UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II
DIPARTIMENTO DI SCIENZE POLITICHE

CORSO DI LAUREA MAGISTRALE IN
SCIENZE STATISTICHE PER LE DECISIONI

TESI DI LAUREA IN
MODELLI LINEARI GENERALIZZATI

ITEM SELECTION, PENALIZED ESTIMATING FUNCTION AND VARIABLE SELECTION IN ITEM RESPONSE THEORY MODELS

CANDIDATO:
ANTONIO COLA
MATRICOLA:
M10/399

RELATRICE:
CH.MA PROF.SSA MARIA IANNARIO
CORRELATORE:
CH.MO PROF. ALFONSO IODICE D'ENZA

ANNO ACCADEMICO 2022/2023

To Ilaria

Contents

Introduction

In the field of psychometrics and quantitative assessment of individual responses, the Partial Credit Model (PCM) represents a polymorphic and versatile model, adaptable for analyzing ordinal data and investigating complex latent constructs. This thesis is focused on the in-depth exploration of the PCM, with a particular focus on item selection using 'bruteforce' and 'backward' methods. These approaches have been analyzed to determine their contribution in selecting the most informative items and constructing more robust and parsimonious item response models.
Attention then shifts to covariates in the PCM, emphasizing the importance of integrating contextual variables that can provide additional explanations for response variability and enrich the understanding of the latent construct under examination. In addition, the penalized estimation function is examined as a means to control model complexity and prevent overfitting, comparing classical and new penalties to determine their impact on model fit.
A simulation chapter is dedicated to validating the penalization techniques. The thesis proceeds with the application of PMC models, delving into item selection and model fitting based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), as well as log-likelihood (logLik) results. The discussion extends to a comparison between selected PCM models, highlighting the effectiveness of different penalty terms and their ability to improve model precision and parsimony.
In conclusion, this thesis aims to provide a significant contribution to the field of psychometric measurement and quantitative assessment by optimizing item selection and integrating new penalization strategies within the Partial Credit Model, with a specific application to the 'Earthquake Stress' construct.

1 Item Response Theory Models in Ordinal Data

> In numerous applications, items are characterized by more than two response categories, which are arranged in a specific order. Consider items scored polytomously, represented by values Y_{pi} which belong to the set $\{0,1,\ldots,k\}$, where p ranges from 1 to P, and i from 1 to I. The set $\{0,1,\ldots,k\}$ symbolizes the sequential arrangement of these categories. For the sake of clarity, we adopt a constant number of response categories, $k+1$, when formulating the models. However, it should be noted that the actual number of response categories may differ among items, with the specific categories for item i being denoted as $\{0,1,\ldots,k_i\}$.

All ordinal models must differentiate between low categories, which typically represent weak performance, and high categories, indicating strong performance, especially in the context of achievement tests. As a result, ordinal models can be constructed by leveraging binary models that distinguish between weak and strong performance. The three primary types of ordinal models commonly used can be characterized based on how they utilize binary models as their building blocks. When dealing with a response variable categorized into ordered levels $\{0,1,\ldots,k\}$, there are various approaches to creating an ordinal model using binary models. Binary models can be employed to compare specific categories or groups of categories within $\{0,1,\ldots,k\}$. Specifically, one can:

- Compare groups of categories resulting from splitting the categories into subsets like $\{0,1,\ldots,r-1\}$ and $\{r,\ldots,k\}$;
- Compare (conditionally) between two categories, such as adjacent categories $\{r-1,r\}$;
- Compare (conditionally) between a category and a set of adjacent categories, for instance, $\{r-1\}$ and $\{r,\ldots,k\}$.

The various methods for comparing categories correspond to the most commonly employed ordinal models, namely cumulative models, adjacent categories, and sequential models. Table 1 illustrates the three types of dichotomizations that define these models (G. Tutz 2022) [33]. With an ordinal response variable Y taking values from $\{0,1,...,k\}$, the variables $Y^{(r)}$ represent the binary variables included in the model. Brackets indicate which response categories are involved in the dichotomization process, and the '|' symbol delineates the (conditional) split. The table displays which binary models are components of ordered latent trait models but can also be utilized to construct models using binary variables. Figure 1 distinguishes between conditional and non-conditional models (G. Tutz 2022) [33]. Among these models, the cumulative or graded response model is the only non-conditional one, while the other two employ some form of conditioning when constructing ordinal models from binary ones.

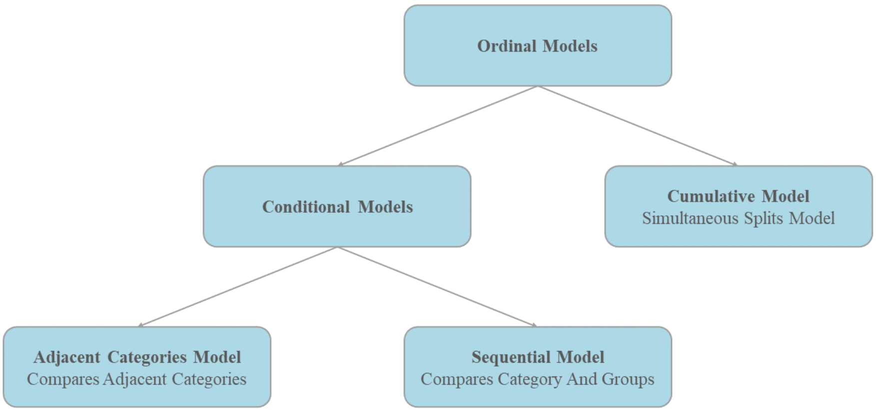

Figure 1: Structure of classical ordinal latent trait models

Cumulative-Type Model, Dichotomization into Groups

$$[0,\dots,r-1|r,\dots,k] \quad Y^{(r)} = \begin{cases} 1 & Y \geq r \\ 0 & Y < r \end{cases}$$

Adjacent-Type Model, Dichotomization given $Y \in \{r, r+1\}$

$$0,\dots,[r-1|r],\dots,k \quad Y^{(r)} = \begin{cases} 1 & Y \geq r \\ 0 & Y < r \end{cases} \quad \text{given} \quad Y \in \{r, r+1\}$$

Sequential-Type Model, Dichotomization given $Y \geq r$

$$0,\dots,[r-1|r,\dots,k] \quad Y^{(r)} = \begin{cases} 1 & Y \geq r \\ 0 & Y < r \end{cases} \quad \text{given} \quad Y \geq r-1$$

Table 1: Types of Ordinal Models for Response $Y \in \{0, 1, \dots, k\}$ and Dichotomous Variables $Y^{(r)}$ that are Contained (G. Tutz 2011)

Before delving into the diverse models derived from employing binary models as fundamental components, it is beneficial to introduce a more structured representation of how categories are partitioned into sets of neighboring categories, thereby distinguishing between weaker and stronger performance. These split variables can be formally defined as follows:

$$Y_{pi}^{(r)} = \begin{cases} 1 & Y_{pi} \geq r \\ 0 & Y_{pi} < r. \end{cases}$$

The split variables are binary variables designed to divide the categories into two subsets, namely $\{0,\ldots,r-1\}$ and $\{r,\ldots,k\}$, based on a specific definition of weak and strong performance. With response categories {0,1,. . . ,k}, there are a total of $k-1$ split variables denoted as $Y_{pi}^{(1)},\ldots,Y_{pi}^{(k)}$. These split variables hold significance because they exhibit a one-to-one correspondence with the ordinal response, as expressed by:

$$Y_{pi} = r \quad \Leftrightarrow \quad (Y_{pi}^{(1)},\ldots,Y_{pi}^{(r)},Y_{pi}^{(r+1)},\ldots,Y_{pi}^{(k)}) = (1,\ldots,1,0,\ldots,0)$$

Here, $Y_{pi}^{(r)}$ represents the last binary variable within the sequence that has a value of 1. The sequence $Y_{pi}^{(1)},\ldots,Y_{pi}^{(k)}$ can be viewed as a specific coding scheme for the multicategorical response Y_{pi}.
These sequences, denoted as $(Y_{pi}^{(1)},\ldots,Y_{pi}^{(k)})$, of split variables possess a unique characteristic where a sequence of ones is followed by a sequence of zeros, resembling a Guttman scale. In a broader context, random variables $Y_{pi}^{(1)},\ldots,Y_{pi}^{(k)}$ with $Y_{pi}^{(r)} \in \{0,1\}$ and $Y_{pi}^{(r)} \geq Y_{pi}^{(r+1)}$ have been referred to as Guttman variables, as they exhibit a Guttman structure.

1.1 Item Selection

Item selection is a crucial component in constructing Item Response Theory (IRT) models used to measure individuals' abilities and competencies. The main goal of this phase is to identify an optimal set of items that maximizes the quality of measurements while simultaneously minimizing computational burden. The items subject to selection are not included in psychometric scales, and this methodology is applicable only in cases where the latent trait is unidimensional. In this section, we introduce a new algorithm that offers a flexible solution to address the challenge of item selection in IRT models. This algorithm employs two distinct approaches: the 'bruteforce' method and the 'backward' method.
The 'bruteforce' method represents an exhaustive approach that considers all possible combinations of available items to determine the best model. However, its effectiveness is limited by its computational slowness, as the runtime grows exponentially with the increase in the number of items. This method proves to be particularly suitable when the number of items to evaluate is relatively limited ($I \leq 10$).
On the other hand, the 'backward' method is a more parsimonious strategy that proves effective when dealing with datasets with a high number of items ($I > 10$). This approach, instead of searching all possible combinations, works by progressively reducing the set of considered items, retaining those that contribute most to the model's information. This allows for accurate and reliable results without the computational burden associated with the 'bruteforce' method.

1.1.1 Bruteforce Method

The process begins with an initial dataset, from which all possible items combinations are generated. For each combination, we assess internal consistency through the calculation of

Cronbach's alpha.

The Cronbach's alpha, or Cronbach's alpha coefficient, is a measure of internal reliability of a psychometric measurement tool (Cronbach 1951) [9]. It is primarily used in the field of research and psychological assessment to assess the internal consistency of a set of questions or items that constitute a measurement tool, such as a questionnaire or a test. Cronbach's alpha measures how different questions or items within the tool are correlated with each other, providing an indication of internal cohesion. A higher alpha value indicates greater internal consistency, suggesting that the questions effectively measure the same characteristic or construct.

$$\alpha = \frac{I}{I-1}\left(1 - \frac{\sum_{i=1}^{I} \sigma_{y_i}^2}{\sigma_y^2}\right)$$

Where I represents the number of items, $\sigma_{y_i}^2$ the variance associated with each item i and σ_y^2 the variance associated with the total scores $\left(y = \sum_{i=1}^{I} y_i\right)$.
A Cronbach's alpha value of 0.7 is generally considered quite good, keeping in mind that the latter is: $0 < \alpha < 1$. This value suggests that the questions or items included in the instrument are consistent with each other and reliably measure the same characteristic or construct.

If Cronbach's alpha is below the accepted standard of 0.7, the items combination is discarded and the process is repeated with the next combination. When a combination meets or exceeds the threshold value of 0.7, we proceed to apply an IRT model.
From the IRT model, we extract three main infromations: the Bayesian Information Criterion (BIC), Info Ratio, and Info Tot.

The BIC helps us select the optimal model based on a likelihood function that penalizes models with an excessive number of estimated parameters (Schwarz 1978) [26].

$$BIC = -2log(\hat{L}) + m\log(P)$$

Where $\hat{L}$ is the maximized value of the likelihood function of the model, m is the number of parameters in the model, P is the number of observations (or number of respondents).
The Info Ratio, defined as the ratio between the number of items with an average information above one and the total number of items considered.

$$InfoRatio = \frac{I_{InfoMean>1}}{I}$$

Where $I_{InfoMean>1}$ is the number of items with an average information greater than one

and I is the number of items considered in the model.
The Info Tot, which represents the sum of the average information of individual items, serve as indicators of the model's informational quality. These metrics are calculated for each item combination.

$$InfoTot = \frac{\sum_{i=1}^{I} \sum_{j=1}^{P} Info_{ij}}{P}$$

Where $Info_{ij}$ is the value of information that the i-th item and the j-th person contribute to the latent variable.

The process of extracting and evaluating the informations is carried out for all possible items combinations. Once the cycle for all combinations is completed, we select the model that presents the most favorable informations. This best model is then paired with a dataframe that summarizes the results of the items selection and the model evaluation informations.
It can be confirmed that for I items, the number of combinations (n_{comb}) considered is as follows:

$$n_{comb} = \sum_{x=2}^{I} \frac{I!}{x!(I-x)!}$$

The result is a rigorous and systematic items selection method that ensures the inclusion of consistent and informative items in the IRT model, thereby facilitating the creation of reliable and valid measurement tools. This approach allows for a stringent selection based on solid statistical informations, ensuring that the items included in the model effectively contribute to the measurement of the construct of interest.

1.1.2 Backward Method

The algorithm starts with a set of I initial items to evaluate. Subsequently, for each possible combination of $I-1$ items, an IRT model is estimated. The InfoRatio is calculated for each combination, which is the ratio between the number of items with an average information greater than or equal to one and the total number of items considered. Similarly, Cronbach's alpha is calculated to measure the internal consistency of individuals' responses to the items. All combinations of items are then sorted in descending order in a data frame, first based on the InfoRatio and then on Cronbach's alpha. The algorithm proceeds until the desired condition is met, which is when the best model (the first one in the data frame) has an InfoRatio equal to 1 and a Cronbach's alpha greater than 0.7. Otherwise, the item not included in the best model is discarded, reducing the number of items considered by 1, and the process is restarted by considering all combinations of $I-1$ items. The algorithm stops when the desired condition is met or when the number of items considered, which decreases by 1 in each cycle, is reduced to 2.
Before proceeding with combinations, a complete model with all I items is estimated, and only after evaluating it, if necessary, the algorithm proceeds with the combinations.

It can be verified that for I items, the maximum number of combinations ($n_{maxcomb}$) considered is as follows:

$$n_{maxcomb} = 1 + \sum_{x=3}^{I} \frac{x!}{(x-1)!(x-x+1)!} = 1 + \sum_{x=3}^{I} x = 1 + \left(\frac{(I+3)(I-2)}{2} \right) =$$

$$= \left(\frac{I^2 + I - 6}{2} \right) + 1 = \frac{I^2 + I - 4}{2}$$

However, in many experiments, the algorithm identifies the best model by reducing the number of items to not more than half. Therefore, it is more plausible that the number of combinations considered is:

$$n_{comb} = \frac{3I^2 + 6I + 8}{8} \quad \text{If } I \text{ is even,} \qquad n_{comb} = \frac{3I^2 + 8I + 5}{8} \quad \text{If } I \text{ is odd.}$$

This algorithm provides a structured and efficient method for item selection in IRT contexts, facilitating the search for optimal item combinations that improve the precision of individual ability measurements. Its utility becomes particularly evident when dealing with large sets of items.

1.1.3 Method Comparison

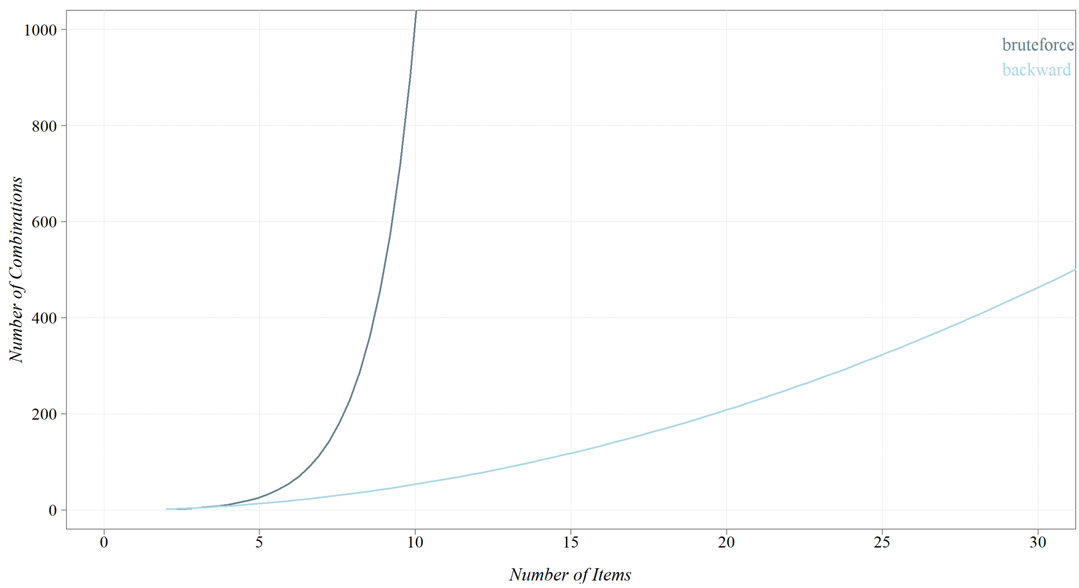

Figure 2: Comparison between Bruteforce and Backward Methods

The graph presents a comparison between the two item selection methods, 'bruteforce' and 'backward', regarding their impact on the number of possible combinations as the number of items varies. As can be observed, the 'bruteforce' method exhibits an exponential increase in the number of combinations as the number of items grows. This phenomenon is expected, as 'bruteforce' examines all possible item combinations, and the number of combinations grows exponentially with the increase in the number of items. Consequently, the computation time for this method can become prohibitive even with a relatively small number of items, limiting its practical applicability, especially in large datasets. In stark contrast, the 'backward' method shows a much more moderate growth in the number of combinations. This method, which starts with a complete model and then progressively removes less significant items, is clearly more efficient in terms of computation time, allowing the analysis of datasets with a greater number of items without encountering the computational hurdle imposed by the 'bruteforce' method. Efficiency in computation time is not only a matter of convenience but also feasibility. In research contexts where time is a precious resource and datasets are extensive, the use of the 'backward' method becomes essential. The ability to conduct analyses in a reasonable amount of time without sacrificing model accuracy makes the 'backward' method the superior choice for item selection in many practical applications.

In conclusion, while the 'bruteforce' method may be considered for small datasets or when maximum accuracy is indispensable despite the high computational cost, the 'backward' method emerges as the preferred choice for most applications due to its optimal balance between computational efficiency and precision in item selection.

2 The Partial Credit Model

The Partial Credit Model (PCM) stands out for its flexibility and relative simplicity, making it an attractive choice for many educational and psychometric applications (Masters 1982) [15] (Masters 1984) [16]. Unlike the Graded Response Model (GRM) (Samejima 1969) [23] (Samejima 2016) [24], which requires the estimation of multiple item threshold parameters, the PCM operates with a single item difficulty parameter, simplifying the estimation and interpretation process. The decision to delve into the Partial Credit Model is motivated by its greater simplicity and lower computational complexity in parameter estimation, making it more accessible. Furthermore, the PCM is well-suited for situations where test or questionnaire items are designed to measure gradations of skills or attitudes, offering a more flexible and intuitive modeling of ordered responses.

A method to derive an ordinal model involves the conditional application of binary models. An approach aligning with the sequential nature of categories is to employ binary models that differentiate between adjacent categories, contingent on the response falling within these specific categories. This technique naturally leads to the formulation of the general adjacent categories model:

$$P(Y_{pi} = r | Y_{pi} \in \{r-1, r\}, \theta_p, \alpha_i, \delta_{ir}) = F(\alpha_i(\theta_p - \delta_{ir})), \quad r = 1, \dots, k \tag{1}$$

The general adjacent categories model encompasses three distinct types of parameters:

- Person Parameters θ_p: These parameters represent individual ability or attitude. They play a crucial role in modeling how the characteristics or capabilities of persons influence their responses;
- Discrimination Parameters α_i: These parameters determine the slopes of the response functions. They are indicative of how sensitively an item differentiates between persons with different levels of the trait being measured (e.g., ability or attitude);
- Location Parameters δ_i: These parameters are on the same scale as the person parameters. They signify the positioning or 'location' of each item on the continuum trait, which is selected, among thems, for robustness issues (M. Iannario et al 2017) [12] essentially providing a reference point against which person abilities or attitudes are measured.

One of the commonly used models in the realm of the IRT framework for ordinal data is the Partial Credit Model, which incorporates the logistic distribution function $F()$. Given its prominence in the field, our subsequent discussion will primarily concentrate on this model. The generalized version of the partial credit model can be articulated as follows:

$$P(Y_{pi} = r | \theta_p, \alpha_i, \delta_i) = \frac{exp(\sum_{l=1}^{r} \alpha_i(\theta_p - \delta_{il}))}{\sum_{s=0}^{k} exp(\sum_{l=1}^{s} \alpha_i(\theta_p - \delta_{il}))}, \quad r = 1, \dots, k \tag{2}$$

To simplify notation, the model's definition implicitly incorporates the term $\sum_{r=1}^{k}(\theta_p - \delta_{ir})$. Additionally, the item location parameters are consolidated into the vector $\delta_i^T = (\delta_{i1}, \ldots, \delta_{ik})$. The formula (3) defines the probabilities of responses, while the broader structure of adjacent categories models, as outlined in (1), describes the conditional binary models that underpin the overall model. It is evident that the general partial credit model aligns with the framework of an adjacent categories model, as demonstrated by the following analysis:

$$\begin{gathered} P(Y_{pi} = r | Y_{pi} \in \{r-1, r\}, \theta_p, \alpha_i, \delta_{ir}) = \frac{\pi_{pir}}{\pi_{pi,r-1} + \pi_{pir}} = \\ = \frac{exp(\alpha_i(\theta_p - \delta_{ir}))}{1 + exp(\alpha_i(\theta_p - \delta_{ir}))} = F(\alpha_i(\theta_p - \delta_{ir})) \end{gathered}$$

Where $\pi_{pir} = P(Y_{pi} = r | \theta_p, \alpha_i, \delta_{ir})$ and $F(\cdot)$ is the logistic function. Deriving the partial credit model (3) from the general model (1) is more challenging, but this is feasible when the logistic function is selected for $F(\cdot)$.

When $k = 2$, we obtain the most basic form of the partial credit model that transcends binary categorization. The probabilities in this scenario are defined as follows:

$$\begin{aligned} &P(Y_{pi} = 0 | \theta_p, \alpha_i, \delta_i) = 1/\kappa_{pi}, \\ &P(Y_{pi} = 1 | \theta_p, \alpha_i, \delta_i) = exp(\alpha_i(\theta_p - \delta_{i1}))/\kappa_{pi}, \\ &P(Y_{pi} = 2 | \theta_p, \alpha_i, \delta_i) = exp(\alpha_i(\theta_p - \delta_{i1}) + \alpha_i(\theta_p - \delta_{i2}))/\kappa_{pi}, \end{aligned}$$

where $\kappa_{pi} = 1 + \exp(\alpha_i(\theta_p - \delta_{i1})) + \exp(\alpha_i(\theta_p - \delta_{i1}) + \alpha_i(\theta_p - \delta_{i2}))$ serves as a normalizing factor.

In the context of the generalized partial credit model (GPCM), as conceptualized by Muraki in 1992 [17] and further discussed by Muraki and Muraki in 2016 [18], the discrimination parameters are allowed to vary across items. The fundamental version of this model, wherein $\alpha_i = 1$ for all items, is recognized as the partial credit model (PCM).

$$P(Y_{pi} = r | \theta_p, \alpha_i, \delta_i) = \frac{exp(\sum_{l=1}^{r}(\theta_p - \delta_{il}))}{\sum_{s=0}^{k} exp(\sum_{l=1}^{s}(\theta_p - \delta_{il}))}, \quad r = 1, \ldots, k \tag{3}$$

The partial credit model is often referred to as a "divide-by-total" model, a characterization stemming from the denominator present in its probability expressions. This model is essentially equivalent to the polytomous Rasch model; the distinction lies mainly in the parameterization. For further insights into this equivalence, one can refer to works like Andrich (2010) [5] and Thissen and Steinberg (1986) [28].

A defining characteristic of the partial credit model becomes apparent when examining adjacent categories. By focusing on the comparison between two such categories, one can derive the following result:

$$\log\left(\frac{P(Y_{pi}=r|\theta_p,\alpha_i,\delta_i)}{P(Y_{pi}=r-1|\theta_p,\alpha_i,\delta_i)}\right)=\alpha_i(\theta_p-\delta_{ir}),\quad r=1,\ldots,k \tag{4}$$

This implies that the Partial Credit Model (PCM) directly contrasts two adjacent categories, where θ_p signifies the strength of preference for the higher category. In certain contexts, it might be more beneficial to consider the following representation:

$$\log\left(\frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)+P(Y_{pi}=r)}\right)=\alpha_i(\theta_p-\delta_{ir}),\quad r=1,\ldots,k \tag{5}$$

Which demonstrates that the model, locally (given the response categories $r-1,r$), functions as a binary Rasch model. In this context, θ_p acts as the person parameter and δ_{ir} represents the item difficulty.

2.1 Covariates in Partial Credit Model

The fundamental structure of the Partial Credit Model (PCM) can be presented as follows:

$$log\left(\frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)}\right)=\theta_p-\delta_{ir},\quad r=1,\ldots,k,$$

involving person parameters θ_p and item parameters δ_{ir}.
In the following let $x_p^T=(x_{p1},\ldots,x_{pm})$ denote a subject-specific vector of covariates that is assumed to potentially modify the response. In a variant of the Partial Credit Model that incorporates covariates, the person parameter θ_p is substituted by $\theta_p+x_p^T\gamma_i$, resulting in the following formulation:

$$log\left(\frac{P(Y_{pi}=r)}{P(Y_{pi}=r-1)}\right)=\theta_p+x_p^T\gamma_i-\delta_{ir},\quad r=1,\ldots,k \tag{6}$$

The item-specific parameter $\gamma^T=(\gamma_1,\ldots,\gamma_m)$ encapsulates the influence of covariates on response behavior.
A clearer understanding of the impact becomes evident when we take into account the odds at the local level:

$$\eta_{pir}=P(Y_{pi}=r|x_p)/P(Y_{pi}=r-1|x_p)$$

If we increase the j-th component of the vector x_p by one unit while keeping all other variables constant, the local odds change by a factor of $e^{\gamma j}$. This can be formally observed through:

$$\frac{\eta_{pir}(x_{p1},\ldots,x_{pj}+1,\ldots,x_{pm})}{\eta_{pir}(x_{p1},\ldots,x_{pm})} \tag{7}$$

For $\gamma_j > 0$, all local odds increase when the j-th variable increases, indicating that respondents with larger covariate values are inclined to choose higher categories. Conversely, for $\gamma_j < 0$, lower categories are expected.
This model encompasses a multitude of parameters; each item is associated with its unique vector γ_i. In simpler versions of the model, a common approach is to assume uniformity across these parameters, setting $\gamma_1 = \cdots = \gamma_i = \gamma$.

2.2 Inferential Issues

Marginal Maximum Likelihood (MML) estimation is a viable method for deriving estimates in the context of the Partial Credit Model. This model facilitates the estimation of parameters via Conditional Maximum Likelihood, owing to the existence of sufficient statistics for item parameters.
In MML estimation, the parameters of items are derived under the assumption of a fixed distribution. To effectively utilize this method, it is beneficial to represent the responses $Y_{pi} \in \{0,1,\ldots,k\}$ using indicator variables, which are defined as follows:

$$Y_{pir} = \begin{cases} 1 & \text{se } Y_{pi} = r \\ 0 & \text{se } Y_{pi} \neq r \end{cases}$$

The $0-1$ indicator variables effectively represent the responses in a vectorized format,

$$Y_{pi} = r \quad \Leftrightarrow \quad Y_{pi}^T = (Y_{pi0},\ldots,Y_{pik}) = (0,\ldots,0,1,0,\ldots,0).$$

In this representation, the vector Y_{pi} contains a single '1' entry. The indicator variables are closely associated with the concept of split variables, defined as $Y_{pi}^r = 1$ if $Y_{pi} \geq r$ and $Y_{pi}^r = 0$ if $Y_{pi} \leq r$. It is straightforward to see that $Y_{pi}^{(r)} = \sum_{j=r}^{k} Y_{pij}$ and $Y_{pir} = Y_{pi}^{(r)} - Y_{pi}^{(r+1)}$, where $Y_{pi}^{(k+1)} = 0$. Split variables and $0-1$ indicator variables are essentially two different methods of representing observations in a vector format.
Using these indicator variables, the marginal likelihood can be expressed as follows:

$$L_m = \prod_{p=1}^{P} \int \prod_{i=1}^{I} \prod_{r=1}^{k} P(Y_{pi} = r)^{y_{pir}} f(\theta_p) d\theta_p,$$

Where $f(\theta_p)$ represents the density of the assumed distribution for the person parameters. Commonly, a normal distribution with a mean of zero and a variance of σ^2, denoted by $f_{0,\sigma_\theta}(.)$, is presumed for these parameters. The corresponding log-likelihood can be articulated as follows:

$$\ell_m = \log(L_m) = \sum_{p=1}^{P} \log\left(\int \prod_{i=1}^{I}\prod_{r=1}^{k} P(Y_{pi} = r)^{y_{pir}} f(\theta_p) d\theta_p\right)$$

2.3 Penalty-Based Methods

Penalized maximum likelihood estimation (PMLE) is an approach used to estimate the parameters of a model in the presence of complexity or specific constraints. This technique modifies the traditional maximum likelihood estimation (MLE) by adding a penalty term to the likelihood function. The goal is to obtain parameter estimates that are not only plausible but also conform to certain desired properties, such as the parsimony of the model or the regularization of the parameters.

Given a set of data and a statistical model, MLE seeks the parameter values that maximize the likelihood function, that is, making the observed data as probable as possible.

However, in situations with high data dimensions or sparse data, MLE can lead to overfitting, where the model fits the sample data too well at the expense of its ability to generalize to new data. This is where penalized maximum likelihood estimation comes into play, introducing a penalty term into the likelihood function.

Maximize the penalized log-likelihood:

$$\ell_p(\{\gamma_i\},\{\delta_{ir}\}) = \ell(\{\gamma_i\},\{\delta_{ir}\}) - P_\lambda(\{\gamma_i\})$$

Where $P_\lambda(\{\gamma_i\})$ is a penalty term. This penalty term can take various forms.

Despite its advantages, PMLE presents some challenges. The choice of the penalty term and its weight (the regularization parameter) is crucial and often requires an empirical approach or one based on cross-validation criteria. Furthermore, the introduction of the penalty can lead to bias in the parameter estimates, although such bias is generally accepted as a compromise for greater robustness and generalization of the model.

2.3.1 Classical Approach

LASSO [29], an acronym for Least Absolute Shrinkage and Selection Operator, is a regularization technique used in statistical regression, particularly notable for its ability to produce parsimonious models through variable selection. Although commonly associated with minimizing the sum of squared residuals (RSS) in linear regression, its use in Maximum Likelihood Estimation (MLE) offers important prospects, especially in high-dimensional contexts. In an MLE context, the LASSO method aims to maximize the penalized log-likelihood function. The LASSO penalty term:

$$P_\lambda(\{\gamma_i\}) = \lambda \sum_{j=1}^{m} \sum_{i=1}^{I} |\gamma_{ij}| \tag{8}$$

Where λ represents the regularization parameter that controls the degree of penalty, imposes a penalty proportional to the absolute magnitude of the model coefficients, thus promoting a solution with many coefficients exactly equal to zero, resulting in natural variable selection. The primary advantage of LASSO in the MLE context lies in its ability to produce simplified models that maintain high predictive capacity. This is particularly valuable in situations where model parsimony is essential for interpretation or to avoid the curse of dimensionality. However, the choice of the regularization parameter λ is critical, and there is no standard method for its selection. Cross-validation methods are often used for this purpose. Additionally, when there are multiple highly correlated variables, LASSO tends to select one at the expense of others, which can be a limitation in terms of interpretation.

Ridge regression [11], also known as Tikhonov regularization, is a regularization method used to address multicollinearity issues in statistical models. While originally developed for linear regression with the goal of minimizing the sum of squared residuals (RSS) plus a penalty term proportional to the squares of the coefficients, the application of Ridge regression in the context of Maximum Likelihood Estimation (MLE) deserves careful consideration, especially for complex models and high-dimensional data.
In the MLE context, the objective of Ridge regression is to maximize a penalized log-likelihood function. The Ridge penalty term:

$$P_\lambda(\{\gamma_i\}) = \lambda \sum_{j=1}^{m} \sum_{i=1}^{I} \gamma_{ij}^2 \tag{9}$$

aims to reduce the magnitude of the model coefficients, particularly penalizing high values, which helps mitigate the multicollinearity problem.
Ridge regression is particularly useful in scenarios where the number of variables is high compared to the number of observations or when explanatory variables are strongly correlated. In such situations, standard MLE estimation can lead to unstable or non-unique solutions. By introducing the Ridge penalty term, a more stable and reliable solution can be obtained.
Unlike LASSO, Ridge regression does not lead to variable selection, as the coefficients are reduced but rarely exactly reach zero. This makes Ridge regression particularly suitable for situations where all variables are considered relevant, but it is necessary to control for the effect of multicollinearity. The main advantage of Ridge regression in the MLE context lies in its ability to produce stable and regularized parameter estimates, especially in the presence of multicollinearity. Furthermore, it can improve the predictive capacity of the model in high-dimensional settings.

2.3.2 New Approach

An alternative method (Fusion with Zero: Grouped LASSO on Differences and Selection (G. Tutz (2024)) [34]) is obtained by incorporating the following penalty term into the regularized maximum likelihood estimation:

$$P_{\lambda}(\{\gamma_i\}) = \lambda_f \sum_{j=1}^{m} \left(\sum_{r<s} |\gamma_{rj} - \gamma_{sj}| + \sum_{r} |\gamma_{rj}| \right) \tag{10}$$

If $\lambda_f = 0$, the estimate acquired is the conventional Maximum Likelihood (ML) estimate. However, when λ_f is finite and positive for each variable, it leads to the formation of clusters of effects, which could potentially be zero. This results in several items exhibiting the same effect.

Some alternative approaches (G. Tutz (2024)) [34] are the method that incorporates the only term: $\lambda_s \sum_{j=1}^{m} \sqrt{\sum_{i=1}^{I} \gamma_{ij}^2}$, namely 'Selection Only'.

An alternative is the 'Only Fusion' where the penalty term is: $\lambda_f \sum_{j=1}^{m} \sum_{r<s} |\gamma_{rj} - \gamma_{sj}|$.

The last approach combines the two previous alternatives by summing them.

3 Simulation

A simulation was conducted to assess the performance of the various proposed penalization methods, both classic and new. This simulation involves the generation of 1000 datasets and the application of a Partial Credit Model (PCM) with covariates to each of them, using the MultOrdRS package in R [25]. In particular, a total of 6000 models were estimated, as a PCM was applied to each of the 1000 datasets for each of the proposed types of penalization, resulting in 6 different penalizations in total. The MultOrdRS package was used to fit the models to the data and estimate the parameters. However, it is important to note that the package was modified to include five out of the six proposed penalizations. After estimating all the models on each dataset with the various penalizations, the Bayesian Information Criterion (BIC) was used to evaluate the model performance. This simulation was conducted to assess which of the proposed penalizations worked best for the specific data considered and to provide a basis for selecting the best model among the available options.
Before presenting the results, let's describe the general settings that were used in the simulation.

3.1 Simulation Settings

For the simulation, 1000 datasets were generated with the following characteristics: $P = 500$ observations, $I = 5$ items with $k = 5$ categories ranging from 1 to 5, 5 variables drawn from standard normal distributions, and 10 binary variables from a Bernoulli distribution with a probability parameters $p =$(0.42, 0.44, 0.58, 0.41, 0.6, 0.5, 0.57, 0.44, 0.5, 0.47).

3.2 Simulation Results

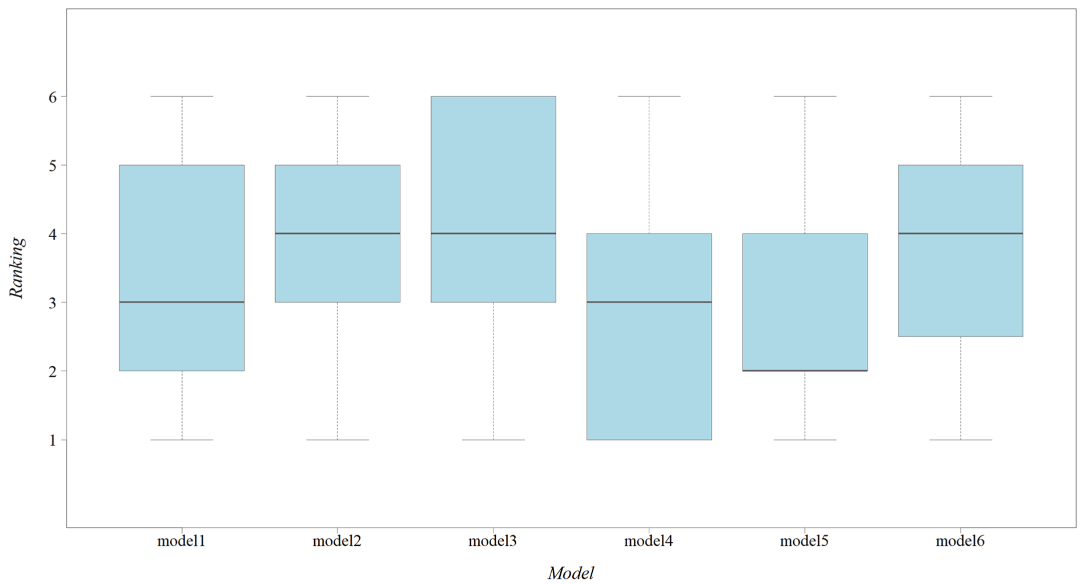

Figure 3: Ranking Model by BIC

The presented graph is a boxplot displaying the distribution of ranks assigned to six different models (model1 (LASSO), model2 (Ridge), model3 (Fusion With Zero), model4 (Selection Only), model5 (Fusion Only), model6 (Combined: Selection and Fusion) based on different regularization methods in a simulation. Each model was estimated on 1000 different datasets and compared to the other models using the Bayesian Information Criterion (BIC) to assess their performance. Models 2, 3, and 6 occupy the lower positions in the ranking and have very similar medians, indicating comparable average performance among them. Models 1 and 4 hold a middle-ranking position and show better performance than the previous three models. Model 5 has the highest median ranking, suggesting that it generally performed better than the other models.

It's important to note that the BIC is a criterion that penalizes model complexity, favoring simpler models if they perform comparably to more complex ones. Therefore, a lower rank here implies a better balance between the model's complexity and its ability to fit the data.

4 Application

The survey we conducted, myself, Antonio Cola, and my colleague Rosario Urso, at LASt[1], aimed at understanding individuals' perception in relation to seismic events. The research was carried out shortly after the onset of tremors in the Campi Flegrei area, from October 31st to November 16th, 2023, using a snowball sampling approach. The sample was selected in the Italian territory as reported in Figure 27 (Appendix).
The survey questions covered various socio-demographic aspects, in addition to specific inquiries about perception, mobility, information, trust in institutions, and economic-financial aspects related to seismic events. This multidimensional approach allowed for a complex and detailed analysis of the population's reactions and attitudes in response to seismic events.
It was observed that the results reflected a range of perceptions and reactions that varied significantly based on different factors such as age, gender, and others. Analyzing this data provides a clear picture of people's concerns, fears, and expectations regarding seismic events, emphasizing the importance of a personalized approach in information and emergency management. This case study could provide valuable insights for authorities and organizations responsible for seismic emergency management, suggesting the importance of considering the diversity of individual perceptions in communication and intervention strategies.

4.1 Exploratory Analysis

It is crucial to conduct a preliminary exploratory analysis to understand the data distribution and identify trends, patterns, and potential associations among the variables of interest. In the first part of the analysis, we focused on the distribution of 7 items related to earthquake perception. It emerged that the perception of seismic events varies considerably among the study participants. Subsequently, covariates of different nature were examined. Socio-demographic variables revealed differences among age groups, genders, and educational levels of the participating individuals. Mobility was examined in terms of residence, distance from seismic areas, and travel habits. The information available to the participants and their level of trust in information sources were found to be influential variables. Additionally, the economic and financial conditions of individuals, including income and employment, were considered.
The exploratory analysis revealed a wide range of responses and behaviors among individuals regarding earthquake perception and the analyzed covariates. This heterogeneity underscores the importance of subsequently applying Item Response Theory (IRT) models to gain a deeper understanding of how these variables influence earthquake perception and how they can be used to improve the accuracy of individual ability measurements in relation to seismic events. The exploratory analysis provided a solid foundation for the subsequent analysis based on IRT models, thus contributing to a more comprehensive understanding of the phenomenon under

[1]Laboratory for Statistical Data Analysis (LASt) – University of Naples Federico II
Scientific Director: Maria Iannario – Domenico Vistocco
Main activities: Individual research consultancy, comprehensive training, e-learning tools, and a learning management system. Managing aspects related to research and publishing the results of the research conducted through the Centre; research and consultancy activities (including training) in data analysis and visualization.

study. The questionnaire underlies a single latent trait supported by preliminary statistical analyses.

4.1.1 Description of the Dataset

The original dataset comprises 47 variables. Subsequently, to make the dataset more manageable and meaningful, several modifications were made, including the removal of variables deemed not useful. The result is a dataset that comprises 226 observations and 44 variables. The econding of variable is avaible in the Appendix.

4.1.2 Item Distribution

Let's proceed with the analysis of the distribution of 7 items (shocks, fear, anxiety, physiological_symptoms, decision_timeliness, insomnia, seismic_concern) belonging to the same set of questions related to earthquake perception. These items were selected to assess and contribute to the determination of a latent construct ('Earthquake Stress') related to individuals' perception and reaction to earthquakes. This latent construct represents the underlying variable that we will seek to measure with subsequent Item Response Theory models, in order to gain a deeper and more precise understanding of individuals' earthquake perception. This latent construct would reflect the psychological and emotional tension that people experience in response to seismic events, including concerns, anxieties, physical symptoms, decision-making difficulties, and sleep disturbances caused by earthquakes.

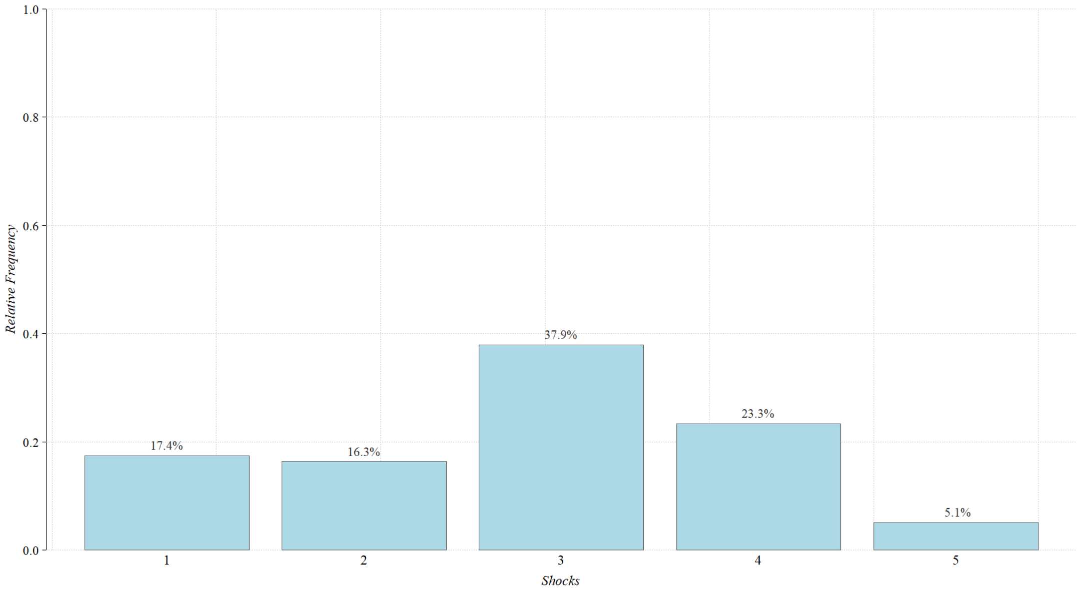

Figure 4: Shocks Distribution

The graph shows that the majority of people (approximately 37.9%) rated the intensity of seismic shocks with a score of 3 out of 5. Ratings 1 and 2, representing a relatively low perception of intensity, were chosen by approximately 17.4% and 16.3% of the respondents, respectively. However, it is interesting to note that a significant 28.4% rated the intensity with a score of 4 or 5, indicating a perception of moderate to high seismic intensity. These results suggest that a significant portion of the population may have experienced or perceived seismic events of higher intensity, which could influence their level of stress and concern. Based on this distribution, it can be stated that perceptions of earthquake intensity vary significantly among individuals, and this could have significant impacts on their emotional reactivity to seismic events.

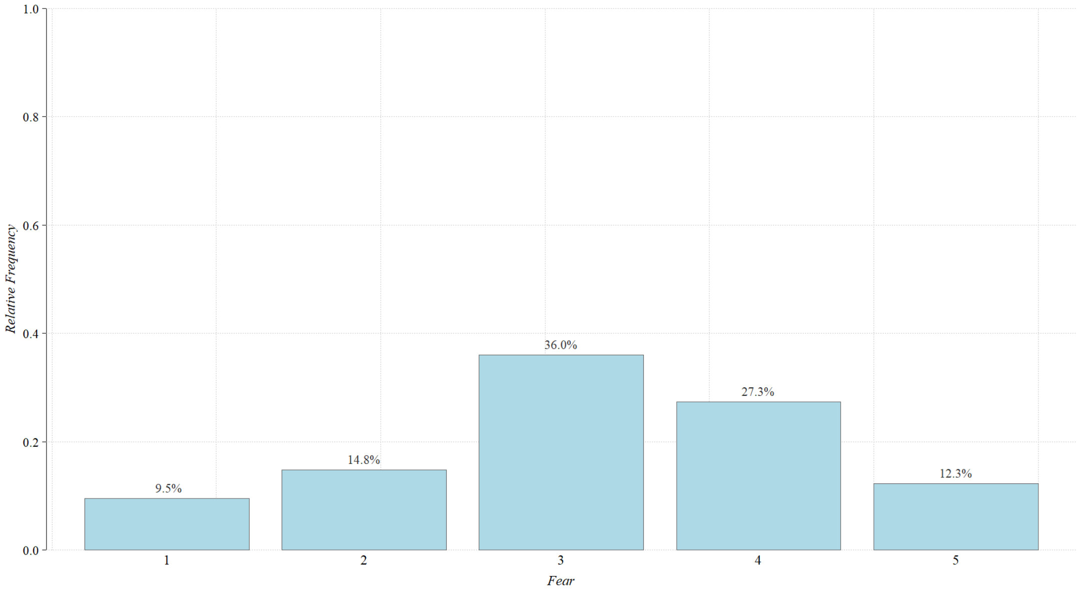

Figure 5: Fear Distribution

In this distribution, the most frequent rating was 3, with 36.0% of respondents indicating experiencing moderate fear during seismic events. However, it is important to note that a significant percentage of individuals reported higher levels of fear, with 27.3% giving a score of 4 and 12.3% giving a score of 5, indicating considerable or very high fear. On the other hand, 9.5% reported a score of 1, indicating a very low perception of fear during seismic events, while 14.8% assigned a score of 2.

Considering these results, it is evident that a significant portion of the population experiences some degree of fear during seismic events, with a non-negligible percentage experiencing higher levels of fear. This suggests that earthquake perception is not uniform, and the level of fear can vary considerably among individuals.

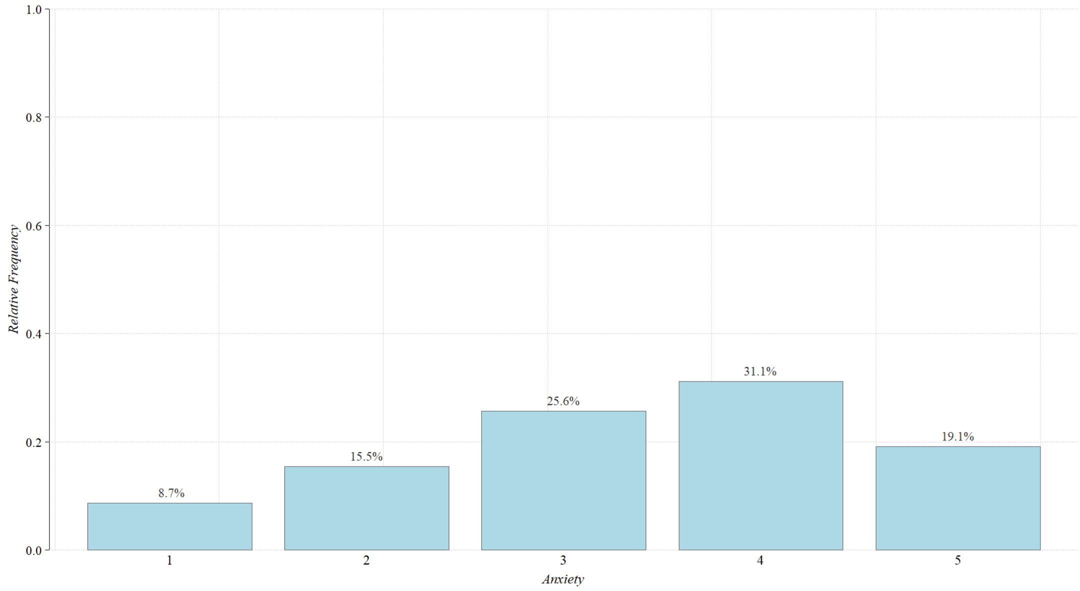

Figure 6: Anxiety Distribution

This distribution shows that a significant portion of the population has experienced various levels of anxiety during seismic events. The most common rating was 4, with 31.1% of respondents indicating experiencing moderate anxiety. However, 25.6% also reported an anxiety level of 3, suggesting that a considerable portion of the population experiences some degree of anxiety during seismic events. Additionally, 19.1% reported a score of 5, indicating a high level of anxiety, while 15.5% assigned a score of 2, and 8.7% a score of 1, indicating minimal anxiety or the complete absence of anxiety during seismic events, respectively.

Considering this data, it becomes evident that anxiety is a common emotion during seismic events, with significant variation in the levels of anxiety perceived by individuals.

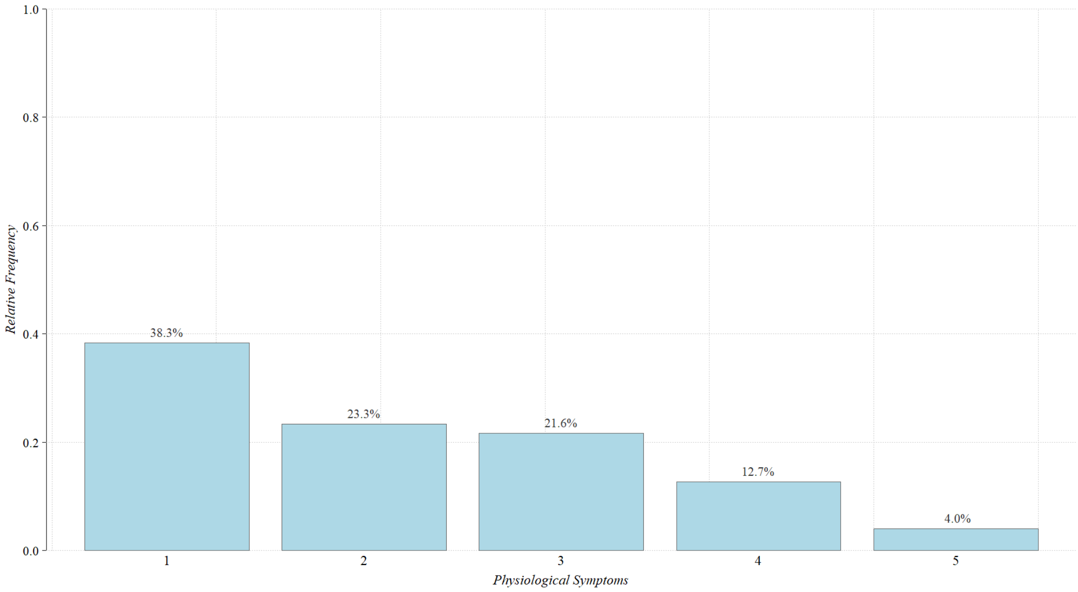

Figure 7: Physiological Symptoms Distribution

This distribution highlights that a large percentage of the population (38.3%) reported a low level of physiological symptoms during seismic events, assigning a score of 1 out of 5. However, a significant portion of the population still reported higher physiological symptoms during seismic events, with 23.3% giving a score of 2 and 21.6% giving a score of 3. On the other hand, 12.7% reported physiological symptoms at level 4, while only 4.0% reported the highest level of physiological symptoms (score 5).

These data suggest that a significant portion of the population experiences physiological symptoms during seismic events, but the majority appears to either not experience such symptoms or experiences them at a low intensity.

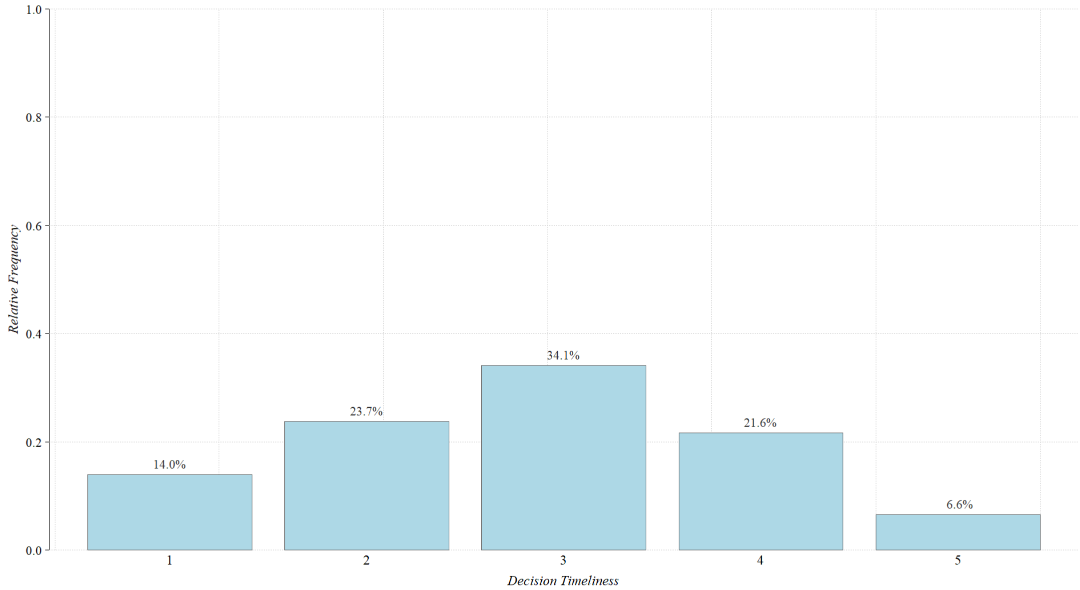

Figure 8: Decision Timeliness Distribution

This distribution highlights that the majority of study participants reported ratings indicating good timeliness in decision-making during seismic events. 34.1% of respondents assigned a score of 3, indicating appropriate timeliness in decisions. However, a significant portion reported varying levels of decision-making timeliness, with 23.7% giving a score of 2 and 21.6% giving a score of 4. Only 6.6% rated decision timeliness with the highest score of 5, indicating excellent timeliness. It's interesting to note that 14.0% of respondents gave a score of 1, indicating very poor decision-making timeliness during seismic events.

Timeliness in decision-making during seismic events varies significantly among individuals. The majority appears to have good timeliness, but a significant portion may face difficulties in making timely decisions. This aspect is important in emergency situations like earthquakes because quick decisions can impact personal safety.

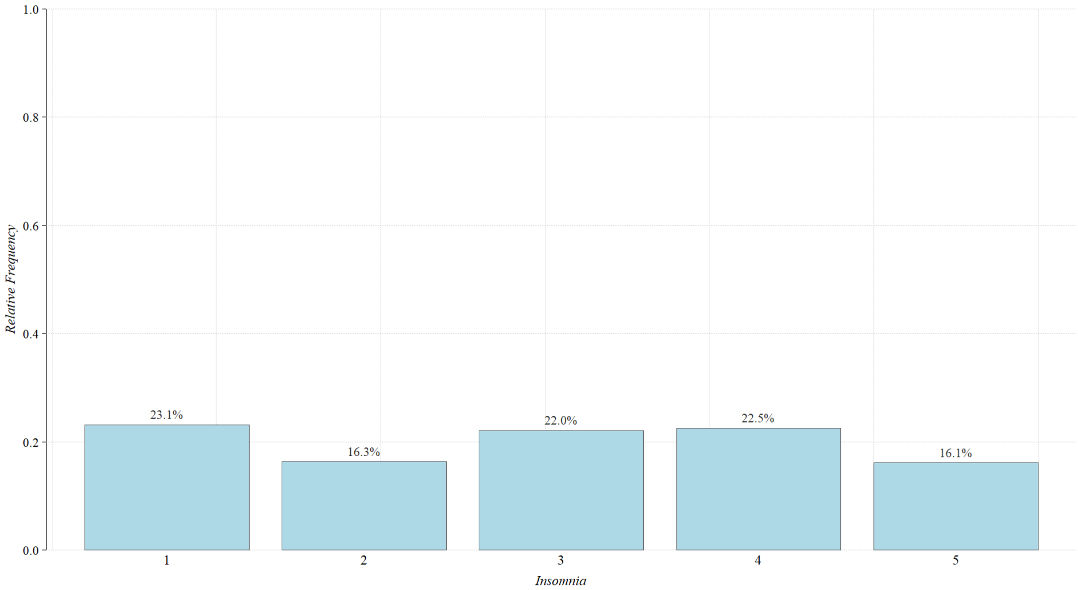

Figure 9: Insomnia Distribution

This distribution highlights that a significant portion of the population has reported levels of insomnia during seismic events. 23.1% of respondents assigned a score of 1, indicating very low insomnia or the absence of insomnia. However, a significant portion reported higher levels of insomnia, with 22.5% giving a score of 4 and 16.1% giving a score of 5, indicating moderate or severe insomnia. Additionally, 16.3% gave a score of 2, while 22.0% gave a score of 3, indicating moderate-level insomnia.

These data suggest that insomnia is a significant issue during seismic events, with a significant portion of the population experiencing difficulty sleeping. This can have consequences for individuals' health and well-being.

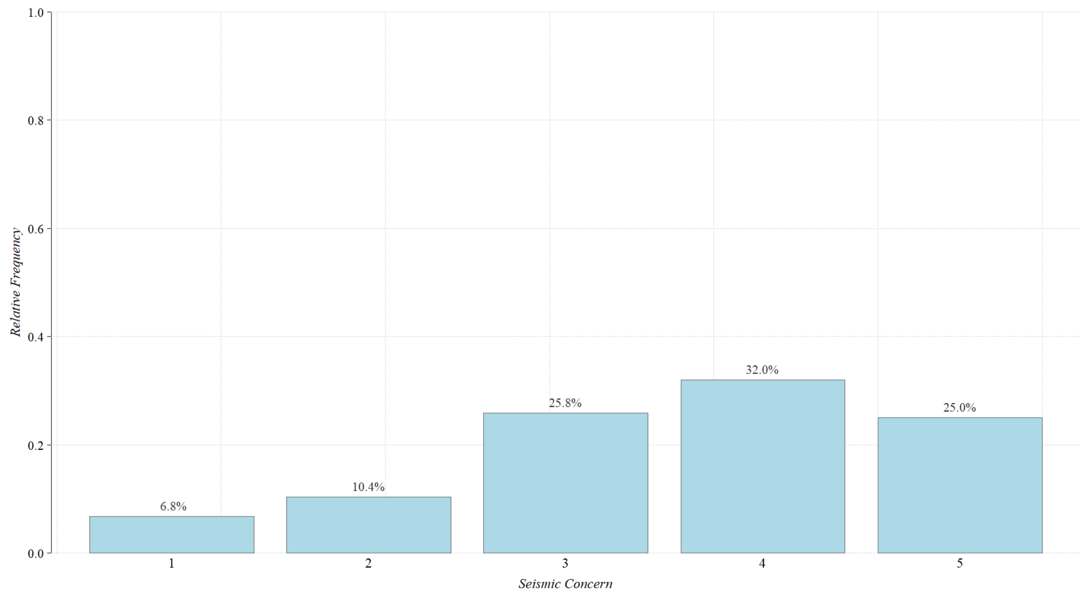

Figure 10: Seismic Concern Distribution

This distribution reflects people's perceptions of their concern related to seismic events. It appears that a significant portion of the population (32.0%) reported high concern (4 out of 5) regarding seismic events. Additionally, 25.8% of respondents assigned a score of 3, indicating moderate concern. However, it's interesting to note that 25.0% assigned a score of 5, indicating very high concern, while 10.4% gave a score of 2, and 6.8% a score of 1, indicating low or very low concern regarding seismic events, respectively.
These data suggest that seismic concern is a significant aspect of people's perception of seismic events. A considerable portion of the population experiences significant concern, which can influence their emotional and behavioral response in seismic situations.

The analysis of the distributions of the 7 items reveals a general trend of variability in individuals' responses regarding their perception of seismic events. Overall, it becomes evident that there is a considerable diversity in individuals' perceptions and reactions to such events. In summary, the data confirm that individuals' perception and response to seismic events vary significantly, highlighting the complexity of emotional and behavioral responses related to earthquakes. These results will serve as an important foundation for further analysis using Item Response Theory (IRT) models to accurately measure the latent construct of 'Earthquake Stress' and to explore the relationships between these 7 items and the construct itself.

4.1.3 Variable Distribution

Before proceeding with the application of Item Response Theory (IRT) models to measure the latent construct of 'Earthquake Stress', it is essential to conduct an analysis of the distribution of the explanatory variables.

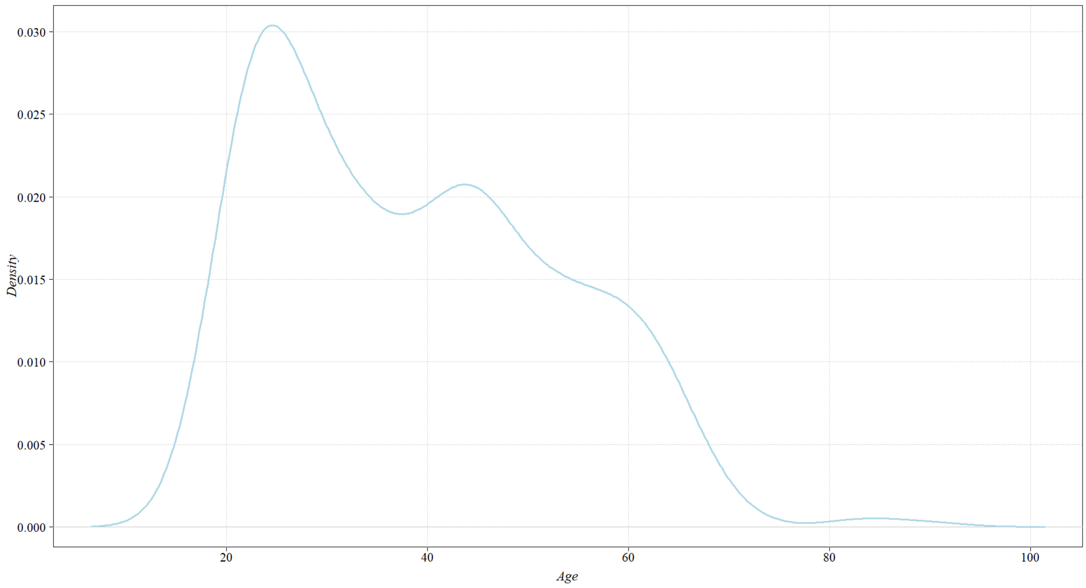

Figure 11: Age Distribution

The probability density function shown in the figure illustrates the distribution of probabilities across different ages. Analyzing the shape of the graph, we observe a pronounced peak around an age value of approximately 25-30 years. This peak indicates that there is a higher concentration of individuals in this age range within the dataset. After the peak, the density gradually decreases as age increases, suggesting that there are fewer individuals in the older age groups.

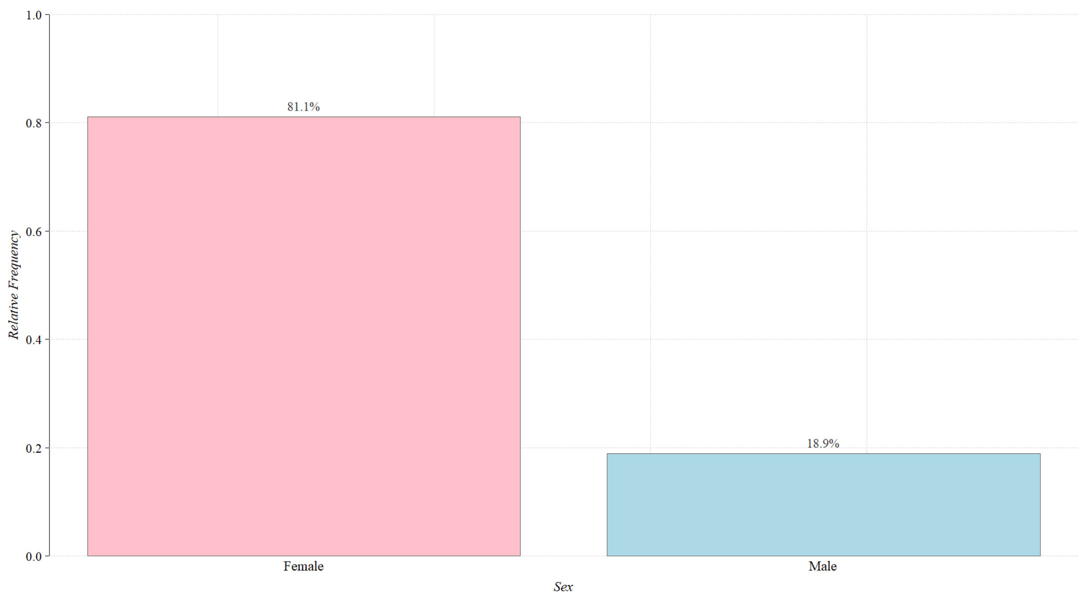

Figure 12: Sex Distribution

The presented graph is a bar chart that visualizes the relative frequency distribution of two gender categories, female and male. The pink bar, representing females, occupies the vast majority of the chart, with a indicated percentage of 81.1%. In contrast, the blue bar representing males is significantly shorter, with a percentage of 18.9%. This highlights that the analyzed population is heavily skewed towards the female gender compared to the male gender.

It has been observed that surveys tend to receive more responses from women compared to men. This phenomenon was highlighted by William G. Smith in 2008 [27], who noted a marked gender difference in participation in online surveys, with women exhibiting significantly higher response rates than men.

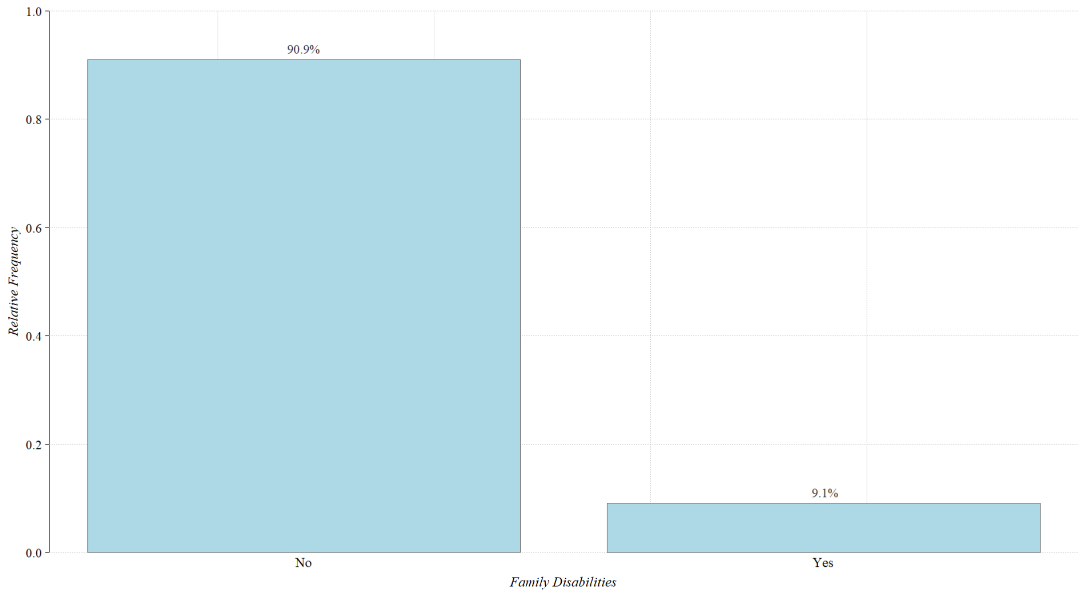

Figure 13: Family Disabilities Distribution

The analysis of the relative frequency graph regarding the presence of people with disabilities in families clearly reveals a predominant pattern. The 'No' category is predominant, holding a position of great significance with a substantial percentage of 90.9%. This data unequivocally indicates that the vast majority of families represented in the data do not include members with disabilities. On the other hand, the 'Yes' category shows a relatively modest presence, with a percentage of 9.1%. This suggests that only a small portion of the involved families reports the presence of people with disabilities.

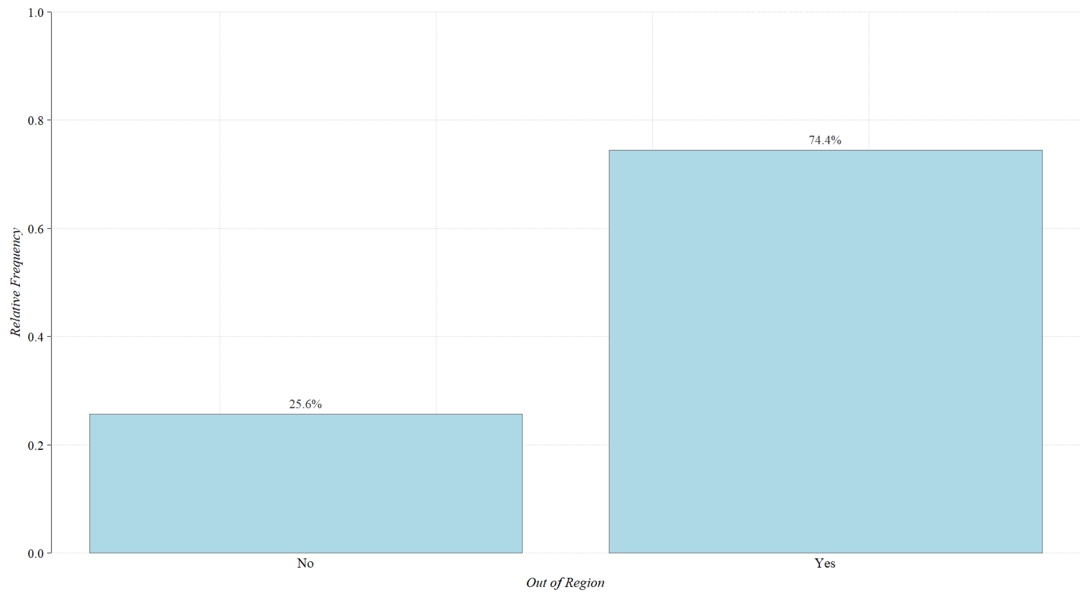

Figure 14: Out of Region Distribution

The bar chart displays the relative frequency distribution for a variable measuring people's consideration of working or studying outside their region of residence. The bar associated with 'No' shows that 25.6% of the individuals under consideration are not willing or have not considered the opportunity to relocate for work or study. In contrast, the bar corresponding to 'Yes' is significantly higher, indicating that a significant majority, 74.4%, has considered or is willing to work or study outside their region. This imbalance could reflect various socio-economic and cultural factors. It might indicate a high level of mobility in the population or the perception that better opportunities for work and study are located elsewhere. It could also suggest a level of dissatisfaction with local opportunities or a greater openness to change and exploration.

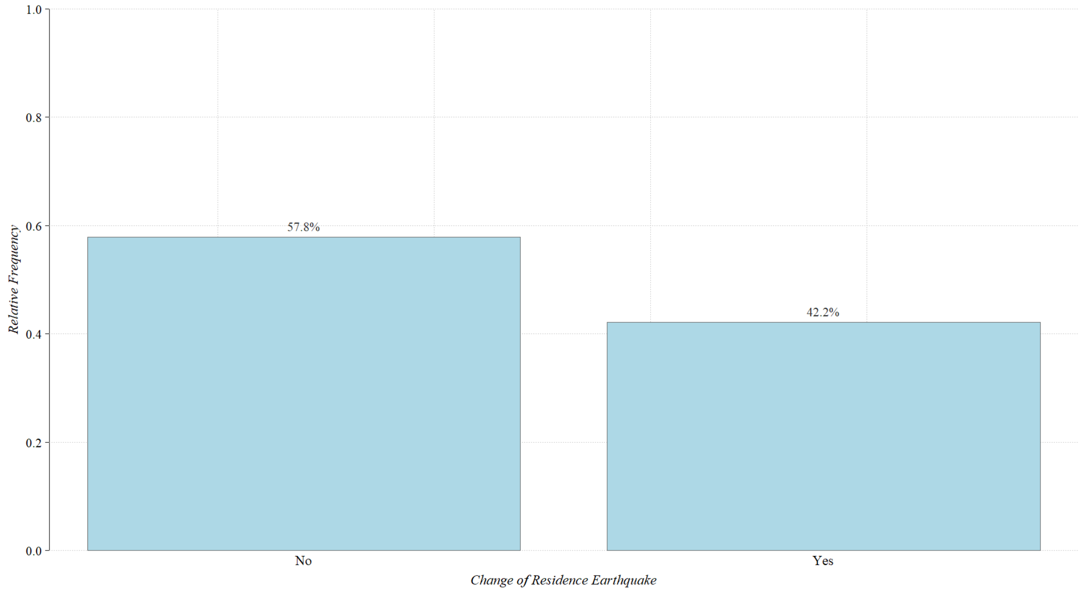

Figure 15: Change of Residence Earthquake Distribution

The chart provides an analysis of people's inclination to change their residence due to seismic events. It can be observed that the majority, 57.8%, has not considered relocating due to earthquakes or other seismic events, as indicated by the 'No' bar. However, a significant portion, 42.2%, has contemplated the possibility of changing their residence, as highlighted by the 'Yes' bar. These data suggest that while it is not the majority, there is still a substantial portion of the population concerned about seismic risks to the extent of considering a move. This may reflect the lived experience of recent earthquakes or an increasing awareness of seismic risks. It could also indicate the perception of infrastructure quality or confidence in local safety measures.

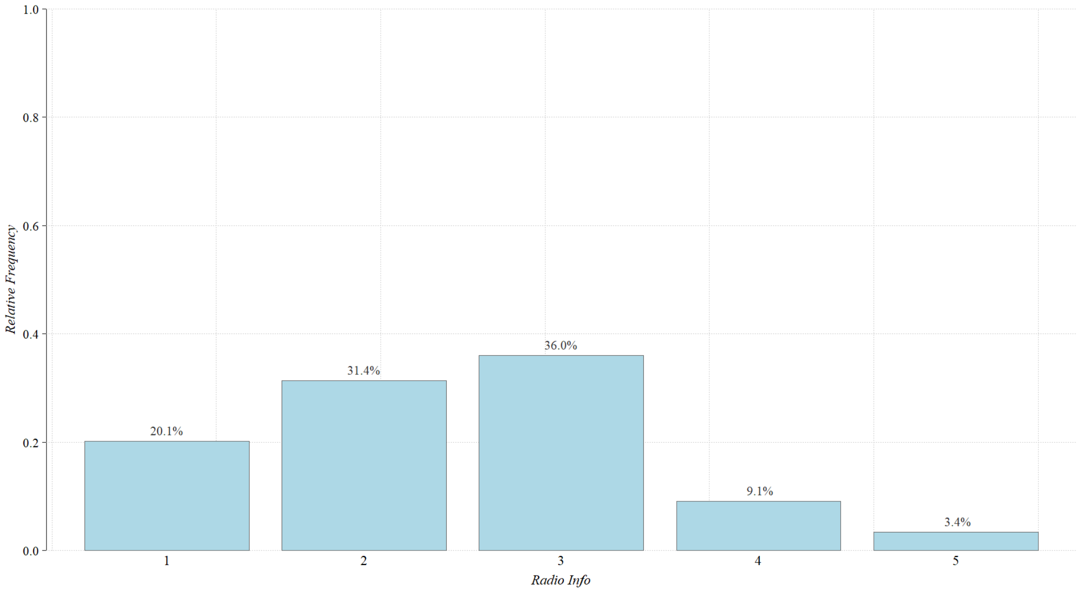

Figure 16: Radio Info Distribution

The chart represents the perception of the usefulness of information received from the radio regarding seismic activities, measured on a scale from 1 to 5, where 1 indicates ’not useful’ and 5 indicates ’very useful’. Most respondents appear to consider the received information to be of moderate usefulness, with 36.0% assigning a score of 3. Percentages decrease as one moves towards higher and lower scores, with 20.1% rating the information as less useful (score 1) and only 3.4% finding it very useful (score 5). It is interesting to note that a significant number, 31.4%, gave a score of 2, suggesting that while they don’t consider the information completely useless, they don’t find it particularly useful. These results may indicate that while the radio is a source of information on seismic activities for some people, there may be room for improvement in the quality and relevance of the information conveyed. It could also reflect the evolution of media, with a potential preference for other more immediate or detailed sources of information like the internet or news apps.

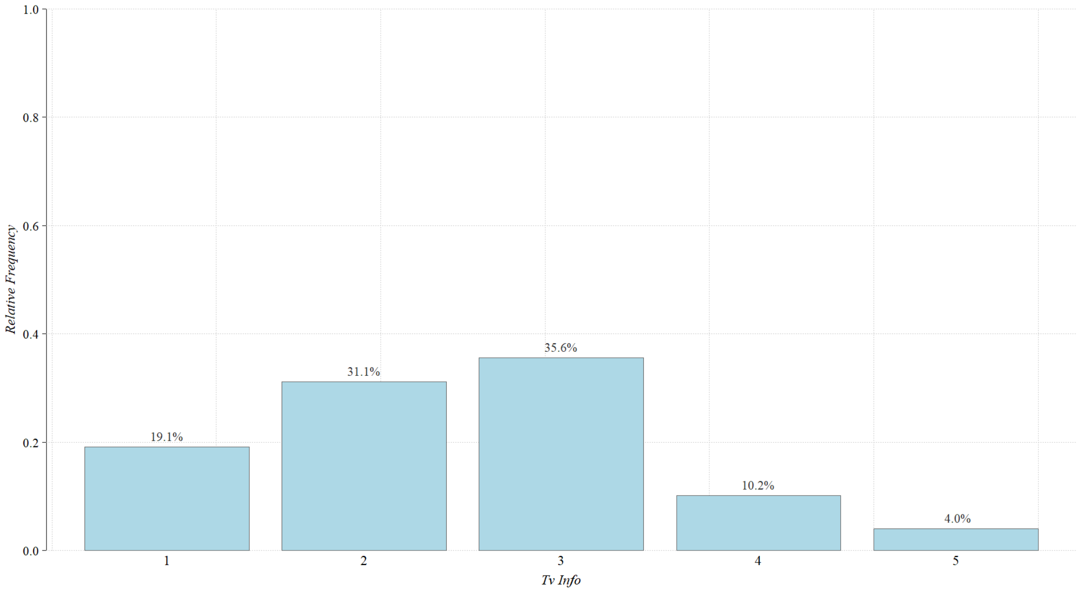

Figure 17: TV Info Distribution

The previous chart provides an analysis of the perception of the usefulness of information received from television regarding seismic activities, evaluated on a scale from 1 to 5. It is noticeable that the majority of respondents assigned a moderate score of 3, with a percentage of 35.6%, indicating that the information is generally perceived as moderately useful. A significant 31.1% gave a score of 2, suggesting that the information provided by TV is not particularly useful but not entirely worthless. A smaller number of people rated the usefulness of the information as high (10.2% for score 4 and 4.0% for score 5), while 19.1% gave the lowest score of 1, indicating a perception of the TV information on seismic activities as unhelpful. These responses may reflect various factors such as the quality and timeliness of the information provided, the relevance of the content to the audience, or trust in television as a reliable source of information in emergency situations.

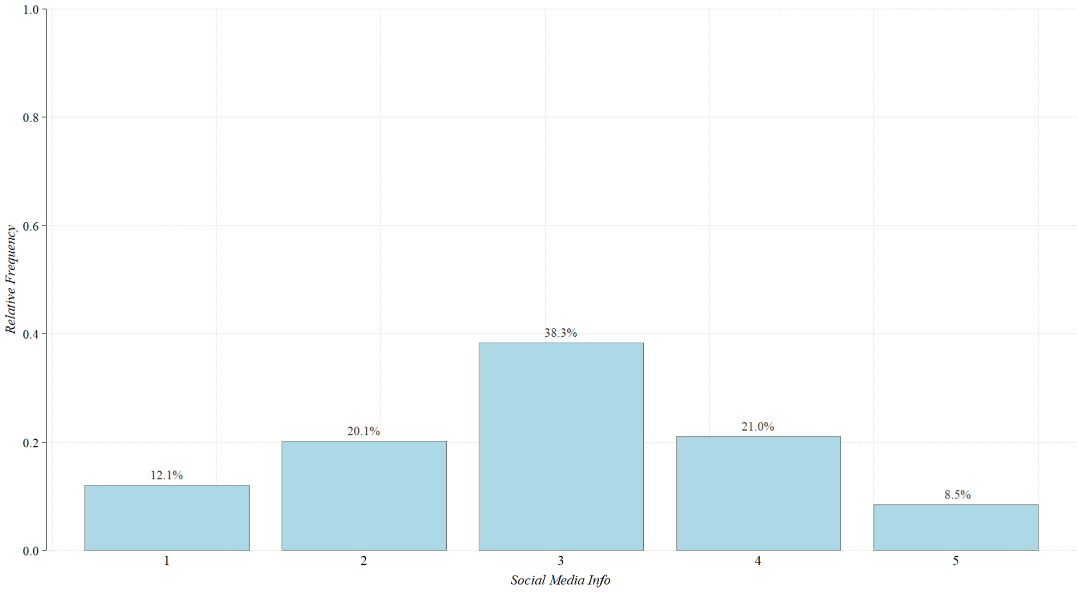

Figure 18: Social Media Info Distribution

The chart illustrates the evaluation of the usefulness of information received from social media regarding seismic activities on a scale from 1 to 5. Most people rated it as 3, with a percentage of 38.3%, indicating that, in general, information on social media is perceived as moderately useful. A significant fraction, 20.1%, found it less useful, giving a score of 2, while a smaller percentage, 12.1%, considered it entirely unhelpful, giving the lowest score of 1. There is then a 21.0% that rated it as 4, suggesting they find the information quite useful, and only 8.5% rated it as very useful with a score of 5. This could reflect the nature of social media as sources of quick and accessible information, but which can vary significantly in terms of reliability and accuracy. The wide spread of scores from the middle towards the two extremes suggests mixed opinions on the credibility and usefulness of information disseminated through these channels. This highlights the need for authorities to provide verified and reliable information via social media during seismic events to ensure the public receives helpful and timely advice.

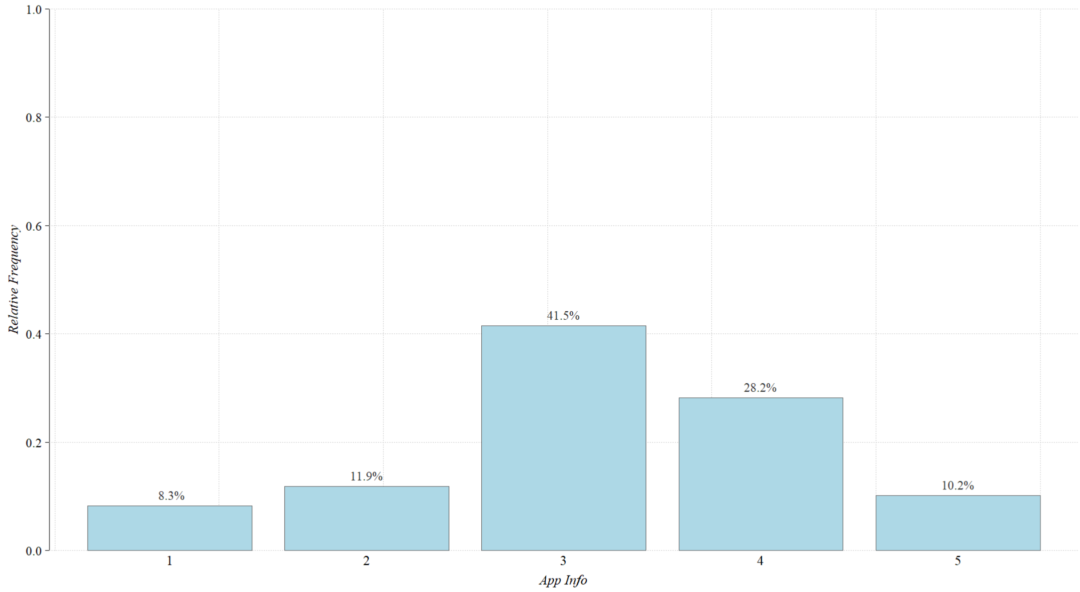

Figure 19: App Info Distribution

The chart displays perceptions of the usefulness of information received through apps related to seismic activities. The general trend shows that most respondents found the information provided by the apps moderately useful, with 41.5% giving it a score of 3. A substantial 28.2% rated the information as quite useful with a score of 4, and a minority but still significant 10.2% considered it very useful with a score of 5. This indicates a relatively positive view of the apps' usefulness. On the other hand, 11.9% gave a rating of 2, and only 8.3% assigned the lowest score of 1, suggesting that a smaller number of people found the information less useful or useless. These ratings may reflect the effectiveness of the apps in providing timely and reliable updates during seismic events. The availability of real-time information and the user-friendliness of the apps may contribute to a more positive evaluation. However, the presence of scores below 3 could indicate the need for improvements in terms of information accuracy, content customization, or app usability.

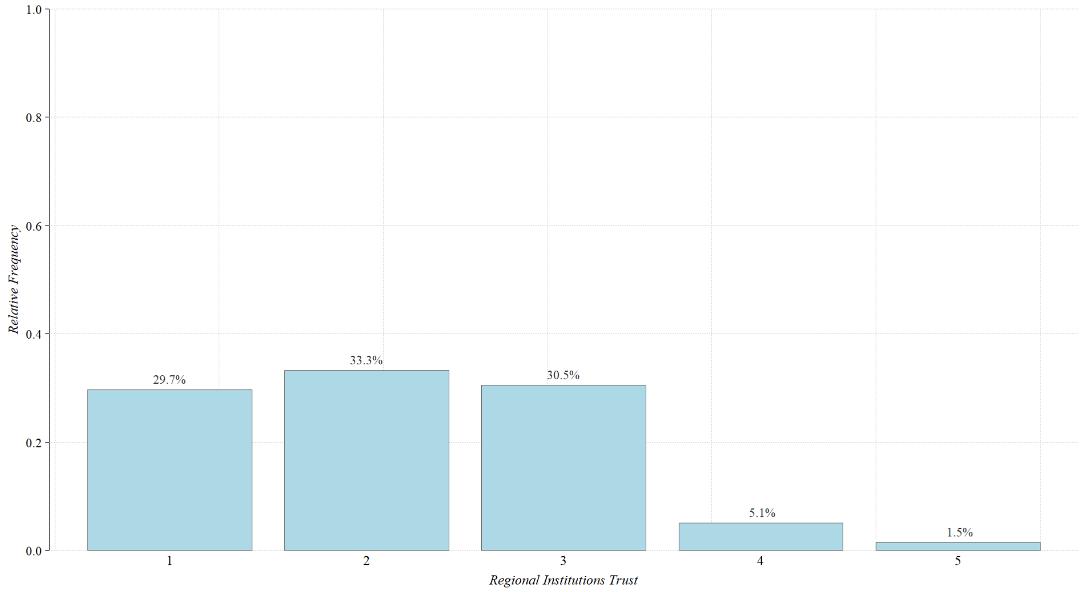

Figure 20: Regional Institutions Trust Distribution

The chart displays the distribution of perceptions of trust in regional institutions, rated on a scale from 1 (minimal trust) to 5 (maximum trust). From the data, it can be observed that there is not a strong inclination towards high levels of trust. Most respondents gave a score of 2 or 3, with 33.3% expressing low trust and 30.5% indicating moderate trust. These scores suggest a general perception of uncertainty or moderate trust in regional institutions. Only a small percentage expressed high levels of trust, with 5.1% choosing 4, and an even smaller 1.5% assigning the maximum score of 5. In contrast, 29.7% expressed the lowest level of trust with a score of 1. This pattern suggests that there may be a sense of distrust or a lack of a strong connection with regional institutions among the represented population.

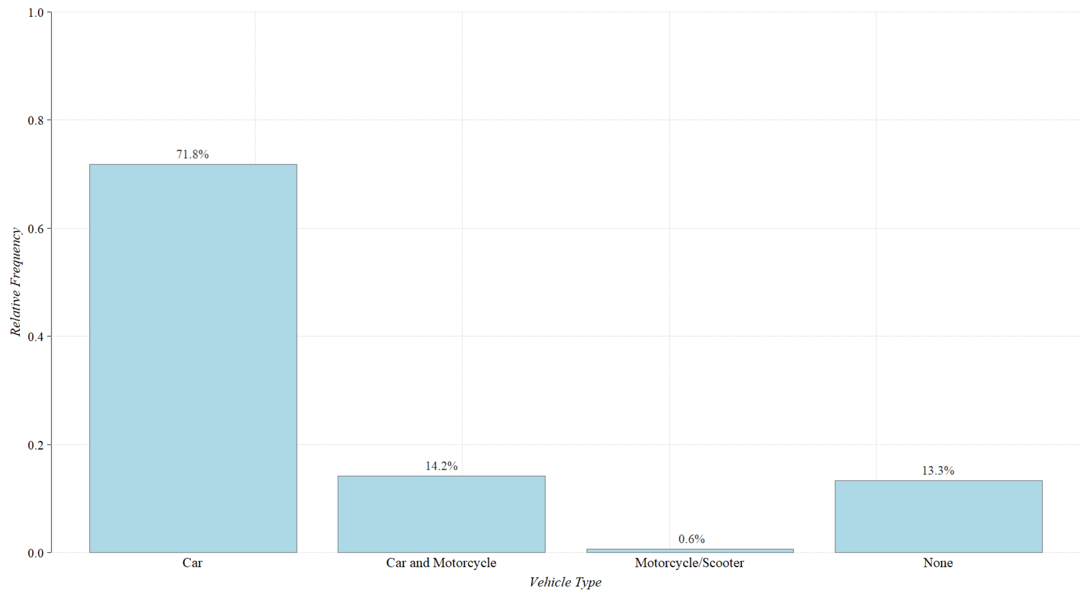

Figure 21: Vehicle Type Distribution

The chart shows the distribution of the type of vehicles owned by the selected sample. The vast majority, 71.8%, own a car, indicating that this mode of transportation is the most common among respondents. A small percentage, 14.2%, own both a car and a motorcycle/scooter, suggesting that a segment of the population prefers to have multiple transportation options. It's interesting to note that only 0.6% of respondents own a motorcycle or scooter, indicating that these vehicles are less popular or perhaps less practical for the daily needs of the survey participants. Finally, 13.3% do not own any type of vehicle, which could reflect a variety of factors such as choosing a more sustainable lifestyle or economic limitations.

To enhance readability, the analysis of the distribution of variables has been focused exclusively on the variables considered significant in the subsequent IRT models. The remaining variables have been presented through graphs in the appendix, section B, and their respective codings have been detailed in section A.

4.1.4 Correlation

The knowledge of the correlation between variables is important because it allows to identify the relationship between two or more quantitative variables. This information is useful for understanding data behavior and can be used to predict the behavior of one variable based on the knowledge of another. In this case, correlations between only the items were analyzed. The results are reported in the following correlation matrix:

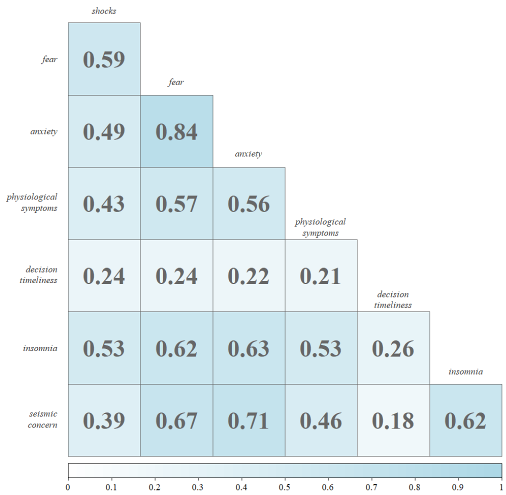

Figure 22: Correlation Matrix

Typically, the correlation matrix is a symmetric matrix, meaning it is a square matrix that is equal to its transpose. This implies that element (i, j) in the matrix is equal to element (j, i) in the same matrix. In this case, to avoid redundant data, the correlation matrix has been reduced to a lower triangular form. Subsequently, data with a unitary value on the main diagonal have been removed. The correlation values of the items are all positive, ranging from a minimum of 0.18 to a maximum of 0.84. It can be stated that there is a positive linear dependence between all pairs of items. In other words, one item increases with the increase of another, to the extent of the correlation coefficient, keeping in mind that the latter is: $-1 < \rho < 1$.
Subsequently, Cronbach's alpha was calculated, resulting $\alpha = 0.86$. A Cronbach's alpha value of 0.86 is generally considered quite good, keeping in mind that the latter is: $0 < \alpha < 1$ and indicates a good internal reliability of the measurement instrument used in the research or evaluation. This value suggests that the questions or items included in the instrument are consistent with each other and reliably measure the same characteristic or construct.
However, a higher Cronbach's alpha value can be obtained by excluding certain items, as we will see later (4.2.1).

4.2 Models

In the continuation of the research, two fundamental steps will be adopted. Initially, a careful item selection will be conducted based on the methodologies previously examined and detailed in the dedicated chapter. This phase will help identify the most informative and relevant items for the latent construct under study, ensuring a more precise measurement of 'Earthquake Stress'.
Subsequently, the estimation of partial credit models will be carried out for each of the previously introduced penalties, as elaborated in the corresponding chapter. The objective will be to assess how different penalties affect the model structure and the measurement of the la-

tent construct. It will be crucial to compare these models using the BIC to determine which model provides the best fit to the data and, consequently, a more accurate representation of 'Earthquake Stress'. The results of this phase will be carefully interpreted to draw meaningful conclusions about the measurement of the construct and its relationship with the explanatory variables.

4.2.1 Item Selection

In the item selection process, two distinct approaches were employed: the 'bruteforce' method and the 'backward' method (1.1.1 and 1.1.2). It is worth noting that both approaches led to the identification of the same combination of items as the best for measuring the 'Earthquake Stress' construct. This combination includes the following items: fear, anxiety, insomnia, and seismic_concern. Selecting a total of 4 items out of the initial 7 considered.
An important aspect of this selection is the observation that Cronbach's alpha, a measure of internal reliability, increased from 0.86 to 0.89 after the optimized item selection. This result suggests that the identified combination of items offers greater internal consistency in individuals' responses, increasing the measurement's internal consistency. Furthermore, the InfoRatio, reflecting the ratio between the number of informative items and the total number of items considered, is equal to 1, confirming that the selected combination of items has maximum informativeness regarding the 'Earthquake Stress' construct.
These results indicate that the optimized item selection has resulted in a more accurate and reliable measurement of the construct, highlighting the importance of careful item selection in the context of analysis using Item Response Theory (IRT) models.

4.2.2 Partial Credit Model

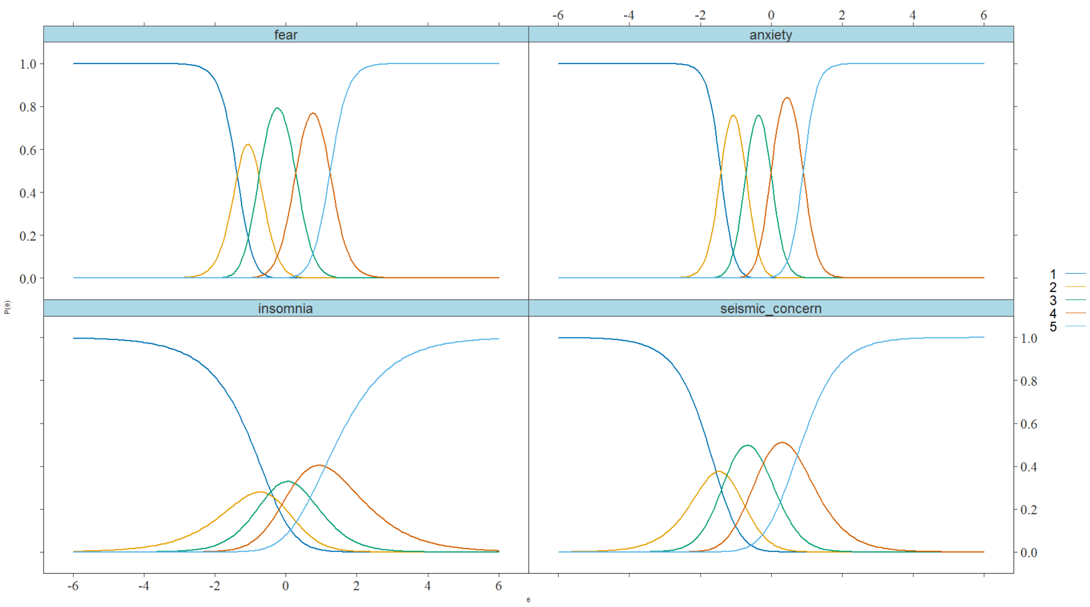

Figure 23: Category Response Curves

It is interesting to examine the probabilities of responding to specific categories in an item's response scale. These probabilities are graphically displayed in the category response curves (CRCs). Each symmetrical curve represents the probability of endorsing a response category. These curves have a functional relationship with θ, as θ increases, the probability of endorsing a category increases and then decreases as responses transition to the next higher category. The CRCs indicate that the response categories cover a wide range of θ.
To assess the model fit, the M2 index, designed specifically for evaluating the fit of item response models for ordinal data, was used (C. Li and M. Hansen 2013)[14]. Additionally, the M2-based root mean square error of approximation served as the primary fit index, along with the standardized root mean square residual (SRMSR) and the comparative fit index (CFI):

M2	df	p	RMSEA	RMSEA_5	RMSEA_95	SRMSR	TLI	CFI
22.05	2	0	0.15	0.09	0.2	0.03	0.95	0.98

The obtained $SRMSR = 0.03$ suggests that data fit the model reasonably well using suggested cutoff values of $SRMSR \leq 0.08$ as suggested guidelines for assessing fit. The $CFI = 0.98$ was above a recommended 0.95 threshold. The $RMSEA = 0.15$ (95% CI[0.09, 0.2]) was above a recommended 0.05 threshold, suggesting a modest fit of the model to the data.

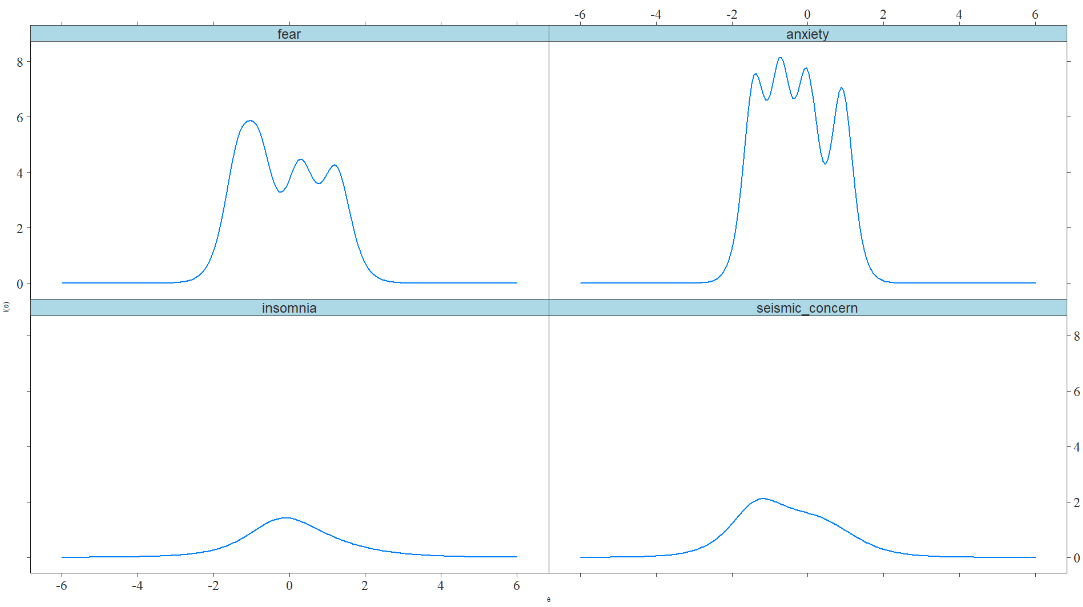

Figure 24: Item Information Curve

In polytomous models, the amount of information an item contributes depends on its slope parameter: the larger the parameter, the more information the item provides. Further, the farther apart the location parameters ($\delta_{i1}, \delta_{i2}, \delta_{i3}, \delta_{i4}$), the more information the item provides. Typically, an optimally informative polytomous item will have a large location and broad category coverage (as indicated by location parameters) over θ.
Information functions are best illustrated by the item information curves for each item as displayed above. These curves show that item information is not a static quantity, rather, it is

conditional on levels of θ. The relationship between slopes and information is illustrated here. Insomnia had the lowest slope and is, therefore, the least informative item. On the other hand, anxiety had the highest slope and provides the highest amount of statistical information. Items tended to provide the most information between [-2, 2] θ range . The wavy form of the curves reflects the fact that item information is a composite of category information, that is, each category has an information function which is then combined to form the item information function.

Item	α_i	δ_{i1}	δ_{i2}	δ_{i3}	δ_{i4}
fear	3,9405	-1,3725	-0,7589	0,2766	1,2408
anxiety	5,2280	-1,4235	-0,7161	-0,0114	0,8976
insomnia	1,0355	-0,3590	-0,4883	0,2509	1,1052
seismic_concern	1,6409	-1,5560	-1,2526	-0,2241	0,7295

4.2.3 Partial Credit Model with Covariates and Penalties

Below, an application of the Partial Credit Model (PCM) considering the presence of covariates has been performed. This application was conducted using the R package MultOrdRS, which has been customized by including new penalty terms, as discussed in previous chapters. Subsequently, for each of the different penalties considered, a version of the PCM was estimated. This procedure was carried out to examine the impact of the new penalties on the PCM model and evaluate which of them provides the best fit to the data and underlying theoretical models. This approach will allow for a more detailed understanding of the relationships between response variables and covariates, as well as assessing the effectiveness of customized penalties in modeling ordinal responses.

Item	**Thresh.1**	**Thresh.2**	**Thresh.3**	**Thresh.4**
fear	1,8672	1,2251	-0,9509	-2,3392
anxiety	1,9284	0,9158	-0,1231	-1,9352
insomnia	0,4379	0,1462	-0,7156	-2,0463
seismic_concern	1,5685	1,6302	0,1194	-1,5746

Table 2: Threshold Coefficients of model1

Variable	Estimate	Std.Err	Z value	P(>z)
age	-0,0242	0,0104	-2,3166	0,0205
sexMale	-0,6222	0,1892	-3,2887	0,0010
family_disabilitiesYes	-0,6429	0,2682	-2,3974	0,0165
shocks	0,4195	0,0844	4,9704	0,0000
physiological_symptoms	0,4491	0,0826	5,4341	0,0000
decision_timeliness	0,3235	0,0762	4,2460	0,0000
out_of_regionYes	-0,5298	0,2272	-2,3314	0,0197
change_of_residence_earthquakeYes	0,4726	0,2248	2,1025	0,0355
social_media_info	0,2120	0,0793	2,6714	0,0076
app_info	-0,3065	0,0965	-3,1770	0,0015
vehicle_typeNone	-0,5850	0,2823	-2,0724	0,0382

Table 3: Location Effects of model1

Item	Thresh.1	Thresh.2	Thresh.3	Thresh.4
fear	2,2028	1,8222	-1,0727	-2,5304
anxiety	2,2424	1,3867	0,0117	-2,0776
insomnia	0,6936	0,5089	-0,6226	-2,1457
seismic_concern	1,8443	2,2996	0,3343	-1,6849

Table 4: Threshold Coefficients of model2

Variable	Estimate	Std.Err	Z value	P(>z)
age	-0,0298	0,0122	-2,4464	0,0144
sexMale	-0,8111	0,2225	-3,6462	0,0003
family_disabilitiesYes	-0,6937	0,3107	-2,2325	0,0256
shocks	0,6083	0,1013	6,0073	0,0000
physiological_symptoms	0,5432	0,0970	5,5999	0,0000
decision_timeliness	0,2259	0,0820	2,7545	0,0059
out_of_regionYes	-0,6651	0,2663	-2,4979	0,0125
change_of_residence_earthquakeYes	0,6137	0,2625	2,3381	0,0194
TV_info	-0,2535	0,1521	-1,6670	0,0955
social_media_info	0,2528	0,0925	2,7320	0,0063
app_info	-0,3107	0,1105	-2,8110	0,0049
vehicle_typeNone	-0,6923	0,3275	-2,1138	0,0345

Table 5: Location Effects of model2

Item	Thresh.1	Thresh.2	Thresh.3	Thresh.4
fear	2,5069	1,8939	-1,1210	-2,9307
anxiety	2,5839	1,4734	-0,0510	-2,3881
insomnia	0,8731	0,5528	-0,7347	-2,5293
seismic_concern	2,1201	2,5137	0,2796	-1,9326

Table 6: Threshold Coefficients of model3

Variable	Estimate	Std.Err	Z value	P(>z)
age	-0,0296	0,0134	-2,2132	0,0269
sexMale	-0,8766	0,2488	-3,5233	0,0004
family_disabilitiesYes	-0,7764	0,3474	-2,2347	0,0254
shocks	0,6606	0,1129	5,8537	0,0000
physiological_symptoms	0,5668	0,1072	5,2851	0,0000
decision_timeliness	0,2471	0,0915	2,6997	0,0069
out_of_regionYes	-0,7157	0,2966	-2,4129	0,0158
change_of_residence_earthquakeYes	0,6741	0,2937	2,2950	0,0217
social_media_info	0,2808	0,1036	2,7113	0,0067
app_info	-0,3174	0,1229	-2,5829	0,0098
vehicle_typeNone	-0,7671	0,3661	-2,0951	0,0362

Table 7: Location Effects of model3

Item	Thresh.1	Thresh.2	Thresh.3	Thresh.4
fear	2,2690	2,0121	-1,2572	-2,6451
anxiety	2,2612	1,5139	0,0333	-2,1611
insomnia	0,5952	0,4725	-0,6855	-2,1779
seismic_concern	1,8882	2,4483	0,4295	-1,7548

Table 8: Threshold Coefficients of model4

Variable	Estimate	Std.Err	Z value	P(>z)
age	-0,0330	0,0122	-2,6976	0,0070
sexMale	-0,8785	0,2235	-3,9307	0,0001
family_disabilitiesYes	-0,7241	0,3104	-2,3328	0,0197
shocks	0,5608	0,0992	5,6518	0,0000
physiological_symptoms	0,5639	0,0972	5,8013	0,0000
decision_timeliness	0,2177	0,0813	2,6786	0,0074
out_of_regionYes	-0,6931	0,2653	-2,6126	0,0090
change_of_residence_earthquakeYes	0,6199	0,2618	2,3680	0,0179
radio_info	0,3125	0,1504	2,0780	0,0377
app_info	-0,2583	0,1087	-2,3764	0,0175
vehicle_typeNone	-0,7162	0,3268	-2,1914	0,0284

Table 9: Location Effects of model4

Item	Thresh.1	Thresh.2	Thresh.3	Thresh.4
fear	2,4466	1,8877	-1,1072	-2,8430
anxiety	2,5226	1,4334	-0,0413	-2,3471
insomnia	0,8514	0,5818	-0,6730	-2,4528
seismic_concern	2,0434	2,4434	0,2561	-1,9167

Table 10: Threshold Coefficients of model5

Variable	**Estimate**	**Std.Err**	**Z value**	**P(>z)**
age	-0,0269	0,0128	-2,1105	0,0348
sexMale	-0,8624	0,2387	-3,6131	0,0003
family_disabilitiesYes	-0,7743	0,3342	-2,3172	0,0205
shocks	0,6819	0,1098	6,2100	0,0000
physiological_symptoms	0,6030	0,1047	5,7578	0,0000
decision_timeliness	0,2553	0,0884	2,8896	0,0039
out_of_regionYes	-0,6996	0,2849	-2,4560	0,0141
change_of_residence_earthquakeYes	0,6719	0,2822	2,3814	0,0172
social_media_info	0,2499	0,0988	2,5289	0,0114
app_info	-0,3197	0,1182	-2,7038	0,0069
vehicle_typeNone	-0,7629	0,3518	-2,1685	0,0301

Table 11: Location Effects of model5

Item	**Thresh.1**	**Thresh.2**	**Thresh.3**	**Thresh.4**
fear	2,3659	1,8666	-1,1007	-2,7429
anxiety	2,4268	1,4305	-0,0264	-2,2458
insomnia	0,7971	0,5551	-0,6597	-2,3520
seismic_concern	1,9966	2,4148	0,2856	-1,8256

Table 12: Threshold Coefficients of model6

Variable	**Estimate**	**Std.Err**	**Z value**	**P(>z)**
age	-0,0283	0,0133	-2,1325	0,0330
sexMale	-0,8123	0,2449	-3,3165	0,0009
family_disabilitiesYes	-0,7133	0,3429	-2,0805	0,0375
shocks	0,6358	0,1115	5,7013	0,0000
physiological_symptoms	0,5756	0,1070	5,3777	0,0000
decision_timeliness	0,2332	0,0904	2,5787	0,0099
out_of_regionYes	-0,6654	0,2930	-2,2712	0,0231
change_of_residence_earthquakeYes	0,6283	0,2897	2,1690	0,0301
social_media_info	0,2502	0,1018	2,4579	0,0140
app_info	-0,3127	0,1218	-2,5686	0,0102
vehicle_typeNone	-0,7063	0,3612	-1,9552	0,0506

Table 13: Location Effects of model6

Variable	model1	model2	model3	model4	model5	model6
age	-0,0242	-0,0298	-0,0296	-0,033	-0,0269	-0,0283
sexMale	-0,6222	-0,8111	-0,8766	-0,8785	-0,8624	-0,8123
family_disabilitiesYes	-0,6429	-0,6937	-0,7764	-0,7241	-0,7743	-0,7133
shocks	0,4195	0,6083	0,6606	0,5608	0,6819	0,6358
physiological_symptoms	0,4491	0,5432	0,5668	0,5639	0,603	0,5756
decision_timeliness	0,3235	0,2259	0,2471	0,2177	0,2553	0,2332
out_of_regionYes	-0,5298	-0,6651	-0,7157	-0,6931	-0,6996	-0,6654
change_of_residence_earthquakeYes	0,4726	0,6137	0,6741	0,6199	0,6719	0,6283
radio_info	-	-	-	0,3125	-	-
TV_info	-	-0,2535	-	-	-	-
social_media_info	0,212	0,2528	0,2808	-	0,2499	0,2502
app_info	-0,3065	-0,3107	-0,3174	-0,2583	-0,3197	-0,3127
vehicle_typeNone	-0,585	-0,6923	-0,7671	-0,7162	-0,7629	-0,7063

Table 14: Location Effects Comparison

The analysis of the Table 14 displaying estimated coefficients for various models of the Partial Credit Model (PCM) with different penalties reveals how various variables influence the latent construct of 'Earthquake Stress'. These models, identified as:

- model1: it refers to the penalty term of equation (LASSO)
- model2: it refers to the penalty term of equation (Ridge)
- model3: it refers to the penalty term of equation (Fusion With Zero)
- model4: it refers to the penalty term of equation (Selection Only)
- model5: it refers to the penalty term of equation (Fusion Only)
- model6: it refers to the penalty term of equation (Combined: Selection and Fusion)

highlight how different demographic, exposure, and behavioral factors can have significant impacts on the experience of stress following an earthquake.
The variable 'age' shows a consistently negative effect in all models, suggesting that older age may be associated with slightly lower levels of earthquake stress. Although the differences in coefficients are minimal between models, the consistent negative sign indicates a general trend. Male gender has a significantly negative effect, indicating that men may experience less earthquake stress compared to women. The variability in the coefficient across models reflects the impact of variable selection and applied penalties, but the consistent negative sign underscores a gender difference in stress experience.
The presence of disability within the family emerges as a significant factor associated with lower levels of stress. This suggests that families with disabled members may cope with stress more effectively, as they are compelled to adapt to unique circumstances and develop greater resilience.
Variables indicating exposure to shocks, physiological symptoms, and timeliness of decisions all show positive effects, indicating that direct experience of shocks, the presence of physiological stress symptoms, and delays in decision-making can increase earthquake-related stress. Interestingly, there is the inclusion of variables related to information sources (such as 'radio_info,' 'TV_info,' 'social_media_info,' 'app_info') in some models, suggesting how different

information channels can differently influence the level of stress, with some sources even potentially reducing stress (as indicated by the negative coefficients for 'app_info' and 'TV_info' in some models).
The decision to change residence following an earthquake and the presence of a vehicle are associated with higher stress, which could reflect the challenges and uncertainties related to relocation and limited mobility following a disaster.

Model	ℓ_m	**BIC**
model1	-1044,59	2504,39
model2	-1038,94	2493,09
model3	-1035,71	2486,64
model4	-1045,16	2505,53
model5	-1035,58	2486,38
model6	-1036,33	2487,89

Table 15: Model Comparison

The table presented provides a comparison based on two statistical criteria: the logarithm of the likelihood (ℓ_m) and the Bayesian Information Criterion (BIC), for six different models of the Partial Credit Model (PCM) that incorporate various covariates and penalties. These criteria are commonly used to assess the goodness of fit of statistical models, with particular attention to their complexity and penalties for overfitting.
log-likelihood is a measure of how well the model fits the data, with higher values indicating a better fit. However, ℓ_m alone does not account for the complexity of the model, which is regulated in BIC through penalties for the number of estimated parameters.
In analyzing the table, we observe that model5 has the highest logLik (-1035.58), suggesting that it has the best fit to the data among the models considered. This model also has the lowest BIC values (2486.38), indicating that it is the most preferable model in terms of both fit and complexity compared to the other models presented.
It is interesting to note that, although models model3 and model6 have similar log-likelihood values to model5, model5 maintains its advantage in BIC, which might suggest that it has an optimal number of parameters that provide the best balance between fit and complexity.
In summary, based on the criteria presented in this table, model5 stands out as the most effective model for describing the latent construct 'Earthquake Stress' in the context of the available data.

The results obtained from the application of Partial Credit Models (PCMs) with various penalties provide important empirical confirmation of the theoretical predictions discussed in previous chapters. In particular, model5 emerges as the optimal model, reflecting superior fit to the data when compared to the other models. Despite its supremacy, it is noteworthy that the other models also show commendable performance, suggesting that the variations introduced through different penalties have been beneficial in refining the goodness of fit and understanding of the 'Earthquake Stress' construct. The consistency between the simulation results and practical application reinforces the validity of the methodologies adopted and supports the utility of introducing new penalty terms. These terms not only improve the parsimony and accuracy of PCM models but also offer greater flexibility in modeling the complexity of psychometric phenomena, demonstrating their value in advanced statistical analysis.

5 Conclusion

In conclusion, the analysis of different models of the Partial Credit Model (PCM) for assessing the latent construct 'Earthquake Stress' has provided significant and wide-ranging results. The item selection[2] process employed 'bruteforce' and 'backward' methods, resulting in the identification of an optimal combination of items that accurately reflects the construct under examination. This rigorous selection process, supported by the increase in Cronbach's alpha value, has highlighted the importance of selecting the most informative and consistent items for measuring the construct.

Penalizations, both classical and new, played a crucial role in refining the models. The results of simulation and practical application confirmed that the introduction of new penalty terms improves the parsimony and accuracy of PCM models. This is particularly evident with model5, which showed the best fit to the data, as indicated by lower values of AIC and BIC, and maintained this superiority even compared to models with similar logLik values, suggesting an optimal balance between data fit and model complexity.

It is important to note that the entire study can also be extended to the case of 'Response Style' (RS) [13]. In fact, the modification of the MultOrdRS package involved penalized estimation functions for both the Response Style case and the case without RS.

[2]The items subject to selection are not included in psychometric scales, and this methodology is applicable only in cases where the latent trait is unidimensional.

References

[1] A. Agresti. *Analysis of Ordinal Categorical Data*. Wiley, 2 edition, 2010.

[2] A. Agresti. *Categorical Data Analysis*. Wiley, 3 edition, 2013.

[3] D. Andrich. A probabilistic irt model for unfolding preference data. *Applied Psychological Measurement*, 13(2):193–216, 1989.

[4] D. Andrich. A rating formulation for ordered response categories. *Psychometrika*, 75:292–308, 2010.

[5] D. Andrich. Sufficiency and conditional estimation of person parameters in the polytomous rasch model. *Psychometrika*, 75:292–308, 2010.

[6] F. B. Baker, S. Kim, and Others. *The Basics of Item Response Theory Using R*. Springer, 2017.

[7] R. P. Chalmers. mirt: A multidimensional item response theory package for the r environment. *Journal of statistical Software*, 48, 2012.

[8] A. Cola, M. Iannario, and G. Tutz. Item selection method in irt models. Technical article, 2024.

[9] L. J. Cronbach. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3):297–334, 1951.

[10] A. Dobson. *An Introduction to Generalized Linear Model*. Springer, 2 edition, 1990.

[11] A. E. Hoerl and R.W. Kennard. Ridge regresssion: Biased estimation for nonorthogonal problems. *Technometrics,*, 12:55–67, 1970.

[12] M. Iannario, A. C. Monti, D. Piccolo, and E. Ronchetti. Robust inference for ordinal response models. *Electronic Journal of Statistics*, 11(2):3407–3445, 2017.

[13] T. R. Johnson. On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style. *Psychometrika*, 68:563–583, 2003.

[14] C. Li and M. Hansen. Limited information goodness of fit testing of hierarchical item factor models. *British Journal of Mathematical and Statistical Psychology*, 66(2), 2013.

[15] G. N. Masters. A rasch model for partial credit scoring. *Psychometrika*, 47(2):149–174, 1982.

[16] G. N. Masters and B. D. Wright. The essential process in a family of measurement models. *Psychometrika*, 49:529–544, 1984.

[17] E. Muraki. A generalized partial credit model: Application of an em algorithm. *Applied Psychological Measurement*, 16(2):159–176, 1992.

[18] E. Muraki and M. Muraki. Generalized partial credit model. In *Handbook of Item Response Theory*, pages 127–137. Chapman and Hall/CRC, 2016.

[19] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT press, 1 edition, 2012.

[20] L. Palazzo, M. Iannario, and F. Palumbo. Integrated assessment of financial knowledge through a latent profile analysis. *Behaviormetrika*, 51:319–339, 2024.

[21] D. Piccolo. *Statistica*. Il Mulino, 3 edition, 2010.

[22] General Data Protection Regulation. *Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio sulla Protezione dei Dati*. European Union Publications Office, 2016.

[23] Fumiko Samejima. Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*, 1969.

[24] Fumiko Samejima. Graded response models. In *Handbook of item response theory*, pages 95–107. Chapman and Hall/CRC, 2016.

[25] G. Schauberger. *MultOrdRS: Model Multivariate Ordinal Responses Including Response Styles*, 2024. R package version 0.1-3.

[26] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[27] W. G. Smith. Does gender influence online survey participation? a record-linkage analysis of university faculty online survey response behavior. *Online submission*, 2008.

[28] D. Thissen and L. Steinberg. A taxonomy of item response models. *Psychometrika*, 51(4):567–577, 1986.

[29] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–88, 2010.

[30] G. Tutz. *Regression for Categorical Data*. Cambridge University Press, 2012.

[31] G. Tutz. On the structure of ordered latent trait models. *Journal of Mathematical Psychology*, 96, 2020.

[32] G. Tutz. Ordinal regression: a review and a taxonomy of models. *WIRES Computational Statistics, to appear*, 2021.

[33] G. Tutz. Ordinal regression: A review and a taxonomy of models. *Wiley Interdisciplinary Reviews: Computational Statistics*, 14(2):e1545, 2022.

[34] G. Tutz. Penalties. Technical article, 2024.

[35] G. Tutz and G. Schauberger. Uncertainty in latent trait models. *Applied Psychological Measurement*, 44:447–464, 2020.

[36] G. Tutz, G. Schauberger, and M. Berger. Response styles in the partial credit model. *Applied Psychological Measurement*, 42:407–427, 2018.

Appendix

A Encoding Of The Variables

- **age**: Participant's age;
- **sex**: Participant's sex;
- **marital_status**: Participant's marital status;
- **residence**: City of residence of the participant;
- **education**: Participant's level of education;
- **occupation**: Participant's occupation;
- **out_of_region_employment**: Indicates whether the participant works or studies outside the Campania region;
- **family_members**: Number of members in the participant's household;
- **family_disabilities**: Presence of disabled persons in the family;
- **house_floor**: Floor of the participant's residence;
- **earthquake80**: Participant's direct experience of the earthquake in the '80s;
- **political_orientation**: Participant's political orientation;
- **shocks**: Perception of the intensity of seismic shocks;
- **fear**: Level of fear experienced during seismic events;
- **anxiety**: Level of anxiety during seismic events;
- **physiological_symptoms**: Presence of physiological symptoms (such as dizziness, nausea, headache) during seismic events;
- **decision_timeliness**: Perception of the timeliness of decisions made during seismic events;
- **insomnia**: Level of insomnia during seismic events;
- **seismic_concern**: Level of concern in relation to seismic events;
- **abroad**: Participant's experience of staying abroad;
- **out_of_region**: Consideration of working or studying out of region;
- **out_of_region_earthquake**: Consideration of working or studying out of region due to seismic phenomena;

- **change_of_residence**: Consideration of changing residence for any reason;
- **change_of_residence_earthquake**: Consideration of changing residence due to seismic phenomena;
- **red_zone_frequency**: Frequency with which the participant is in high seismic risk areas;
- **radio_info**: Usefulness of information received from the radio on seismic activities;
- **TV_info**: Usefulness of information received from the TV on seismic activities;
- **social_media_info**: Usefulness of information received from social media on seismic activities;
- **newspaper_info**: Usefulness of information received from newspapers on seismic activities;
- **app_info**: Usefulness of information received via apps on seismic activities;
- **municipal_institutions_trust**: Trust in municipal institutions;
- **regional_institutions_trust**: Trust in regional institutions;
- **national_institutions_trust**: Trust in national institutions;
- **INGV_trust**: Trust in the National Institute of Geophysics and Volcanology (INGV);
- **security**: Perception of safety in public structures;
- **reception_centers**: Knowledge of emergency reception points;
- **property_house**: Ownership of the dwelling;
- **housing_type**: Type of housing;
- **elevator**: Presence of an elevator in the dwelling;
- **n_vehicles**: Number of vehicles owned;
- **vehicle_type**: Type of vehicle owned;
- **end_of_month**: Ability to make ends meet;
- **salary**: Gross annual income;

B Graphs

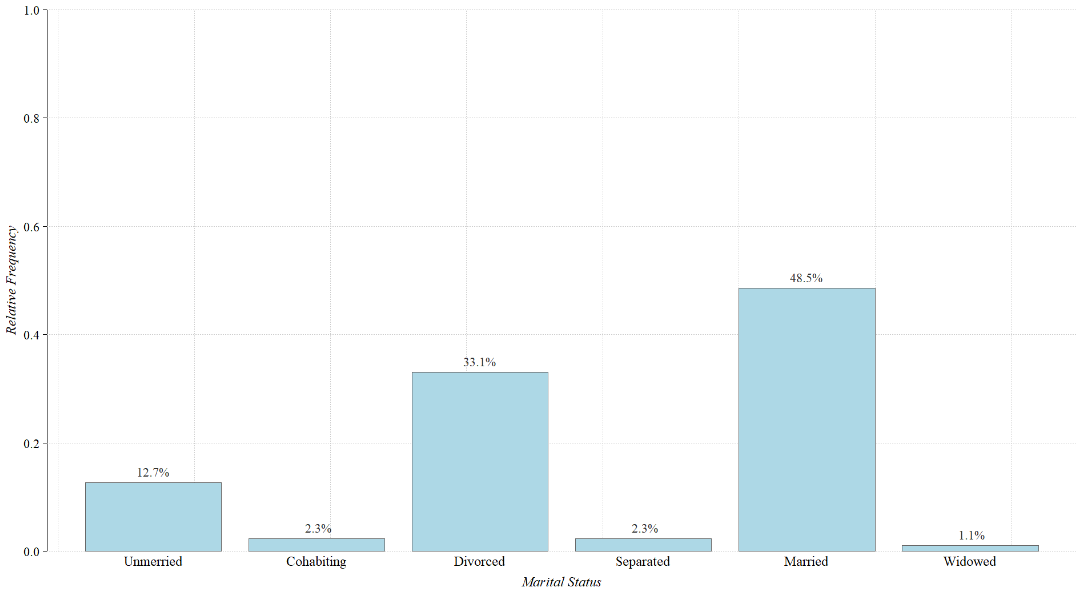

Figure 25: Marital Status Distribution 1

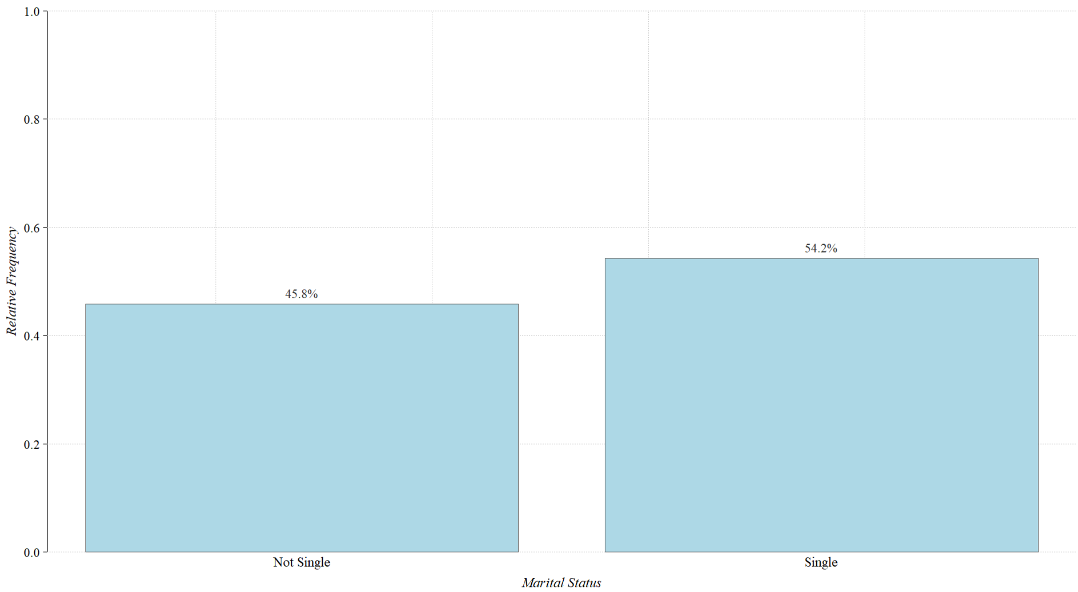

Figure 26: Marital Status Distribution 2

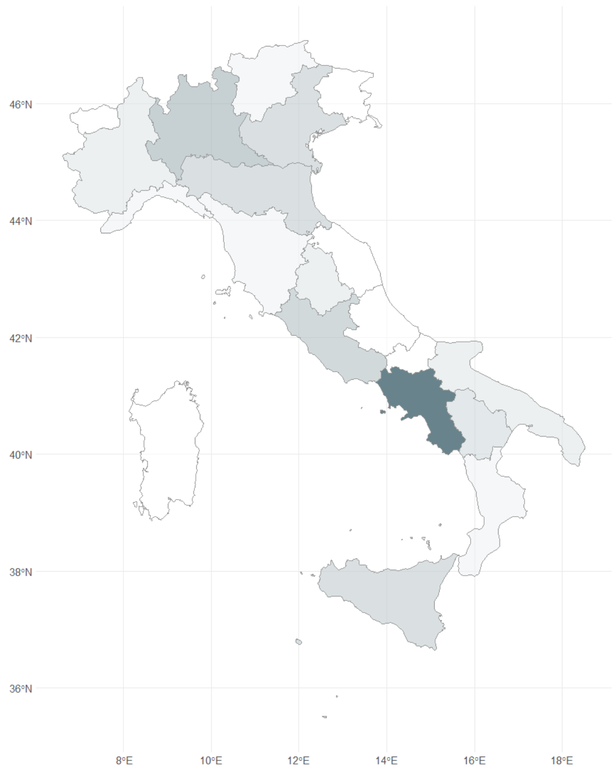

Figure 27: Residence Distribution 1

Figure 28: Residence Distribution 2

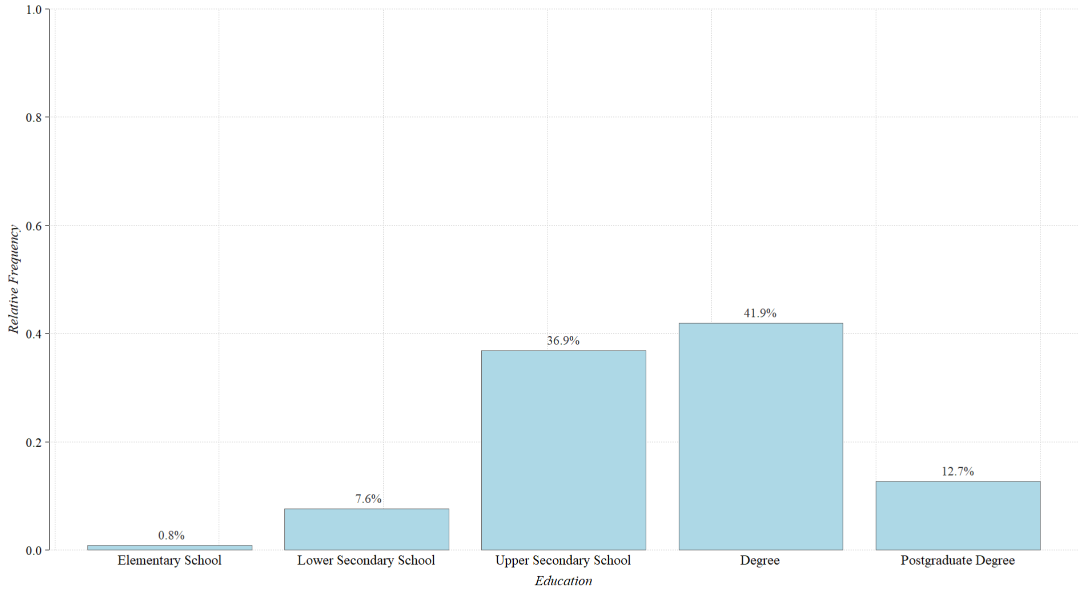

Figure 29: Education Distribution 1

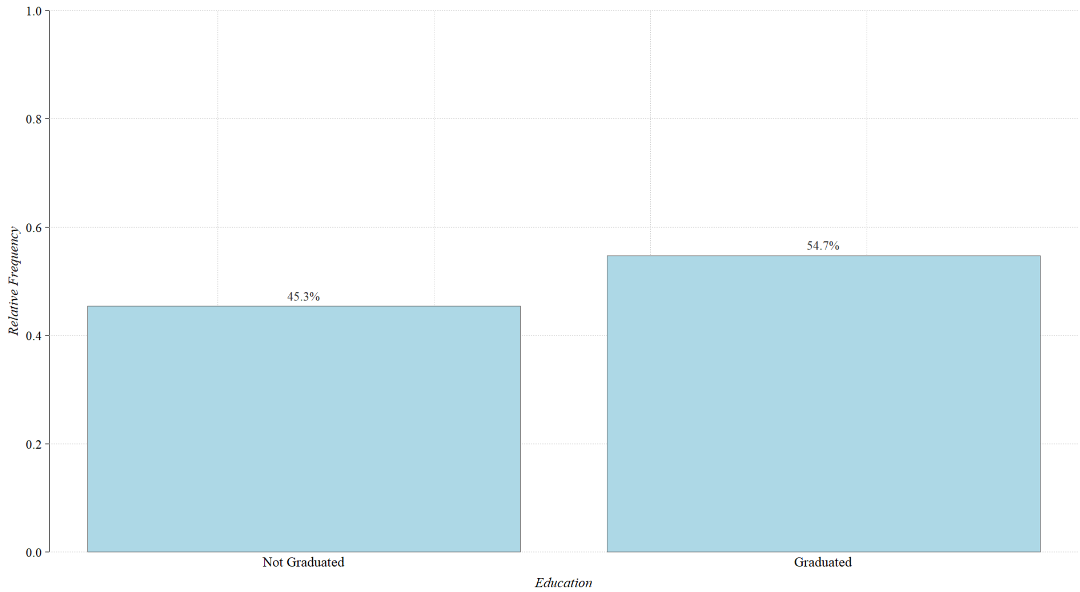

Figure 30: Education Distribution 2

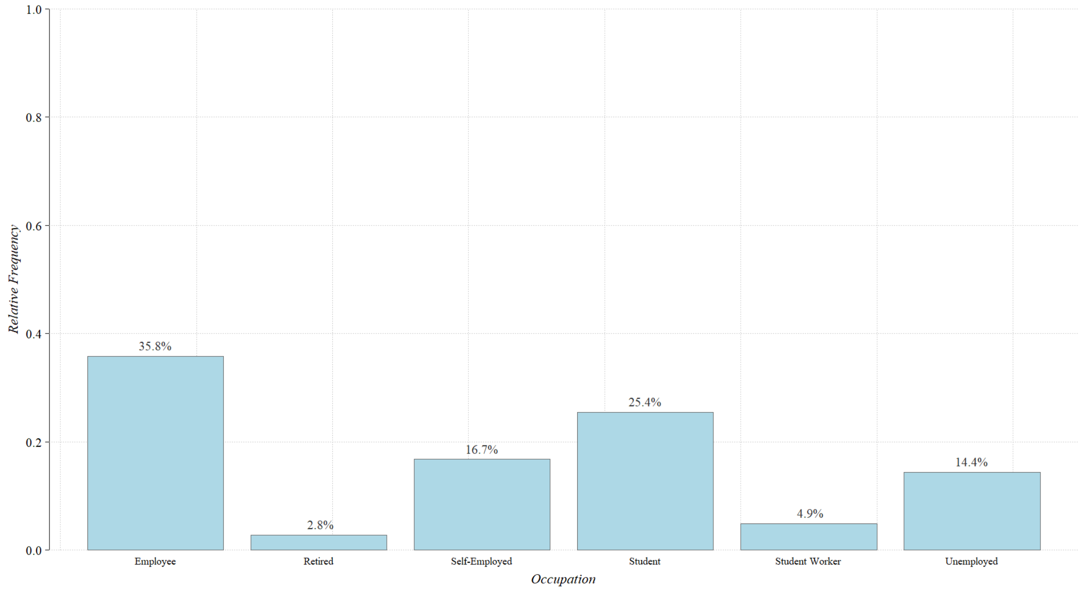

Figure 31: Occupation Distribution 1

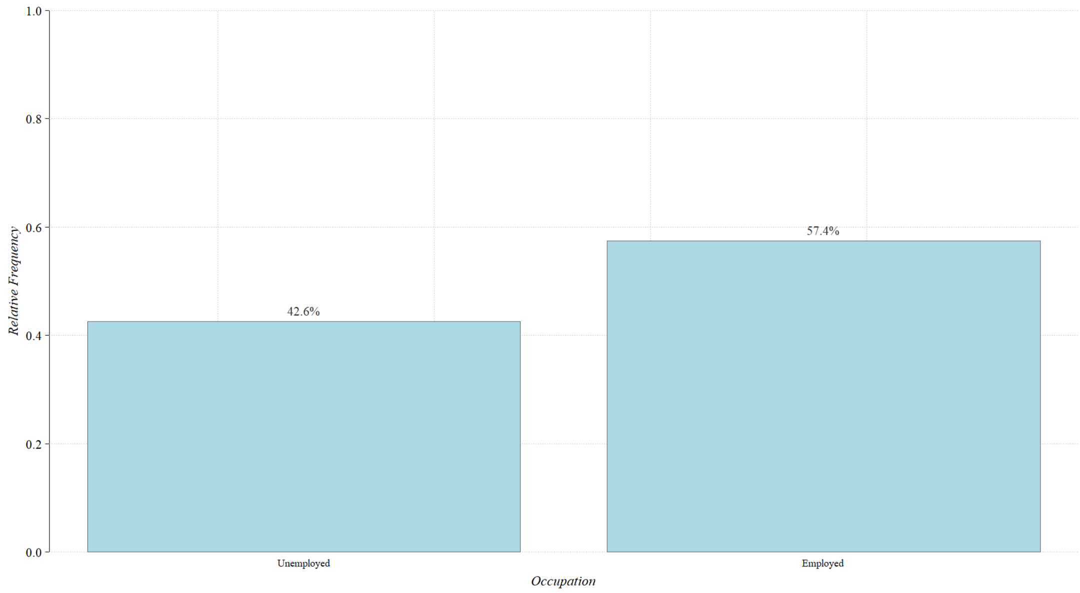

Figure 32: Occupation Distribution 2

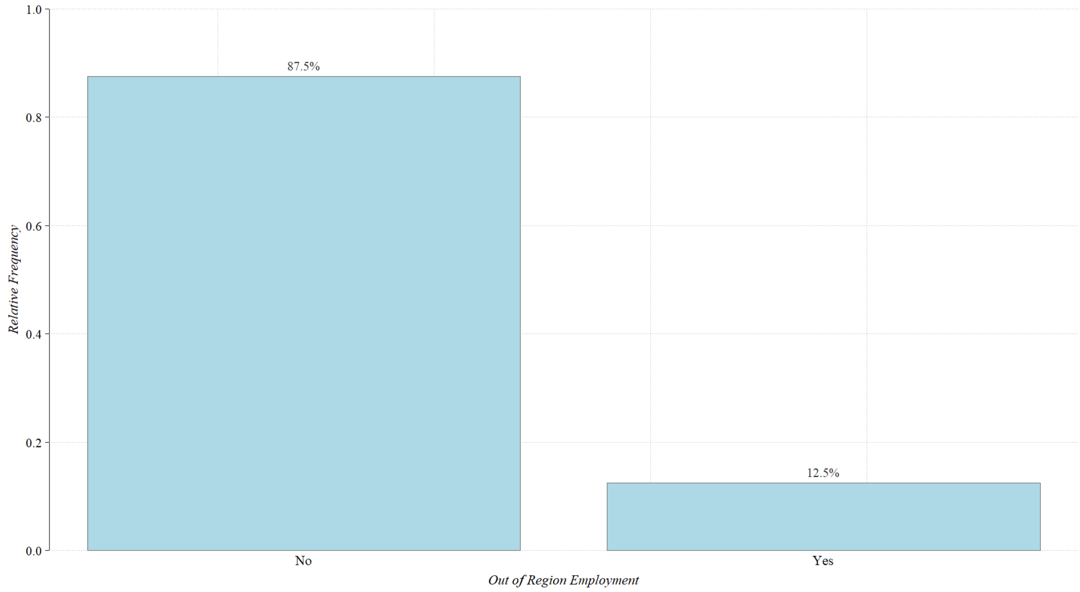

Figure 33: Out of Region Employment Distribution

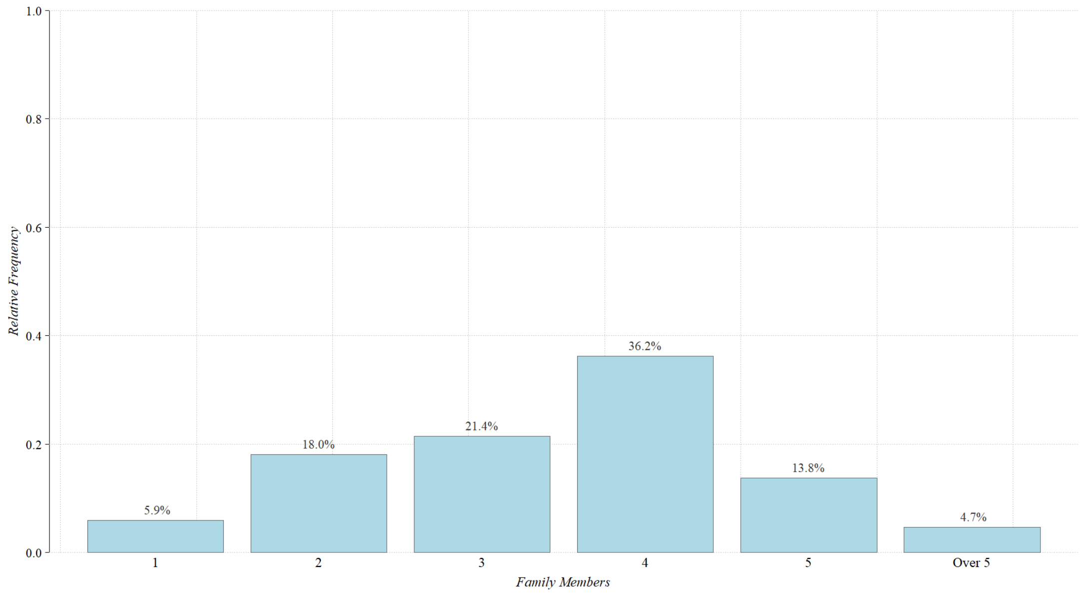

Figure 34: Family Members Distribution 1

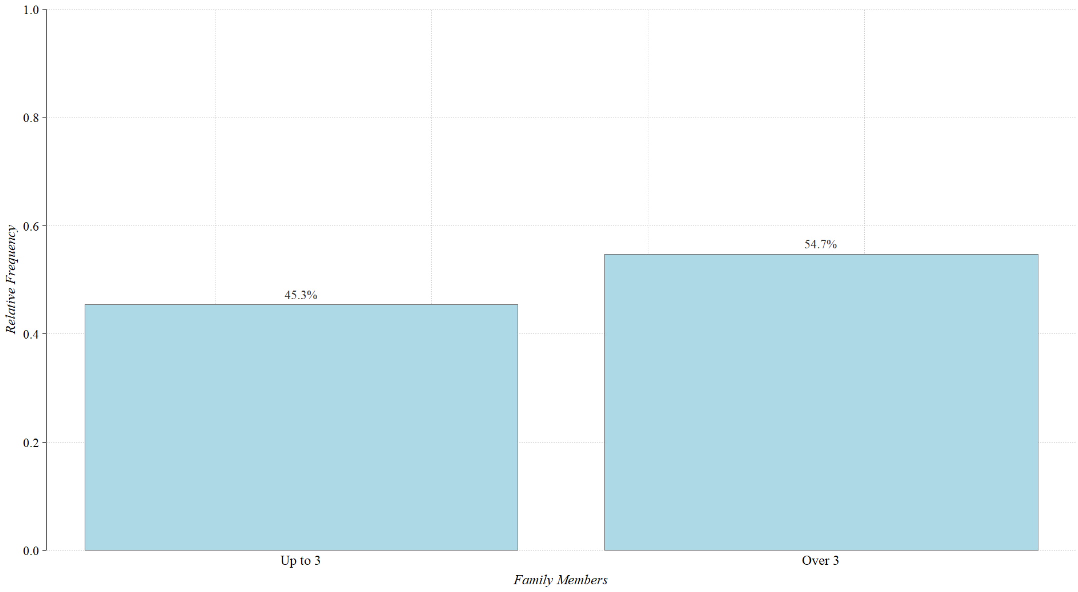

Figure 35: Family Members Distribution 2

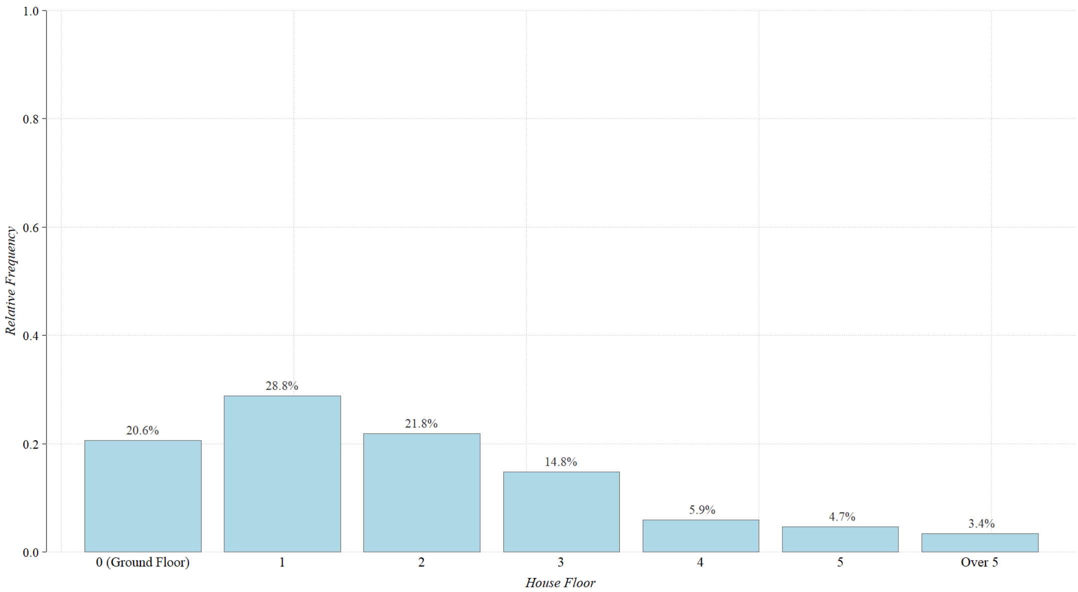

Figure 36: House Floor Distribution 1

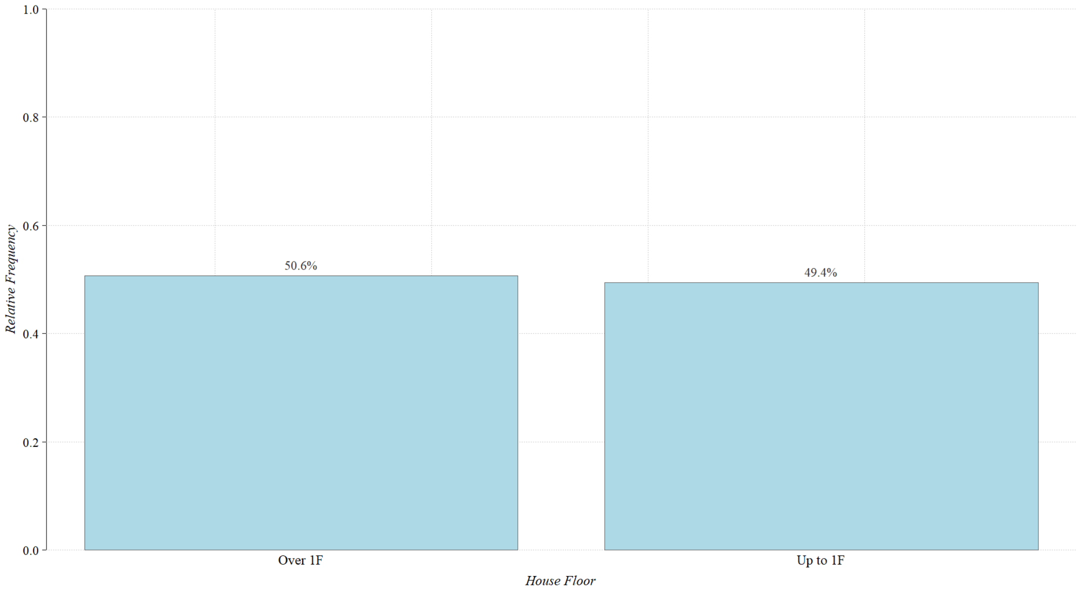

Figure 37: House Floor Distribution 2

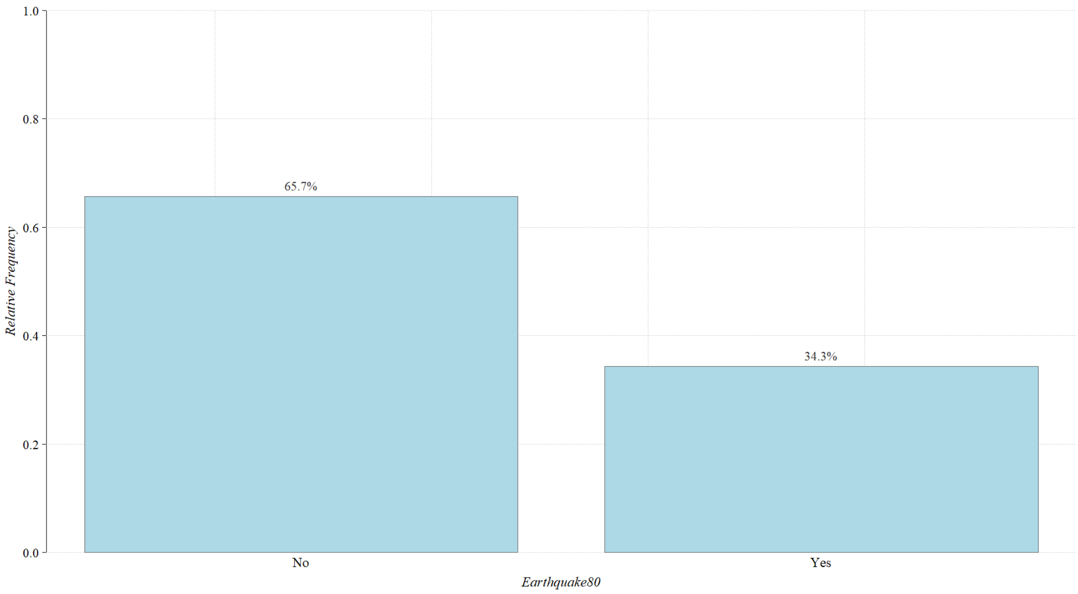

Figure 38: Earthquake80 Distribution

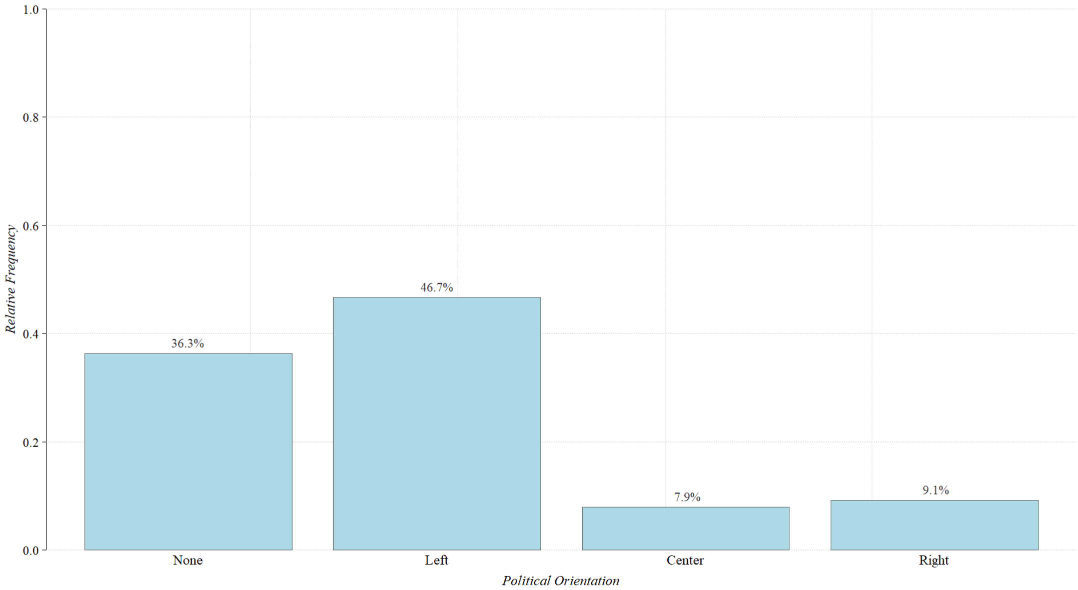

Figure 39: Political Orientation Distribution

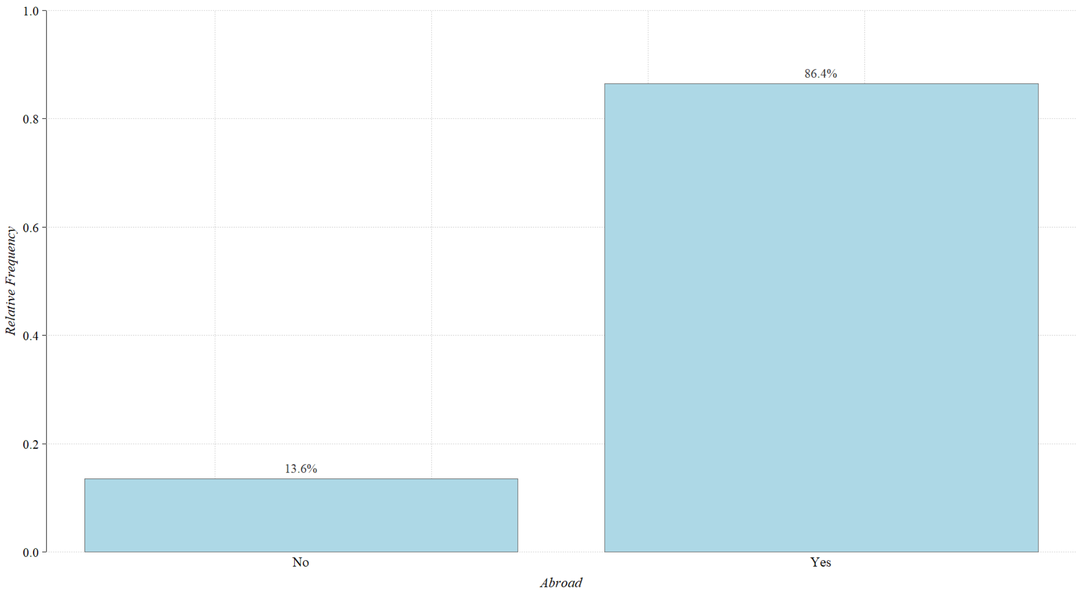

Figure 40: Abroad Distribution

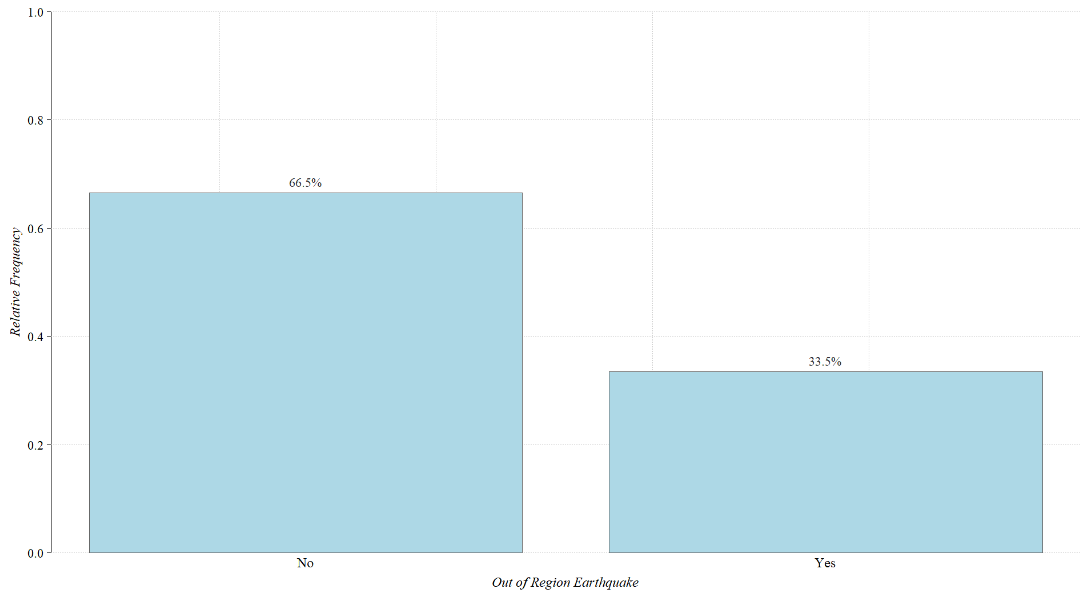

Figure 41: Out of Region Earthquake Distribution

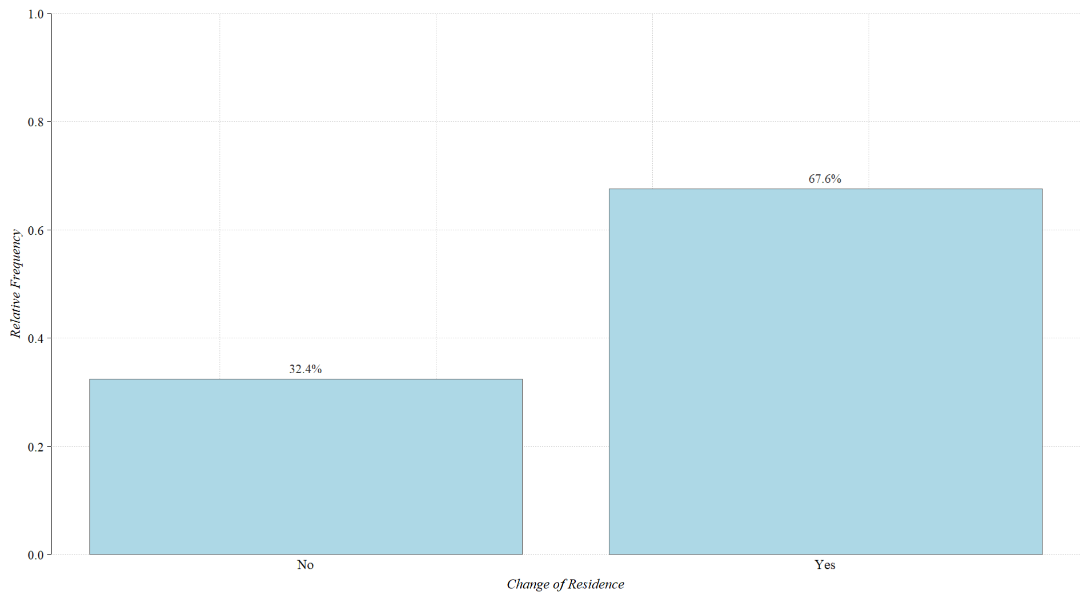

Figure 42: Change of Residence Distribution

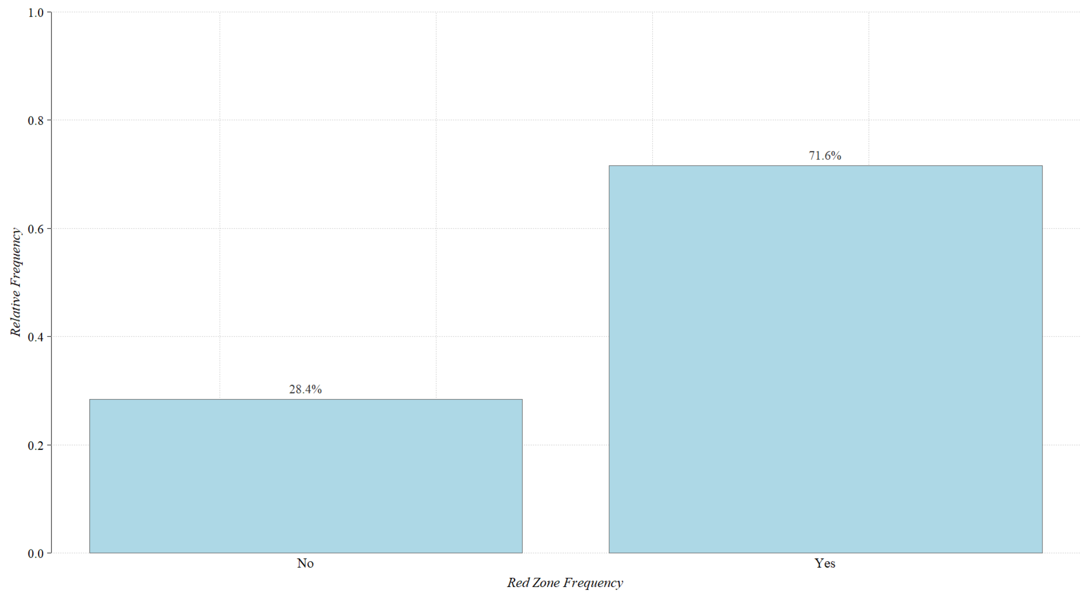

Figure 43: Red Zone Frequency Distribution

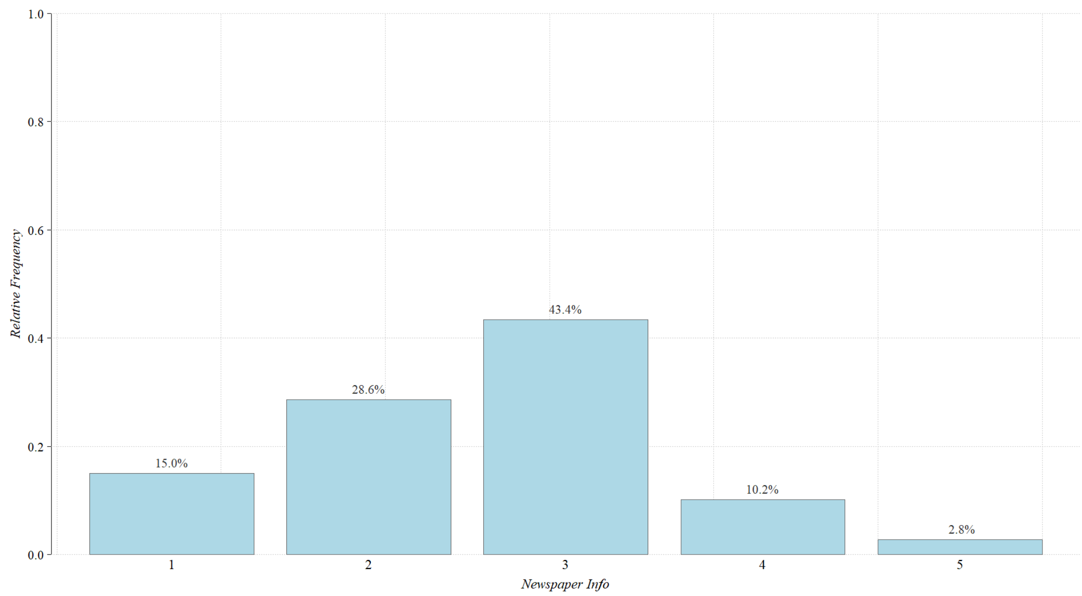

Figure 44: Newspaper Info Distribution

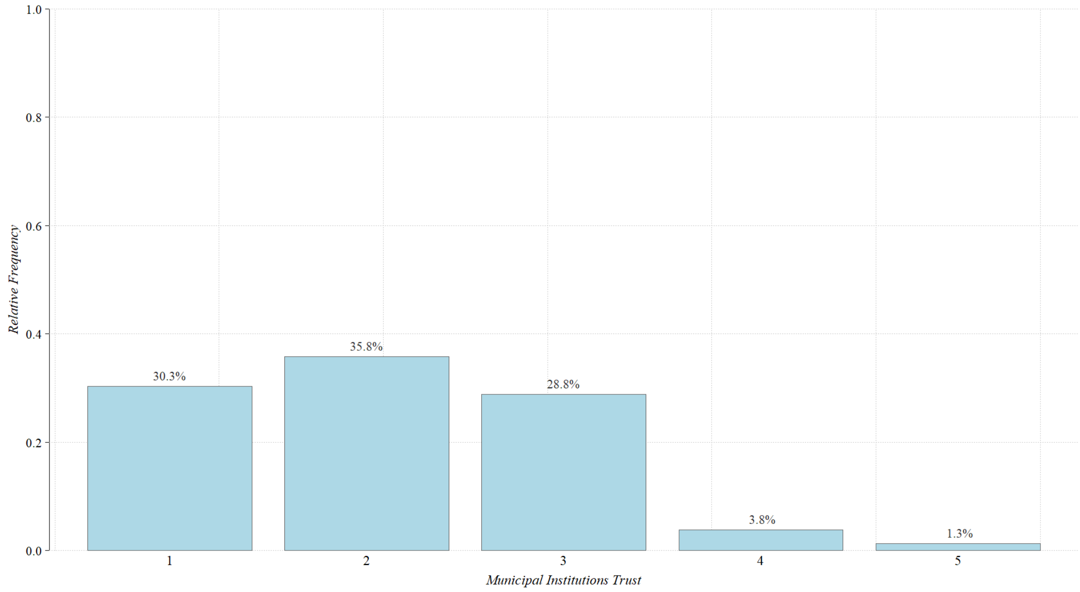

Figure 45: Municipal Institutions Trust Distribution

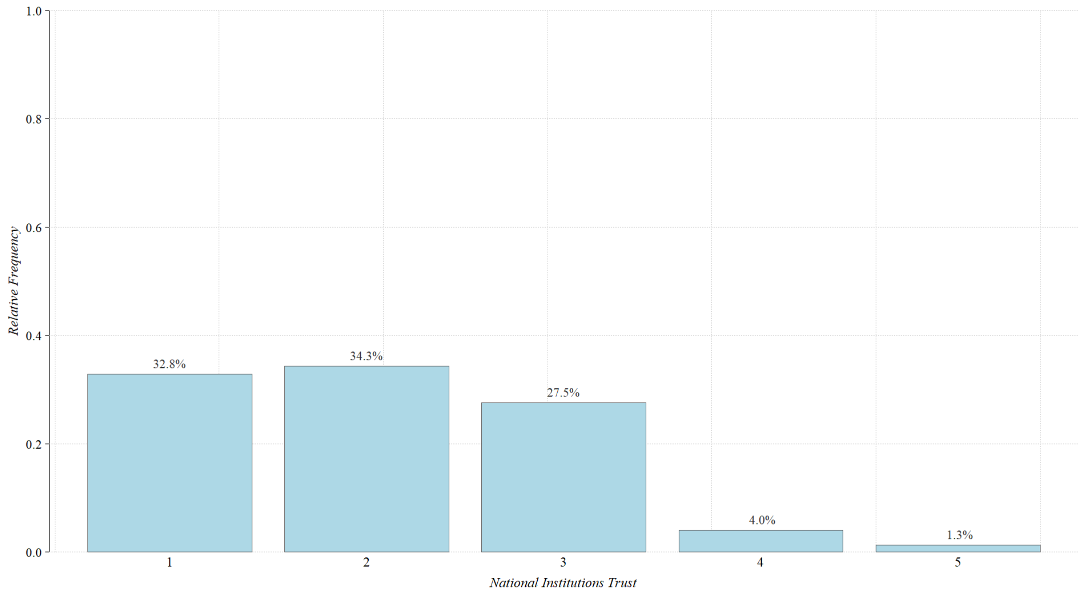

Figure 46: National Institutions Trust Distribution

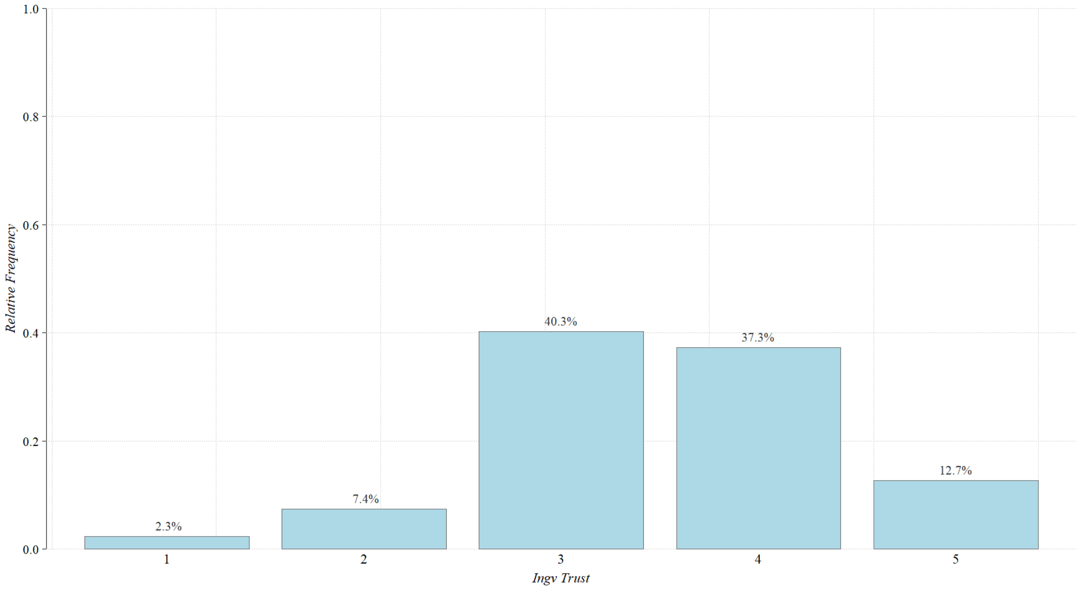

Figure 47: Ingv Trust Distribution

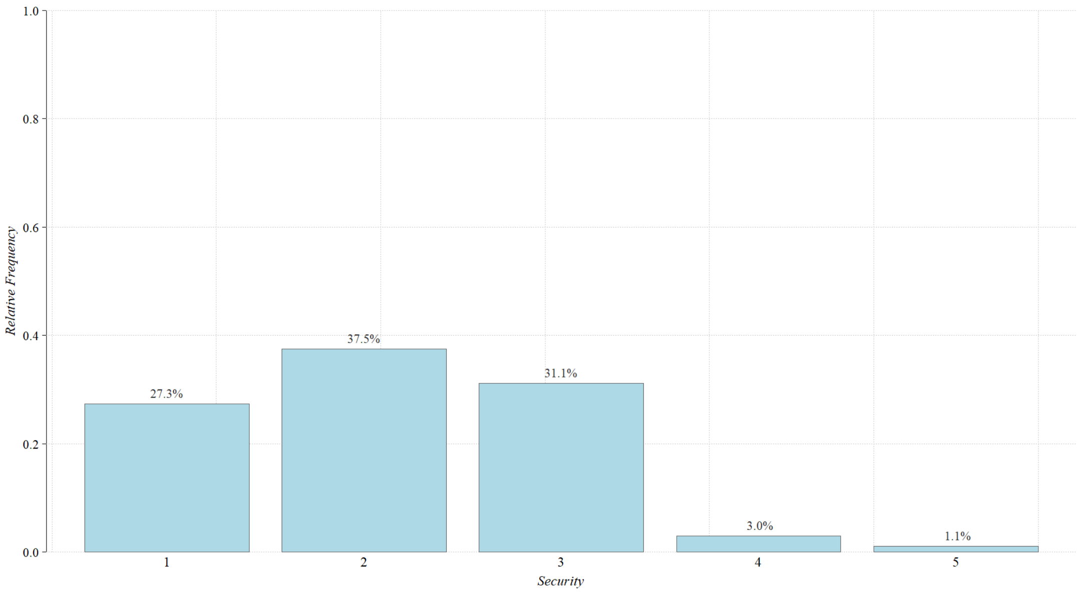

Figure 48: Security Distribution

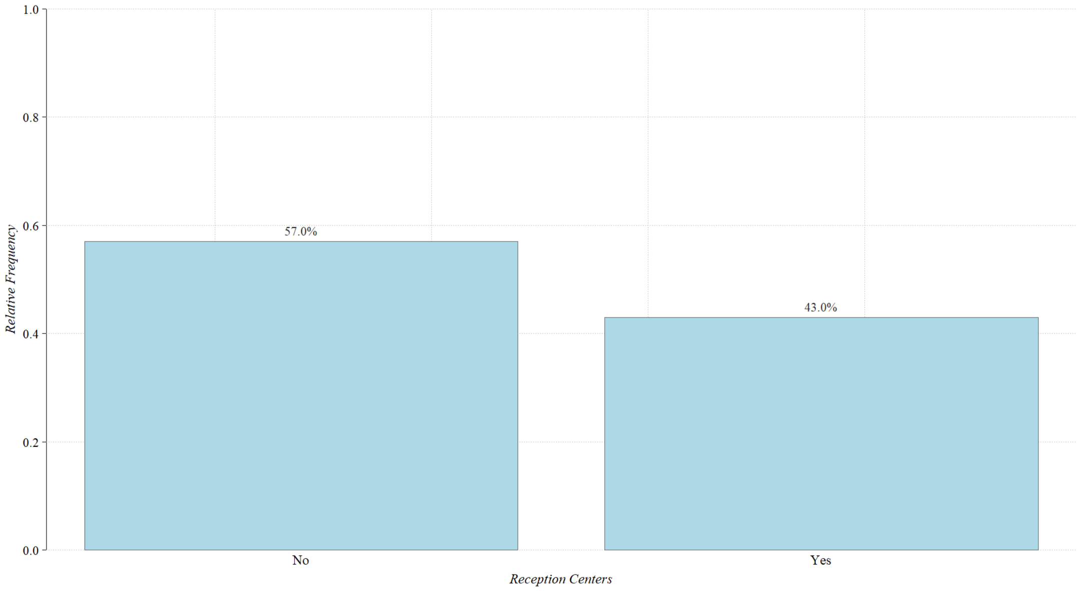

Figure 49: Reception Centers Distribution

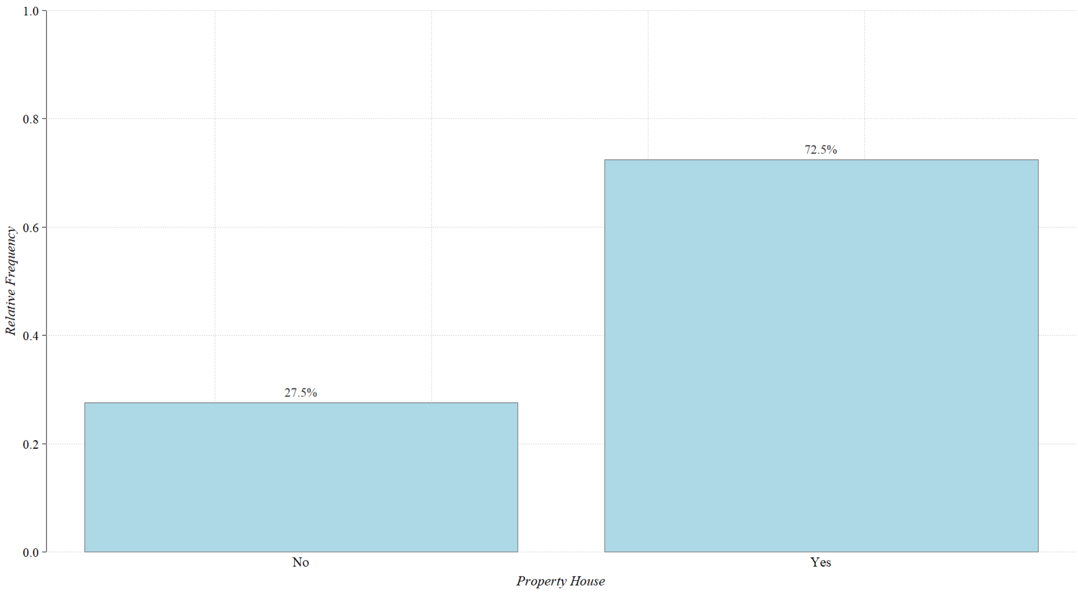

Figure 50: Property House Distribution

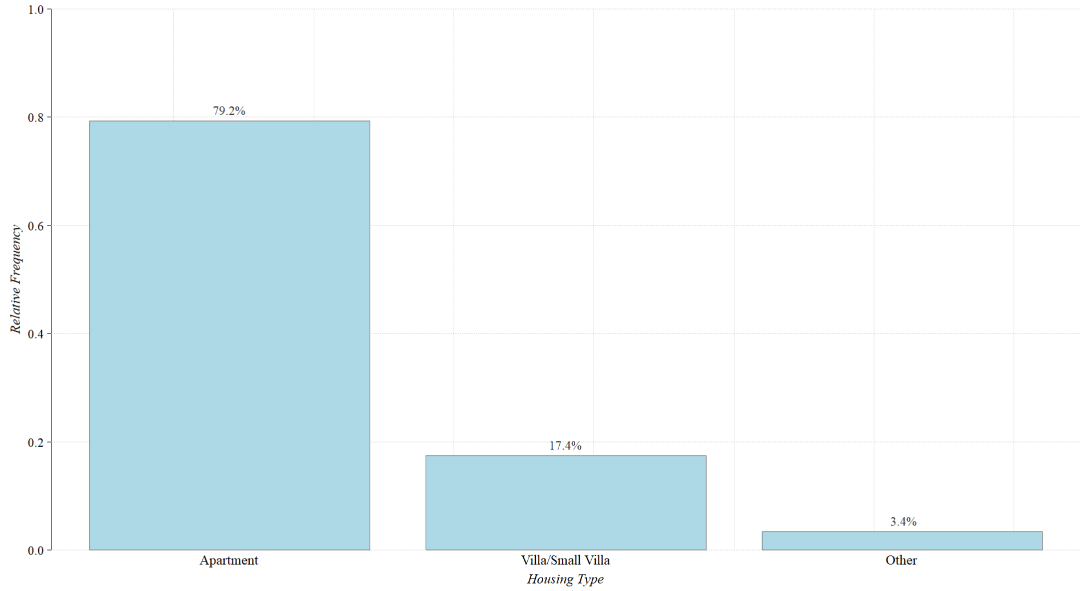

Figure 51: Housing Type Distribution 1

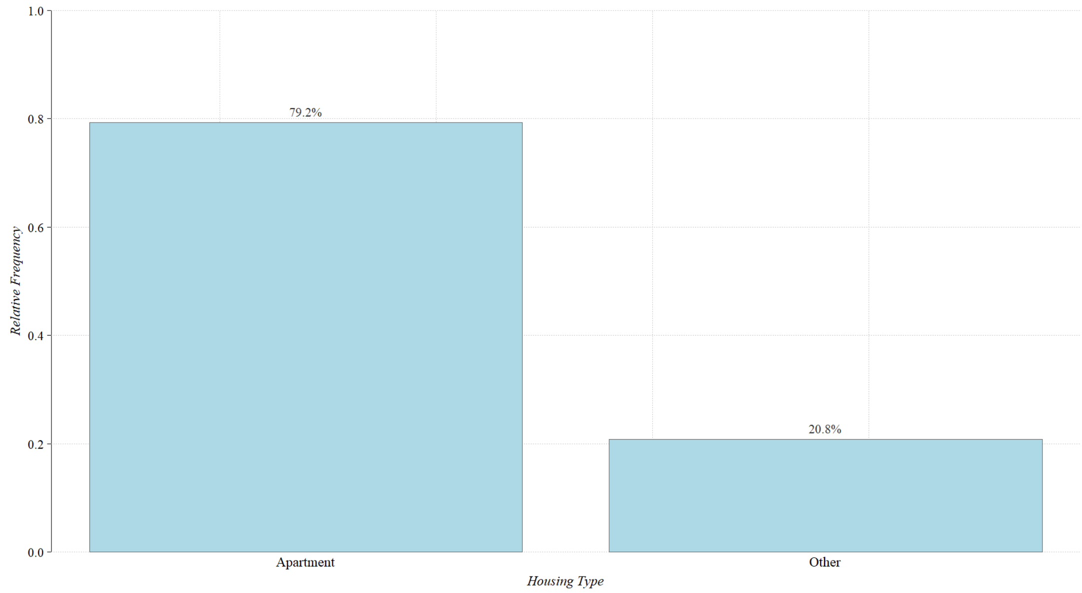

Figure 52: Housing Type Distribution 2

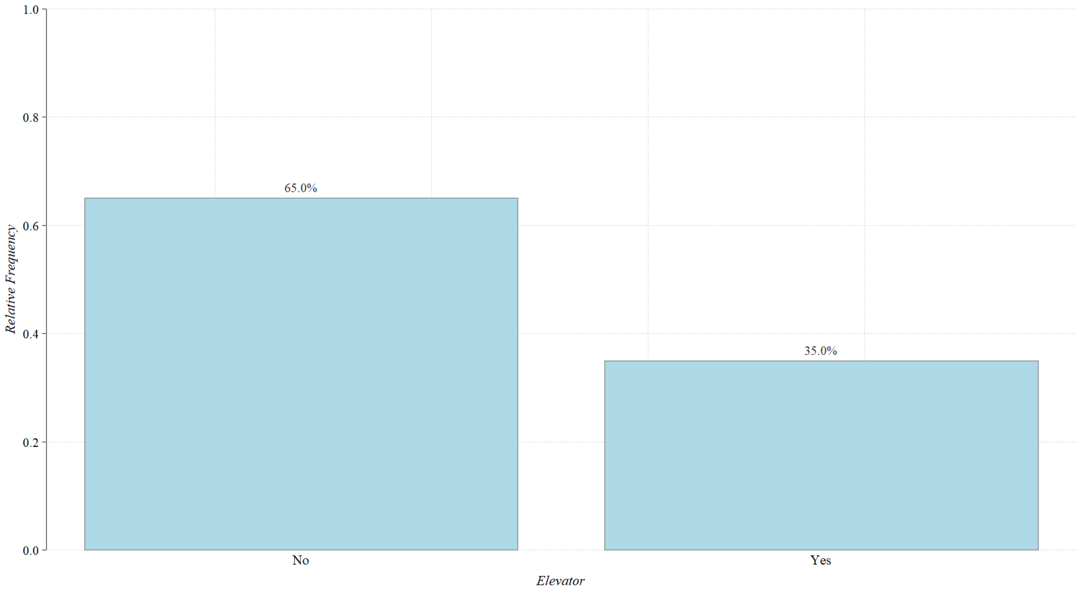

Figure 53: Elevator Distribution

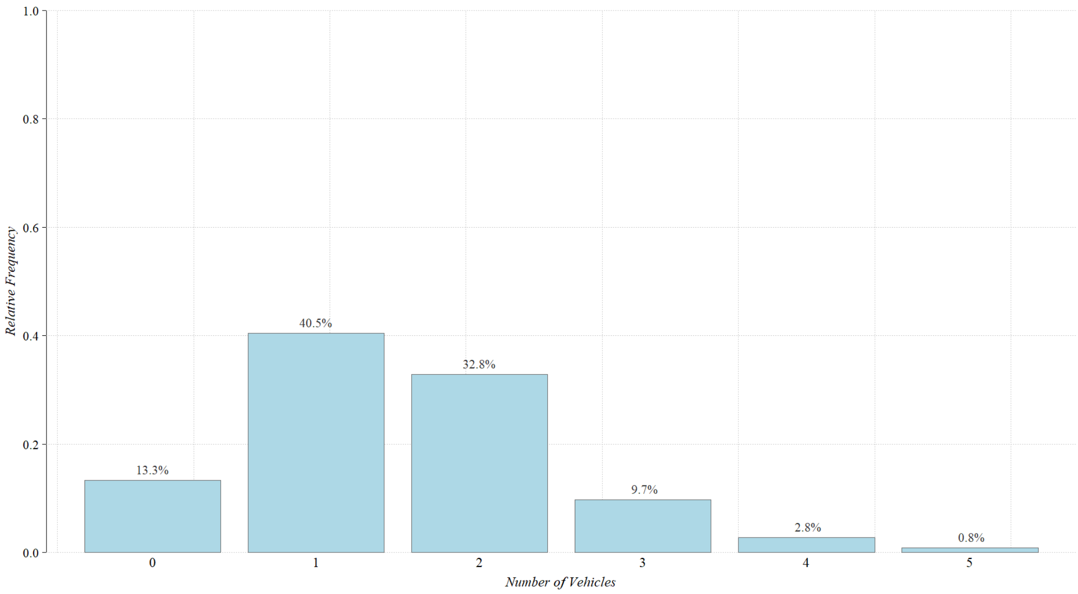

Figure 54: Number of Vehicles Distribution

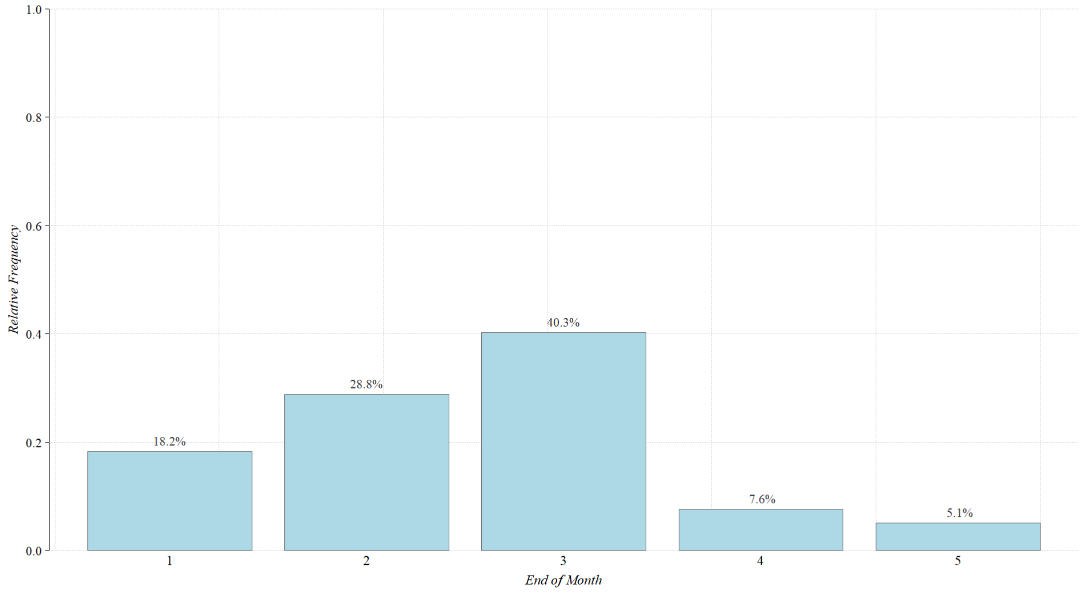

Figure 55: End of Month Distribution

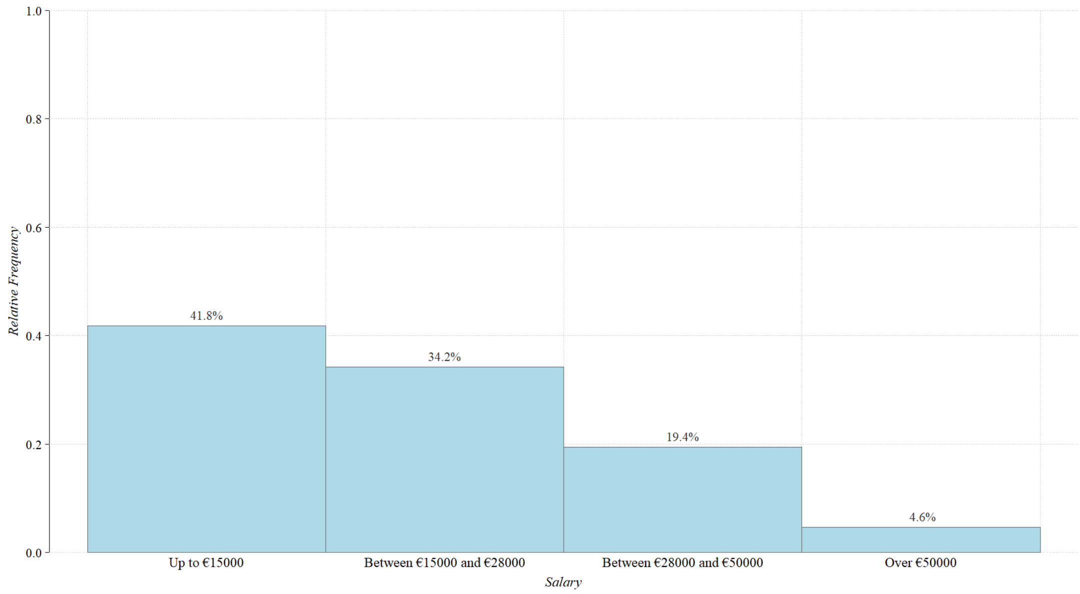

Figure 56: Salary Distribution

C Test

Item	Variable	Chi-Squared	p-value
fear	sex	34,52	0,0000
fear	occupation	13,65	0,0085
fear	political_orientation	22,38	0,0335
fear	shocks	269,67	0,0000
fear	physiological_symptoms	205,94	0,0000
fear	decision_timeliness	58,76	0,0000
fear	abroad	16,86	0,0021
fear	out_of_region_earthquake	33,16	0,0000
fear	change_of_residence_earthquake	54,34	0,0000
fear	radio_info	27,01	0,0414
anxiety	sex	43,13	0,0000
anxiety	shocks	168,71	0,0000
anxiety	physiological_symptoms	204,46	0,0000
anxiety	decision_timeliness	67,67	0,0000
anxiety	out_of_region_earthquake	53,00	0,0000
anxiety	change_of_residence_earthquake	67,14	0,0000
anxiety	social_media_info	32,77	0,0079
anxiety	app_info	31,19	0,0127
anxiety	end_of_month	32,44	0,0088
insomnia	sex	36,28	0,0000
insomnia	marital_status	16,49	0,0024
insomnia	occupation	12,39	0,0147
insomnia	family_members	10,77	0,0293
insomnia	earthquake80	11,21	0,0243
insomnia	shocks	192,42	0,0000
insomnia	physiological_symptoms	199,10	0,0000
insomnia	decision_timeliness	65,12	0,0000
insomnia	out_of_region_earthquake	73,11	0,0000
insomnia	change_of_residence_earthquake	83,41	0,0000
insomnia	red_zone_frequency	11,12	0,0252
insomnia	newspaper_info	29,16	0,0229
insomnia	app_info	44,97	0,0001
insomnia	reception_centers	20,00	0,0005
insomnia	end_of_month	31,24	0,0125
seismic_concern	age	259,57	0,0227
seismic_concern	sex	42,22	0,0000
seismic_concern	education	15,26	0,0042
seismic_concern	out_of_region_employment	16,40	0,0025
seismic_concern	shocks	110,83	0,0000
seismic_concern	physiological_symptoms	126,50	0,0000
seismic_concern	decision_timeliness	57,78	0,0000
seismic_concern	out_of_region_earthquake	76,20	0,0000
seismic_concern	change_of_residence_earthquake	96,21	0,0000
seismic_concern	red_zone_frequency	10,63	0,0311
seismic_concern	app_info	36,94	0,0021
seismic_concern	INGV_trust	29,21	0,0226
seismic_concern	end_of_month	29,19	0,0227
seismic_concern	salary	25,46	0,0128

Table 16: Indipendence Test

D Questionnaire and Dataset

SCAN ME