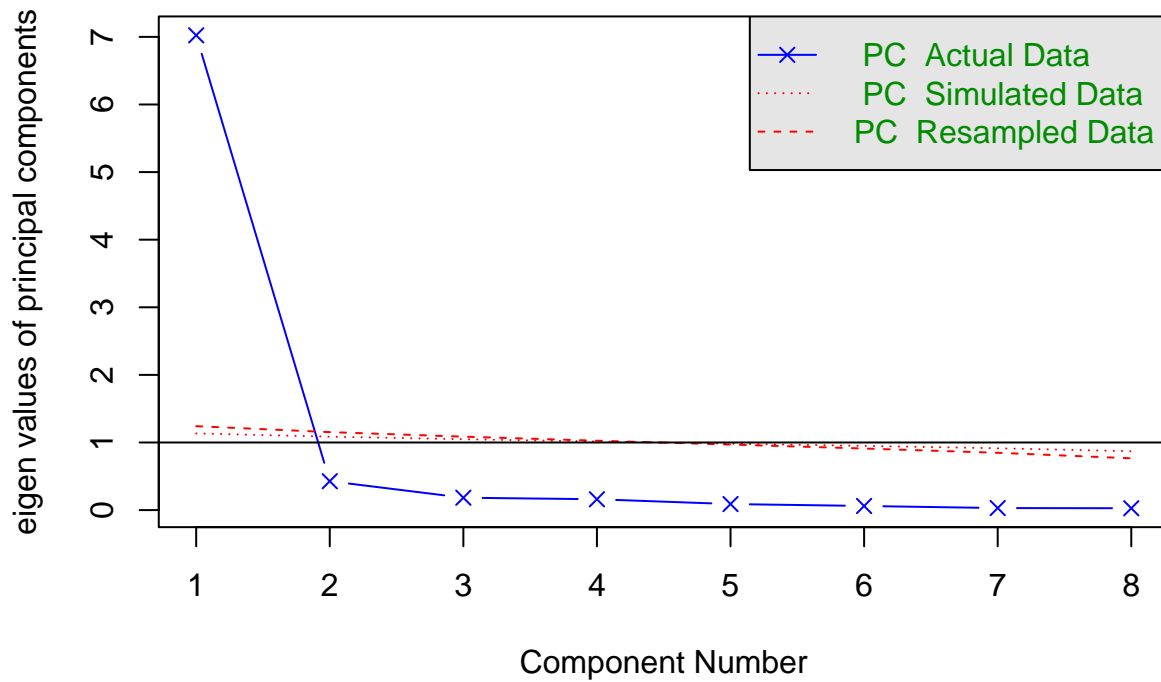

```
## Parallel analysis suggests that the number of factors =  NA  and the number of components =  1
```

At the end we have applied the EFA, which has given us the results in the plot of factorial loads, which shows the factor loadings for each item. All the loadins are extremely high and positive, between 0.87 and 0.97, meaning that each item is highly correlated to this single factor and hence, they contribute to measure the latent construct.

Factorial Loads
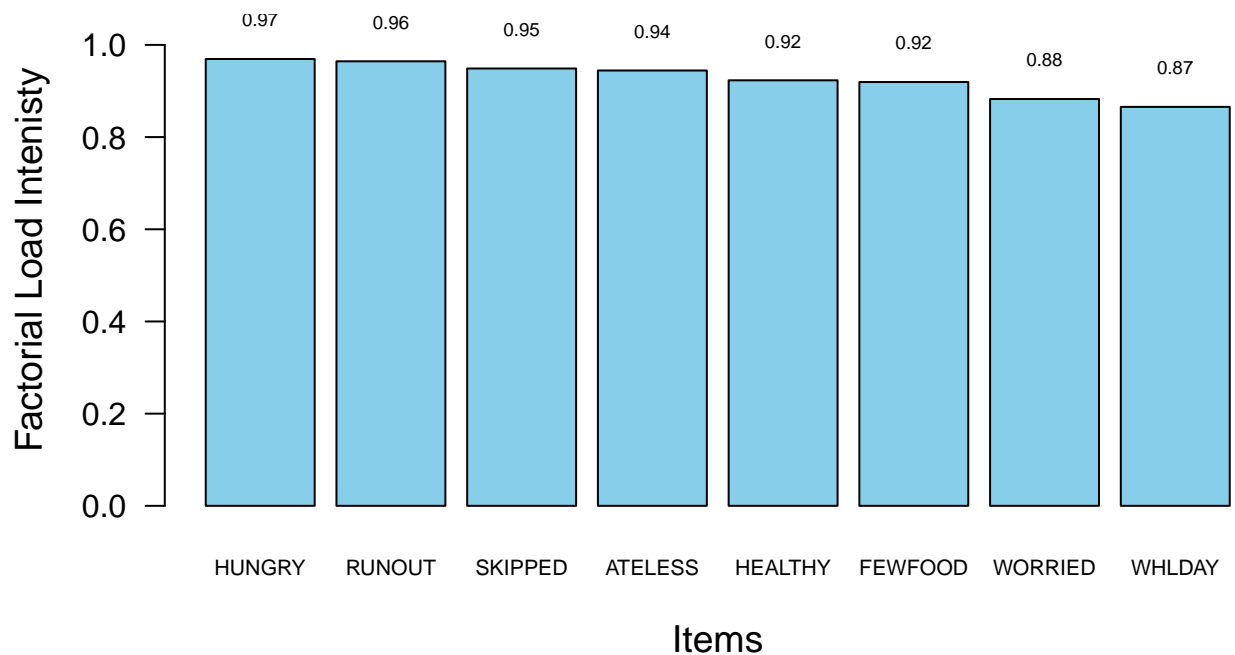
Rasch Model Estimation

The goal in this section is to determine the item difficulty parameters β_j and the person ability parameters θ_i that best fit the observed response data, according to the Rasch model. The most common used method is The Maximum Likelihood Estimate. There are three approaches:

- Joint Maximum Likelihood (JML): Estimates all item and person parameters simultaneously. While straightforward, it can produce biased estimates.

- Conditional Maximum Likelihood (CML): Focuses on estimating item parameters first, by conditioning on the raw scores. Person parameters are then estimated separately. This method yields consistent estimates for item parameters.

- Marginal Maximum Likelihood (MML): Integrates person parameters out of the likelihood function by assuming a distribution for person abilities. This allows for robust estimation of item parameters.

```
##
## Attaching package: 'eRm'

## The following object is masked from 'package:psych':
##
##     sim.rasch

##
## Results of RM estimation:
##
## Call:  RM(X = data_fies, sum0 = TRUE)
##
## Conditional log-likelihood: -846.9221
## Number of iterations: 21
## Number of parameters: 7
##
## Item (Category) Difficulty Parameters (eta):
##              HEALTHY    FEWFOOD    SKIPPED    ATELESS    RUNOUT    HUNGRY
## Estimate -1.9805326 -2.4971689 1.3526367 -0.3791356 1.0658090 1.1185199
## Std.Err   0.1221637  0.1293169 0.1322941  0.1149843 0.1280255 0.1287742
##              WHLDAY
## Estimate 3.2202486
## Std.Err  0.1758936
```

The output above presents the results of the Rasch model on our data and converged after 21 iterations, estimating 7 parameters, since WORRIED is the reference item whose difficulty has been implicitly set to 0.

Analisys of items

Item	Estimate	Interpretation
FEWFOOD	-2.4972	Lowest difficulty; represents a mild form of food insecurity related to lack of dietary diversity. Frequently affirmed even by individuals with low levels of insecurity.
HEALTHY	-1.9805	Also low in severity; inability to eat healthy and nutritious food is one of the earliest signs of food insecurity.
ATELESS	-0.3791	Moderate severity; indicates a reduction in the amount of food consumed.
SKIPPED	1.3526	High difficulty; skipping meals is indicative of more serious food insecurity.
RUNOUT	1.0658	Similar to SKIPPED; refers to running out of food entirely signaling acute food access issues.
HUNGRY	1.1185	High severity; being hungry due to lack of food is a clear indicator of serious deprivation.
WHLDAY	3.2202	Most severe item; going without food for an entire day suggests extreme food insecurity, likely only reported in the most vulnerable cases.

Table 1: Interpretation of Rasch Model Item Difficulty Estimates
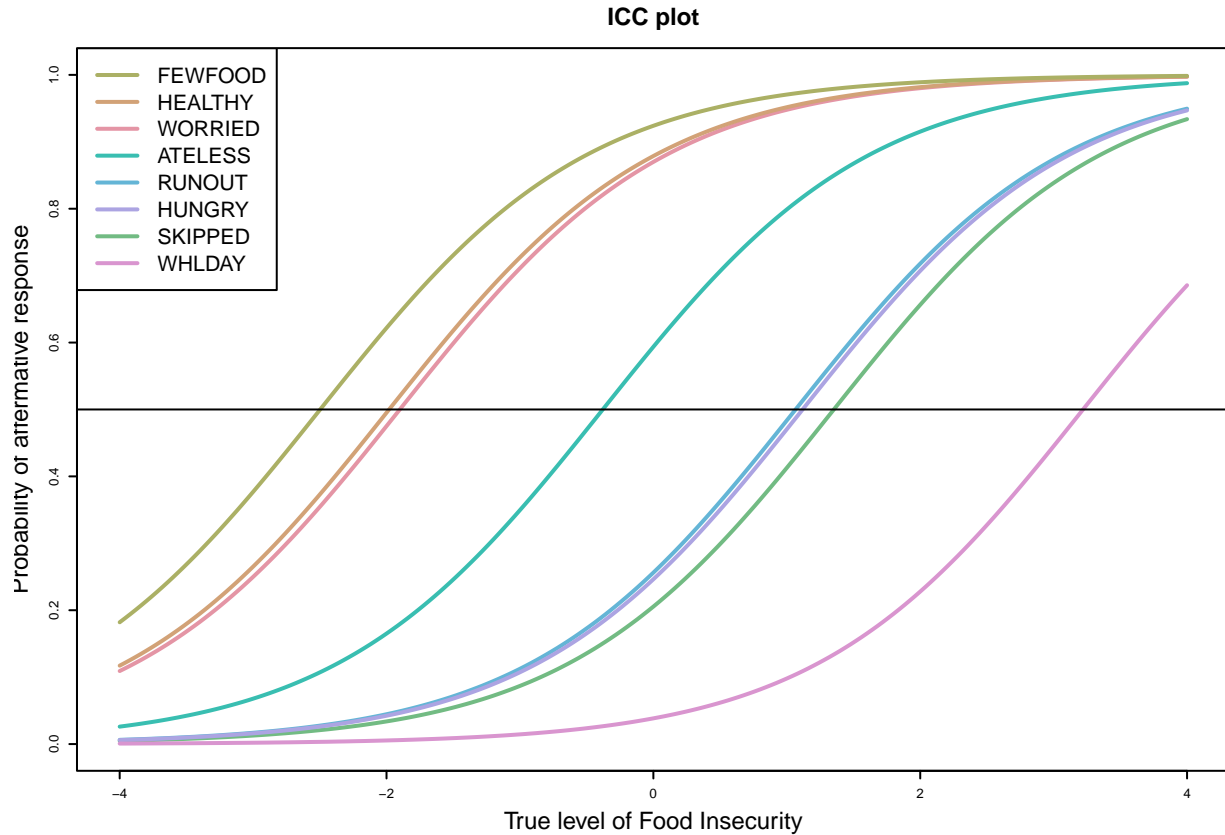
To visualize how each item discriminates along the food insecurity continuum, the plot presents the joint Item Characteristic Curves (ICCs) derived from the Rasch model.

These curves show the probability of an affirmative response to each item across varying levels of the latent trait. The horizontal axis represents the true level of food insecurity, while the vertical axis shows the likelihood of item endorsement. The steepness and position of each curve reflect the severity and discrimination power of the corresponding item.

- The items FEWFOOD, HEALTHY, and WORRIED appear to the left of the plot and rise steeply at lower levels of food insecurity. This indicates they are easier to affirm and reflect milder forms of food insecurity, such as concern about food availability, limited dietary variety, or lack of nutritious food.

- ATELESS, RUNOUT, HUNGRY, and SKIPPED are positioned toward the center of the plot, where the probability of endorsement increases around moderate levels of the latent trait. These items capture more concrete manifestations of food insecurity, such as eating less, running out of food, or experiencing hunger.

- The final item, WHLDAY, appears furthest to the right, suggesting it is the most difficult to endorse, and only affirmed by respondents with the highest levels of food insecurity. This item refers to going an entire day without eating and serves as an indicator of extreme deprivation.

Importantly, the smooth and non-crossing nature of the curves supports the Rasch model assumption of unidimensionality, all items appear to be aligned along a single underlying trait (food insecurity). The gradual progression of curves across the x-axis shows that the scale captures increasing severity levels reliably, from psychological concern to actual food deprivation. Since the probability of answering 'Yes' consistently increases with the level of insecurity, the monotonicity assumption is met too. Overall, this ICC plot provides visual confirmation that the FIES items are well-ordered and function consistently, enabling accurate measurement across a broad spectrum of food insecurity levels.

Analysis of individuals

The number of estimated individuals indicates how many individuals were included in the analysis and received an ability estimate. Note the discrepancy from the original dataset size: 491 people were estimated, but the questionnaire was submitted to 1000 people. This difference is due to people with "extreme scores" (all 0s or all 1s), for whom ability cannot be reliably estimated in the standard Rasch model.

```
## [1] 491
```

```
## [1] -0.3349532
```

The mean value of the latent trait of people in the sample is -0.335, measured on a logit scale. This value means that, on average, the individuals in our sample are slightly below the mean score, 0 logit, of the scale, having assumed sum0=TRUE. On average, the sample is less food insecure than the average item difficulty, thus showing a not very severe level of food insecurity.

In other words, most people in our sample face less severe experiences of food insecurity than the "average severity" of the items describes. It could mean that the scale is more effective in discriminating lower or moderate levels of food insecurity in our sample, while there may be fewer people reaching the higher levels of insecurity covered by the more difficult items.

Table of theta parameters

The table below presents the estimated person parameters for the FIES scale. "RS" represent the number of affirmative responses to FIES scale items. The "Estimate," quantifies the latent level of food insecurity for individuals with a given raw score and it is expressed on the logit scale. A negative value indicates lower insecurity, while a positive value indicates higher insecurity. The "SE" tells us how precise that estimate is. The standard errors reflect measurement precision: generally low for intermediate raw scores, indicating precise estimates for moderate food insecurity. Precision decreases towards scale extremes, with standard errors notably 'NA' for raw scores of 0 and 8. This 'NA' denotes indeterminate person parameter estimates due to perfectly consistent response patterns (all 0s or all 1s), precluding precise localization on the measurement continuum.

RS	Estimate	SE	Interpretation
0	-4.18	NA	This score indicates no affirmed items, suggesting they are very unlikely to experience even the easiest food insecurity items.
1	-3.01	1.18	While still at the lower end, these individuals have affirmed one item, indicating a slight overlap with the easiest items on the scale.
2	-1.88	0.99	These individuals are starting to overlap with more of the easier to moderately difficult items on the FIES scale.
3	-0.93	0.97	These individuals are beginning to overlap with a wider range of items, including those of moderate difficulty.
4	-0.01	0.95	This is a key overlap point, very close to the scale's average (0 logit). Individuals with this score are estimated to have a 50% probability of affirming items of average difficulty. This score often marks a practical threshold for moderate food insecurity.
5	0.88	0.94	These individuals are likely to affirm most easy and moderate items, and are beginning to overlap significantly with more difficult items.
6	1.82	1.01	They affirm most items on the scale, including those of higher difficulty. Their estimated measure overlaps strongly with challenging food insecurity experiences.
7	3.04	1.24	These individuals overlap with almost all FIES items, including the most difficult ones.
8	4.34	NA	These individuals affirmed all items, suggesting an overlap with even the most severe experiences of food insecurity captured by the FIES.

Transformation of Raw Score to Estimated Ability

This graph illustrates the relationship between an individual's total raw score on the food insecurity scale (on the X axis) and his or her estimated level of food insecurity (on the Y axis), expressed in logit units. It is critical because it shows how ordinal raw scores are transformed into severity estimates on a continuous interval scale.

The line in the graph shows a clear increasing monotonic relationship. This means that as the observed raw score increases, thus as the number of reported food insecurity experiences increases, so does the estimated level of food insecurity severity.

**Graph of estimated insecurities as a function of observed scores**
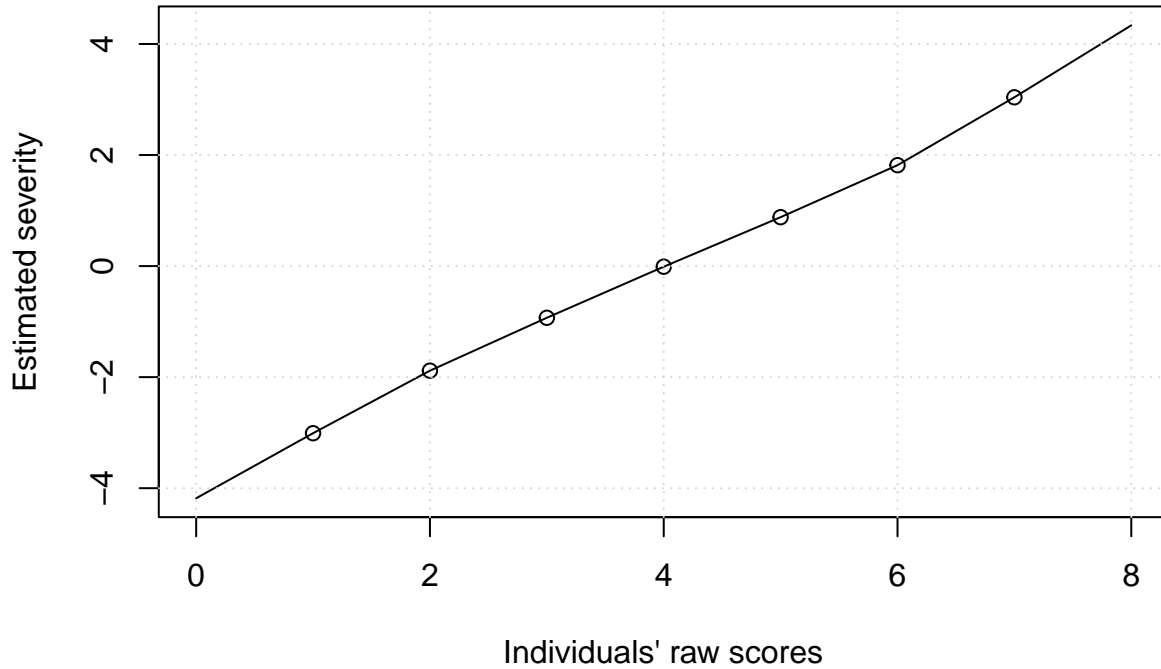
It can be seen that the relationship is not perfectly linear across the scale.

- At the lower ends (low raw scores, e.g., 0 to 2) and upper ends (high raw scores, e.g., 6 to 8), the curve tends to be steeper. This means that even a small change in the raw score (e.g., 0 to 1, or 7 to 8) corresponds to a relatively larger change in the estimated level of insecurity. This reflects the smaller amount of information available at the extremes of the scale to discriminate severity levels.

- In the middle part of the scale (mean raw scores, e.g., 3 to 5), the curve is flatter. Here, each one-point increase in raw score corresponds to a more gradual change in estimated seveirity. This is where the scale, and its component items, generally provide the most information and measurement accuracy.

The graph confirms that the FIES scale functions as expected: individuals with more experiences of food insecurity are estimated to have higher levels of food insecurity severity.

Person Item Map

The figure presents a Person-Item Map. This map plots the estimated item difficulty parameters (denoted β_j) along the vertical axis and overlays the distribution of person parameters (denoted θ_i) along the same latent scale.

Person–Item Map
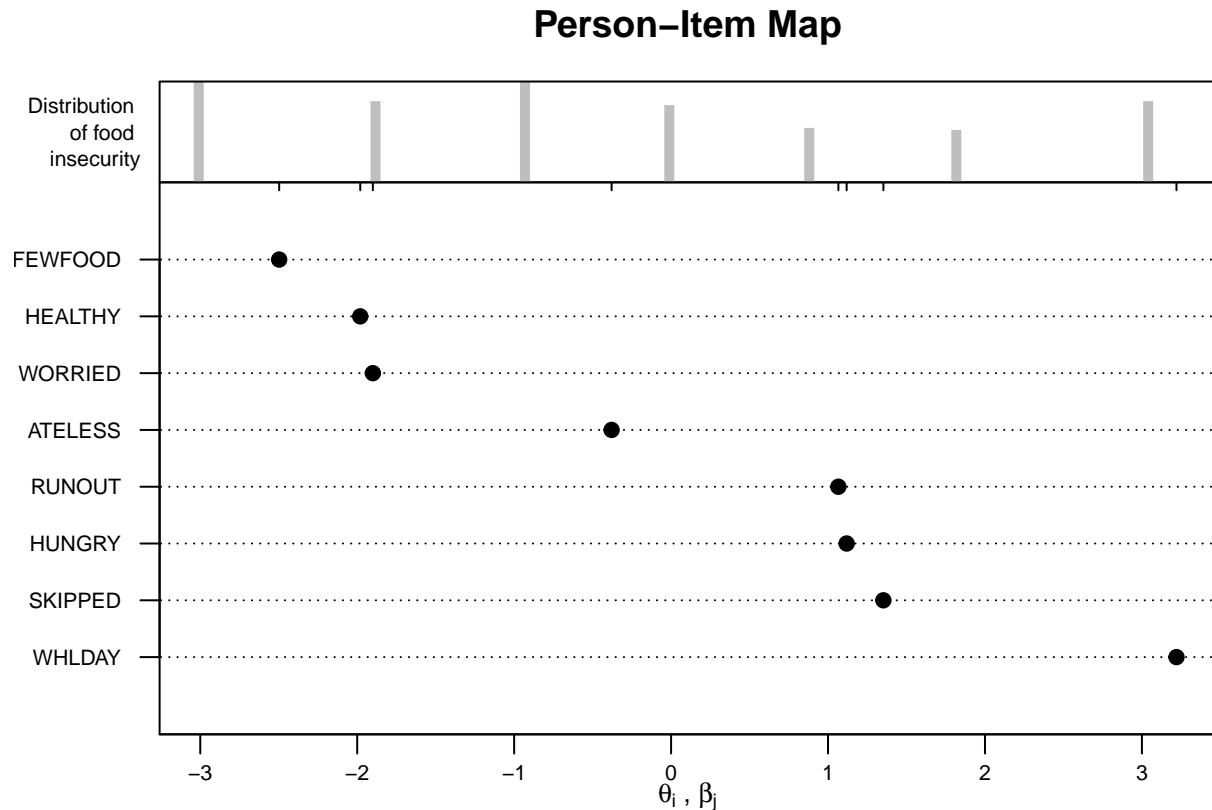
At the bottom of the graphic, the x-axis shows the values of both items and persons along the latent trait of food insecurity severity. The vertical gray bars at the top represent the distribution of respondents (i.e., the density of estimated person severities), while the black dots on the horizontal lines show the location of each item in terms of its difficulty.

- Items such as FEWFOOD, HEALTHY, and WORRIED are placed on the left side of the scale (negative values), indicating they are easier to endorse—appropriate for detecting mild food insecurity.

- On the other hand, items like WHLDAY, SKIPPED, and HUNGRY appear on the right, suggesting they are more severe—endorsed primarily by individuals experiencing more extreme forms of food insecurity.

Diagnostic of the model

In order to evaluate how well the model fits the data, we carried out a global model fit assessment through the Likelihood Ratio Test (LR-Test). This test compares our Rasch model to a saturated or other nested model. A p-value of 0.009, shows that the model fits the data well, as we can see by the output below.

```
##
## Andersen LR-test:
## LR-value: 18.613
## Chi-square df: 7
## p-value:  0.009
```

Interpretation of Item Fit Statistics

The table presents diagnostic statistics from the Rasch model for each of the eight items included in the Food Insecurity Experience Scale (FIES). These indicators are essential to evaluate how well each item fits the expectations of the Rasch model and contributes to the measurement of the latent trait food insecurity severity.

14

```
## 
## Itemfit Statistics:
##              Chisq  df p-value Outfit MSQ Infit MSQ Outfit t Infit t Discrim
## WORRIED  593.845 490   0.001      1.209     1.196    0.923   3.081   0.291
## HEALTHY  693.186 490   0.000      1.412     0.763    1.592  -4.241   0.479
## FEWFOOD  885.886 490   0.000      1.804     0.869    2.177  -2.134   0.354
## SKIPPED  289.407 490   1.000      0.589     0.814   -2.304  -2.480   0.633
## ATELESS  412.146 490   0.995      0.839     0.838   -1.305  -2.773   0.626
## RUNOUT   286.754 490   1.000      0.584     0.677   -2.712  -4.744   0.736
## HUNGRY   260.962 490   1.000      0.531     0.608   -3.067  -5.916   0.769
## WHLDAY  1133.040 490   0.000      2.308     1.044    2.003   0.510   0.153
```

The chi-square test checks whether observed responses significantly deviate from model expectations.

- Items such as FEWFOOD, HEALTHY, and WHLDAY show significant misfit (p < 0.01), meaning they behave less predictably according to the model.

- In contrast, SKIPPED, ATELESS, RUNOUT, and HUNGRY show excellent fit (p-values ≈ 1.000), indicating strong alignment with Rasch assumptions.

Infit and Outfit are used to test hypotheses of items' same discrimination power. Infit is a weighted mean square residual fit statistic sensitive to responses near the item's severity level. It primarily assesses the internal consistency of an item within the model. The recommended interpretation range is:

- 0.8–1.2 = good fit

- 0.7–1.3 = acceptable fit

- < 0.7 or > 1.3 = potential misfit

Items like HUNGRY (0.608) and RUNOUT (0.677) are slightly below the acceptable range, suggesting overfit, too little unpredictability, or maybe redundancy. On the other hand, FEWFOOD (0.869) and WHLDAY (1.044) remain within acceptable bounds, while HEALTHY (0.763) is borderline low.

Outfit is an unweighted mean square residual fit statistic particularly sensitive to unexpected extreme responses. It provides a measure of overall, external consistency. Ideal values must be < 2.0. Here only WHLDAY (2.308) exceeds the threshold, suggesting potential issues with this item, perhaps due to misinterpretation or poor targeting. FEWFOOD (1.804) and HEALTHY (1.412) are high but still within tolerable limits. Several items, including HUNGRY (0.531), RUNOUT (0.584), and SKIPPED (0.589), show low Outfit MSQ, indicating responses that are overly predictable.

Most items show acceptable Infit and Outfit values, indicating a generally good model fit. However, WHLDAY is a clear outlier in terms of Outfit, suggesting it may introduce noise or not work as intended. Items such as HUNGRY, RUNOUT, and SKIPPED consistently show low infit indices, implying possible redundancy or limited discrimination.

Discrimination reflects how well each item differentiates between respondents with different levels of food insecurity. The higher values (closer to 1.0) are better and the values below 0.3 are considered weak. In this case, HUNGRY (0.769), RUNOUT (0.736), and SKIPPED (0.633) show strong discrimination. Instead, WHLDAY, despite being the most severe item, has very low discrimination (0.153), indicating it contributes little to distinguishing between levels of food insecurity.

Reliability and separability

These two indices measure how well the scale is able to distinguish between different levels of people's abilities and different item difficulties.

Our analysis showed good reliability in estimating individuals' ability (reliability = 0.784), suggesting that responses consistently reflect levels of food insecurity. However, the separation index (1.906) indicates a

moderate ability of the instrument to distinguish between different levels of severity, confidently identifying only two main groups, namely low vs. high insecurity.

Reliability (individuals): 0.784

Separability index (individuals): 1.906

In addition, our model showed excellent reliability in estimating item parameters (reliability = 0.996), indicating that the sample is highly informative and allows for accurate placement of questions along the food insecurity severity scale. The item separation index (15.023) suggests the presence of many distinct levels of difficulty among the questions, confirming the instrument's ability to finely differentiate the nuances of respondents' food experience. A very high value relative to the number of items, as in this case, may also be influenced by sample size or low residual variability in the estimates.

Reliability (item): 0.996

Separability Index: 15.023

Differential Item Functioning

DIF occurs when an item performs differently for different subgroups of people. In food insecurity, an item exhibits DIF if people with the same level of food insecurity have a different probability of responding affirmatively to the item, depending on their individual characteristics. Here we have considered the variables: gender, area of residence, income quintiles, and level of education.

Table 2: Results of Anderson LR-test for checking DIF

	GENDER	**URBAN**	**INCOME**	**EDUCATION**
Andersen LR-test	8.114	52.781	41.986	32.721
Chi-square df	7	21	28	14
p-value	0.323	0	0.043	0.003

Andersen's test checks whether the item parameters estimated by the Rasch model are invariant between subgroups of respondents. A high p-value (> 0.05) suggests no DIF, i.e., item invariance between groups. A low p-value (< 0.05) indicates presence of DIF, i.e., that items do not function equally across groups. It can be seen from the table that there is good equity of the test between genders. In contrast, for the other variables the DIF seems significant, indicating that the test does not measure the same severity across groups. In this case, some questions may be more or less severe for certain groups not because of real differences in the observed phenomenon (e.g., food insecurity), but because of cultural, socioeconomic or interpretive factors.

Prevalence Rates

In the context of the FIES, 'prevalence rates' refer to the percentage of the population that falls into each of the different levels of food insecurity severity. It is designed to classify individuals into four categories of food insecurity:

- Safe (Food Secure)
- Mild (Mild Food Insecurity)
- Moderate (Moderate Food Insecurity)
- Severe (Severe Food Insecurity)

Once individual severities are estimated through the Rasch model, these continuous estimates are then used to classify people into the four prevalence categories mentioned above. Prevalence rates were computed for each item to describe the frequency with which participants experienced food insecurity situations. The results

confirm the Rasch model estimates, showing higher prevalence for the 'easier' items and lower prevalence for the 'harder' items

Table 3: Prevalence rates for food insecurity items

Item	PR	Description
FEWFOOD	0.510	51.0% said they had access to little food – most frequent item, it signals a fairly common difficulty.
HEALTHY	0.471	47.1% said they were unable to eat healthily and variously – symptom of qualitative food insecurity.
WORRIED	0.465	46.5% were worried about not having enough food – showing food anxiety.
ATELESS	0.346	34.6% reported eating less than they needed – the first sign of true food restriction.
RUNOUT	0.250	25.0% ran out of food supplies – indicating a critical situation.
HUNGRY	0.247	24.7% felt hungry but did not eat – a true deprivation experience.
SKIPPED	0.234	23.4% skipped meals due to lack of resources – a sign of moderate hardship.
WHLDAY	0.139	Only 13.9% fasted an entire day – it is the most severe and least frequent item, reflecting severe food insecurity.

Items are distributed in a consistent way with the progression of severity: worry → reduced quality → reduced quantity → hunger → fasting. The decreasing prevalence shows a well-functioning scale and clear differentiation in levels of insecurity. The fact that more than 50% report FEWFOOD and nearly half HEALTHY and WORRIED shows a significant prevalence of the problem, although more severe cases (e.g., WHLDAY) are less common.

Prevalence Rates for each Item
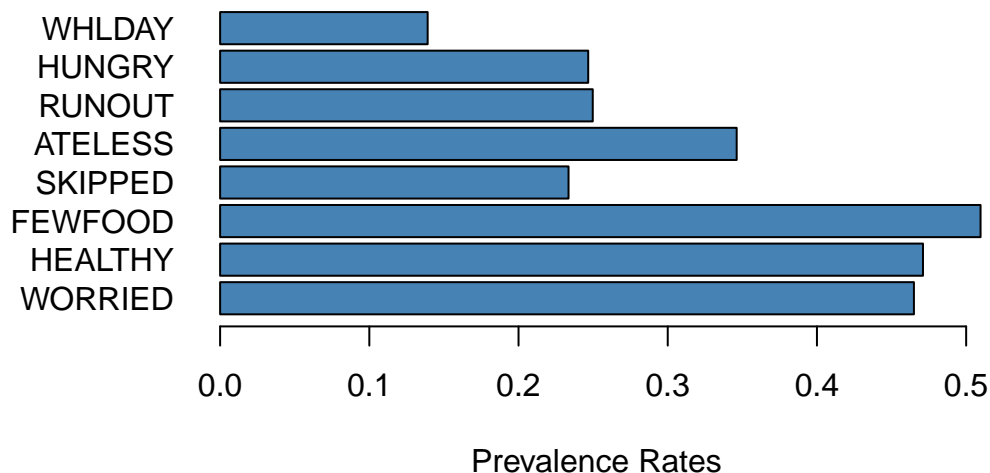
Prevelance Rate Gender

Gender comparisons show generally higher prevalence rates for female than male in almost all items on the scale. This suggests that women are slightly more likely to experience food insecurity, a phenomenon consistent with evidence indicating greater female socioeconomic vulnerability in many contexts.

Prevalence Rate per item and group – Gender
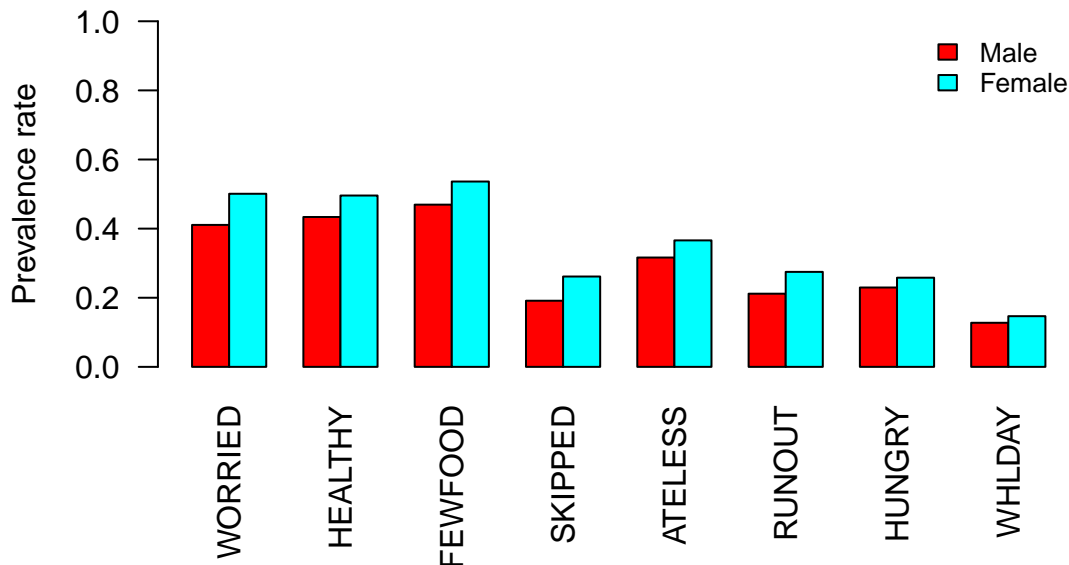
Prevalence Rate Urban

Prevalence rates show a geographic gradient: individuals living in rural and non-metropolitan areas report higher levels of food insecurity than those living in urban areas or the capital. This result may reflect spatial inequalities related to access to services, economic opportunities, and support networks.

Prevalence Rate per item and group – Area
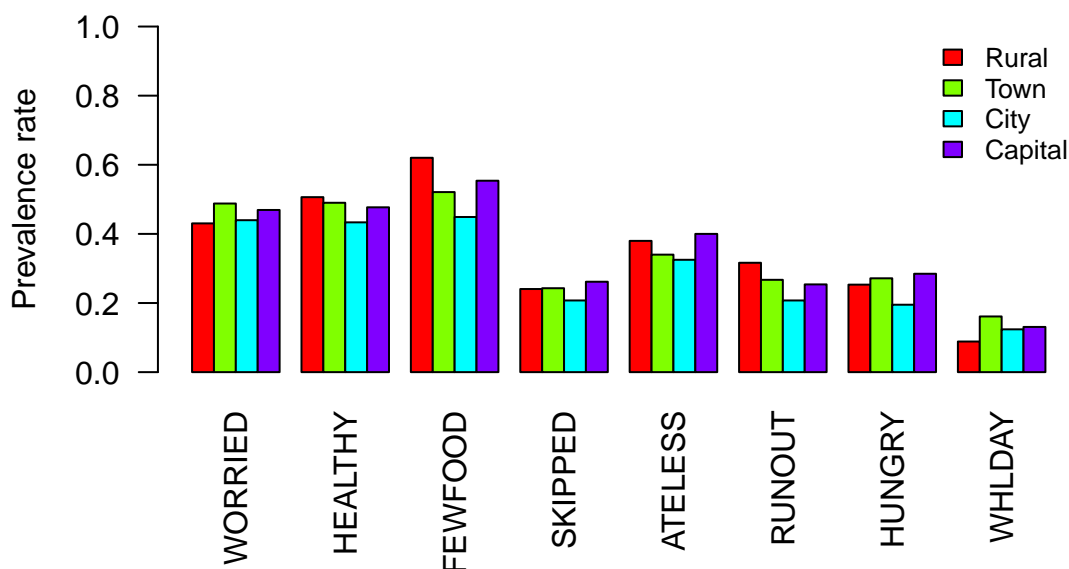
Prevalence Rate Income

The distribution by income quintiles shows an inverse relationship between income level and prevalence of food insecurity. Rates are highest among the "Poorer" and "Low Income" groups, and progressively decrease until the "Richer" group shows minimum values for all items. This trend confirms the strong link between economic hardship and difficulty in accessing food.

Prevalence Rate per item and group – Income Quintiles
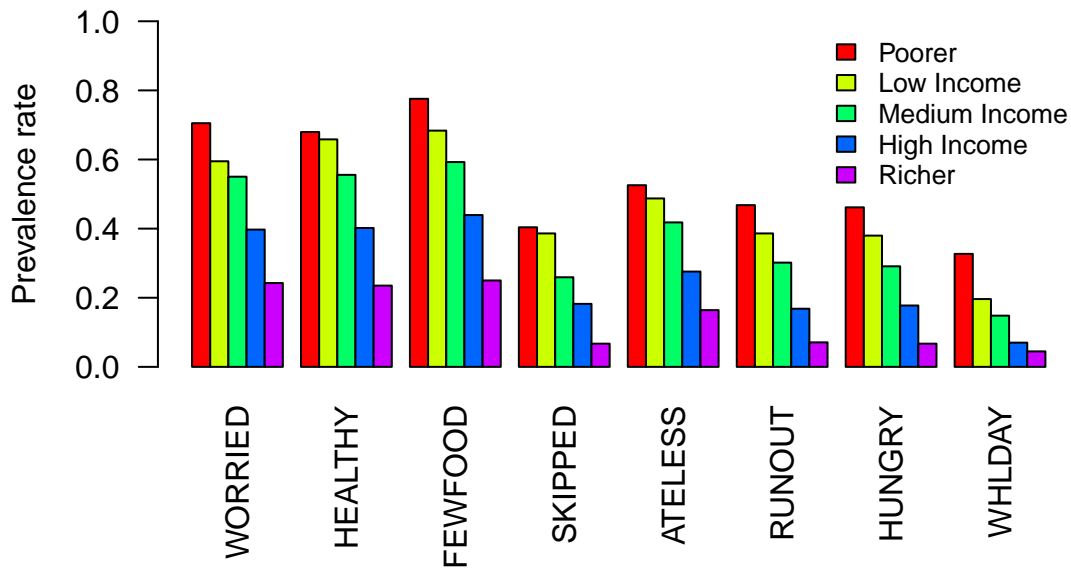
Prevalence Rate Education

Educational level also appears to be strongly associated with food insecurity. Those with low education have significantly higher prevalence rates, while those with high education report much lower values. This highlights how education, probably in relation to economic and cultural capital, plays a protective role with respect to food risk.

Prevalence Rate per item and group – Education Level
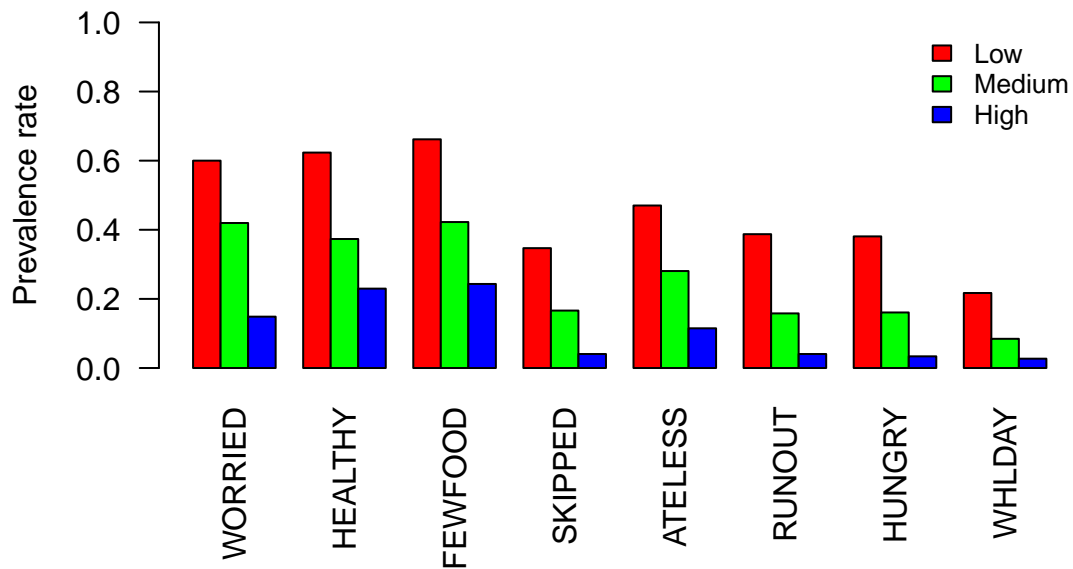
Conclusion

In conclusion, this project provides a comprehensive psychometric validation of the Food Insecurity Experience Scale (FIES) in Albania for the year 2017, leveraging microdata from the FAO and employing the Rasch model. The analysis confirms the FIES scale's robust psychometric properties, demonstrating excellent internal consistency and clear unidimensionality. Item Characteristic Curves further illustrate a logical progression of severity across the eight FIES items, supporting their capacity to reliably measure the latent construct of food insecurity.

The Rasch model estimation successfully determined item difficulty parameters, revealing that items like "FEWFOOD" and "HEALTHY" represent milder forms of food insecurity, while "WHLDAY" signifies extreme deprivation. While the model generally fits the data well, as indicated by a p-value of 0.009 from the Likelihood Ratio Test, some items, particularly "WHLDAY," show potential misfit or low discrimination, suggesting areas for further refinement or careful interpretation in future applications.

Despite the robust overall fit, the analysis of Differential Item Functioning (DIF) highlights significant inequalities across geographic areas, income quintiles, and educational levels, indicating that the FIES may function differently for these subgroups. Conversely, the scale demonstrates good equity across genders. These findings underscore the importance of disaggregated analyses for policy formulation to address specific vulnerabilities.

Finally, the estimated prevalence rates confirm a high incidence of mild food insecurity in Albania, with over 50% of respondents reporting limited dietary diversity ("FEWFOOD"). Severe food insecurity, while less common, still affects a significant subset of the population. The observed gradients by gender, urbanicity, income, and education consistently point to higher food insecurity among females, rural residents, lower-income households, and less educated individuals, reinforcing existing socioeconomic disparities. These insights are crucial for monitoring progress towards Sustainable Development Goals and for designing targeted social protection programs and food interventions in Albania.