LEARNING ABOUT

2 1.2

INDICATOR

SDG Indicator 2.1.2 – Using the Food Insecurity Experience Scale (FIES)

Lesson 3: Statistical validation of FIES data

Text-only version

The interactive version of this lesson is available free of charge at: www.fao.org/elearning

Food and Agriculture Organization of the United Nations

SUSTAINABLE DEVELOPMENT GOALS

working for Zero Hunger

In this lesson

Statistical validation of FIES data

This lesson introduces the basic concepts of the food insecurity experience scale, and the validation of FIES data using the Rasch model.

It explains how to estimate respondent and item parameters, and how to interpret the Rasch model-based outputs.

Learning objectives

At the end of this lesson, you will be able to:

- recognize the characteristics of the food insecurity scale;
- describe the key concepts underlying the validation of FIES data using the Rasch model;
- explain parameter estimation for both item and respondent;
- explain how statistical validation is used to assess the quality of data;
- interpret Rasch model-based statistics: infit, outfit, residual correlation, and reliability.

Introduction to FIES data analysis

Once the FIES survey module has been administered, the **data collected** must be properly **analysed** to produce **meaningful results**. The analysis of FIES data involves:

➲ **Parameter estimation:** To calculate the level of food insecurity severity associated with each question and each respondent.

➲ **Statistical Validation:** To assess whether, depending on the quality of the data collected, the **measure is valid**, (i.e. the data are consistent with the theoretical assumptions that inform the model).

➲ **Calculation of food insecurity prevalence estimates:** To calculate a measure of severity of the food insecurity condition experienced by each respondent, based on their answers to the eight FIES questions. This is then used to **estimate** the **prevalence** of food insecurity at moderate and/or severe levels in the **population**.

This lesson will **focus on parameter estimation** and **statistical validation**, which must precede the calculation of prevalence estimates, covered in Lesson 4.

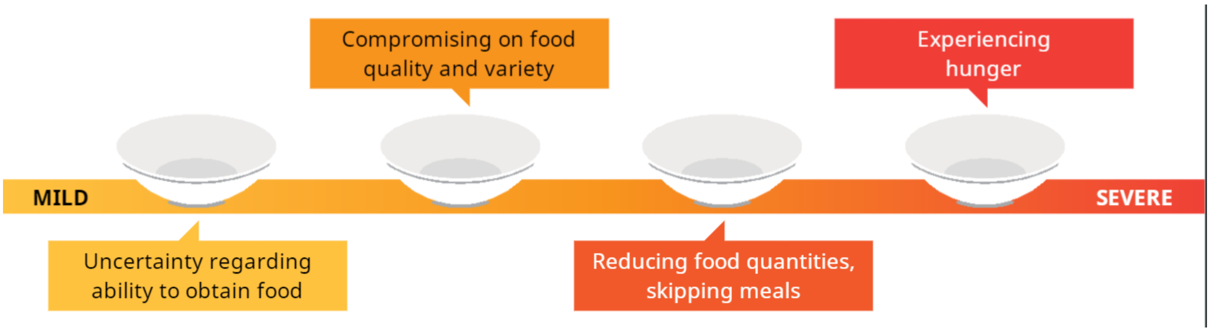
First steps ➥ Denis is the analyst from Grace's team who will use the FIES data to produce an estimate of the national prevalence of food insecurity for SDG monitoring.

🗨 *"Before we can use the FIES data we've collected for national SDG monitoring, they need to be statistically **validated** to make sure that the prevalence estimates we produce are valid and reliable. First let's take some time to look at the **concepts** that underlie this analysis." Denis*

Key concepts

Latent trait and continuum of severity

Food insecurity exists along an underlying **continuum of severity**, and the FIES is intended to reveal information across a range of food insecurity experiences.



What is **challenging** is that food insecurity is an unobservable, or **latent trait**, which means that it **cannot be measured directly**, as is possible with variables such as height and weight. We must instead learn about food insecurity by studying its **observable manifestations**.

Severity of respondent and question

There are **two key concepts** underlying the approach to FIES measurement:

① People answer the questions according to the severity of the food insecurity they **experience**.

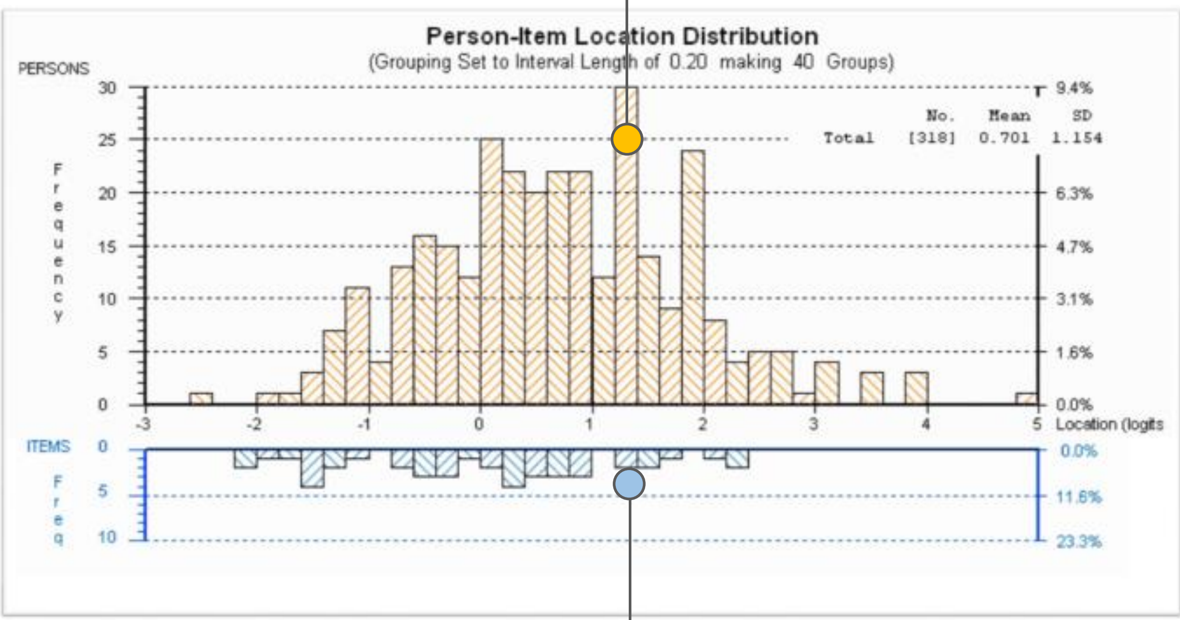
② The **questions** and the **respondents** (individuals or households) are located on the **same underlying continuum** of severity of food insecurity.

An example from educational testing

An example that may help you understand comes from **educational testing**.

Imagine a **hypothetical scale of proficiency** in an academic subject such as mathematics.

Students can be distributed along a continuum of mathematical proficiency, **from lowest to highest skill level**, depending on their answers to exam questions.



Exam questions can also be located on the same scale, as they **represent different levels of proficiency**, depending on their difficulty.

Relative severity of questions

A **fundamental feature** of the FIES is that the **order** of the **questions** along the scale **cannot be considered fixed** across countries. In different countries, or even sub-populations, the relative severity associated with each of the eight FIES questions may vary. There are two main reasons for this:

⇉ food insecurity conditions are **experienced or managed differently** in different cultures and livelihood systems

⇉ **nuances in translation** mean that the same question is interpreted in different ways in different contexts

The **position** of the **questions on the scale** of severity is **not imposed a priori**, but is determined after analysing the specific **data that were collected**.

Example

In some cultures where mild food insecurity is commonplace, **worrying** about having adequate food may not be the first or most common experience people have when facing food insecurity. In such a case, people may make dietary changes more quickly, and worry only when their food insecurity situation becomes more severe.

The **severity** of the FIES item referring to worry would therefore **be higher** in this context.

We will see later in this lesson how **the proportion of yes responses reveals the relative severity** of the questions.

Standard labels for the eight questions

Because the order of the eight questions is not constant, for the purpose of analysis, the FIES methodology uses **standard labels** for each of the questions rather than numbering them.

"Now I would like to ask you some questions about food. During the last 12 months, was there a time when…"

…you were **worried** you would not have enough food to eat because of a lack of money or other resources? ***WORRIED***

…you **ate less** than you thought you should because of a lack of money or other resources? ***ATELESS***

?

….you were unable to eat **healthy** and nutritious food because of a lack of money or other resources? ***HEALTHY***

…your household **ran out** of food because of a lack of money or other resources? ***RANOUT***

….you ate only a **few kinds** of foods because of a lack of money or other resources? ***FEWFOOD***

…you were **hungry** but did not eat because there was not enough money or other resources for food? ***HUNGRY***

…you had to **skip a meal** because there was not enough money or other resources to get food? ***SKIPPED***

…you went without eating for a **whole day** because of a lack of money or other resources? ***WHLDAY***

Important terminology

For the purpose of FIES analysis, the term **item** will refer to the **questions** in the survey module.

A **respondent** is an individual or household responding to the items.

A **case** is the unit for which data are collected and appear in a dataset.

A **parameter** is a **numerical quantity** that **characterizes** a given population or some **aspect of it**, and that can be **estimated** using observed **data**.

In the case of the FIES, **parameters** express the severity of food insecurity of both:

- the FIES questions (**item parameters**);
- the people who answer them (**respondent parameters**).

Preparing the data for analysis

To prepare the data collected through the FIES survey module for analysis, **each item should be coded**, so that: **0** is used for a "**no**" response; **1** is used for a "**yes**" response.

This is an example of FIES data along with the standard labels for the eight items.

	WORRIED	HEALTHY	FEWFOOD	SKIPPED	ATELESS	RUNOUT	HUNGRY	WHLDAY
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	1	1	1	1	0	0	0
4	1	0	1	0	1	0	0	0
5	0	1	1	0	0	0	1	0
6	0	0	1	1	1	0	0	0
7	1	1	1	1	1	1	1	1
8	1	0	0	0	1	0	1	0
9	1	1	1	1	1	1	1	0
10	1	0	0	0	0	0	0	0

The FIES survey module allows recording of "**don't know**" and "**refused**" responses to any of the FIES items. For analytic purposes, such answers are **treated as** "**missing**" and cases with missing responses for any of the FIES items are **excluded from the analysis**.

Example

The question "*During the last 12 months, was there a time you ate only a few kinds of foods because of a lack of money or other resources?*" (FEWFOOD) is always the third most severe item on the food insecurity scale. It is false. A **fundamental feature** of the FIES is that, the **order** of the **questions** and the **distance between them** along the scale are **not constant** across all contexts. In different countries, or even sub-populations, people's responses to the eight FIES questions may vary, which leads to **differences** in their order and relative severity.

FIES analysis methodology

Item Response Theory (IRT) is a methodology used to analyse responses to survey or test questions. The **Rasch model** is one of several models in IRT and is applied for the analysis of FIES data. Item response theory aims to improve the measurement accuracy and reliability of surveys and tests through analysis of response data. The item response theory (IRT) measurement model known as the Rasch model provides a theoretical base and a set of statistical tools to:

- assess the suitability of a set of survey items for scale construction
- create a scale from the items, and compare performance of a scale in various populations and survey contexts.

The Rasch model is widely used in health, education and psychology studies and provides the statistical basis for experience-based food security measurement.

For more details read the FAO publication "[Introduction to item response theory applied to food security measurements](www.fao.org/3/a-i3946e.pdf)" **www.fao.org/3/a-i3946e.pdf**

Assumptions of the Rasch model

The Rasch model is based upon **four key assumptions**:

Only one dimension is represented by the response data. For the **FIES**, this is the **access** dimension of **food security**.

An individual's responses to the eight FIES items are **correlated** with each other only because they are all conditioned by the **severity** of food insecurity of that **individual**.

The **greater the severity** of food insecurity experienced by a respondent, the **higher the likelihood** that he or she will **respond affirmatively** to each item.

All items are **equally strongly related** to the **latent trait** of food insecurity and **differ only in severity.**

A relative scale of severity

The analysis does not generate an **absolute scale** of severity. It **estimates** the **relative position** of the items and respondents along a scale **specific to the given dataset**.

This **relative position** of items and respondents on the scale of severity is **expressed by** their respective **parameters**.

Parameters cannot be compared with those of the same items on a scale produced from a different FIES dataset. Such a comparison requires an additional step called **equating**. This will be covered in Lesson 4 'Using FIES data to calculate food insecurity prevalence rates".

Expected response pattern

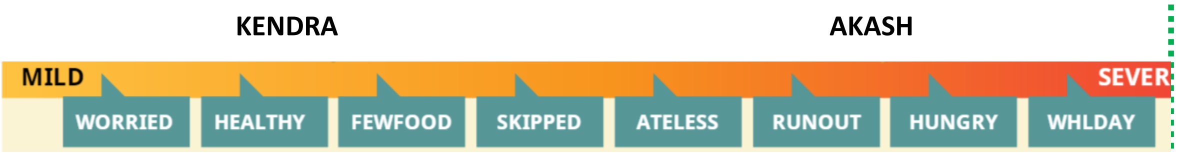
The logic behind the Rasch model is that the **likelihood of a respondent reporting an experience** depends on the distance along the scale between the severity of that respondent and that of the item associated with that experience. The following examples will help you visualize the scale to better understand this concept. The **more severe** a respondent's food insecurity is, **relative to** that of the **item**, the **more likely** they are to answer "yes" (give an **affirmative** response).

- a respondent who answers **yes** to a question can be expected to also answer **yes all less severe questions**.
- a respondent who answers **no** to a question is expected to also answer **no** to **all more severe questions**.

This means that once the **order of severity** of the eight **questions has been established**, specific **patterns of responses** by individual respondents can be considered more or less to "**fit**" the logic of the model.

Example

Which of the two individuals Kendra or Akash is more likely to answer yes to the item "HUNGRY"?



Akash is more likely to answer "yes" to the HUNGRY question since his position on the scale means that he is more severely food insecure than Kendra is and thus more likely to affirm the item HUNGRY, which is one of the most severe.

Introducing the analytical tool for analysis of FIES data

Now that you can **visualize a scale** with item and respondent parameters, you may wonder how they are calculated. This involves **parameter estimation**, which you will learn about next. First, let's look at the **analytical tools** which are available to help you carry out this calculation.

The analytical tools

Voices of the Hungry project provides the following **free analytical tool** to facilitate the **Rasch analysis of FIES data**, although other statistical software can also be used to carry out the analysis.

RM.weights - This package is aimed at food security measurement specialists and those familiar with the R open source statistical software. The "R Manual for the Implementation of Voices of the Hungry Methods to Estimate Food Insecurity" provides instructions for its use. It requires users to download the R open source software on their computer and use R programming language. For the RM.weights package, data can be of any format and columns can have any label (see section 21 of the manual). Download RM.weights here:
https://cran.r-project.org/web/packages/RM.weights/index.html

This tool will produce the parameter estimates and other outputs related to statistical validation. You will learn how to properly interpret these outputs in this lesson.

A National Statistics team scenario

Denis is the analyst from Grace's team who will use the FIES data to produce an estimate of the national prevalence of food insecurity for SDG monitoring.

*"Now I can understand how the items and respondents each have a place along the scale of severity. The **next step is to estimate their parameters** to determine their exact location." - Denis*

Estimating the item parameters

The **item parameter** is estimated based on the overall **pattern of responses** given by all respondents. A question representing a **less severe** experience will have a **smaller parameter** value, whereas a question representing a **more severe** experience will have a **larger parameter** value.
The **relative severity** of the items is determined based upon the understanding that the **more severe** an item is, the **less likely** respondents are to **report it**.

To return to the example of **educational testing**, this is similar to the expectation that the more difficult a question is, the fewer students will answer it correctly.

44% of respondents answered **YES** to the last severe item.
Only **5% of respondents** answered **YES** to the most severe item.
Another way to understand this concept is that the **proportion of affirmative responses** to a given item, in any sample, must be **inversely related** to the **severity** of the item.

Example of item parameter output

This is an example of output showing the item parameters and the errors associated with each of them.

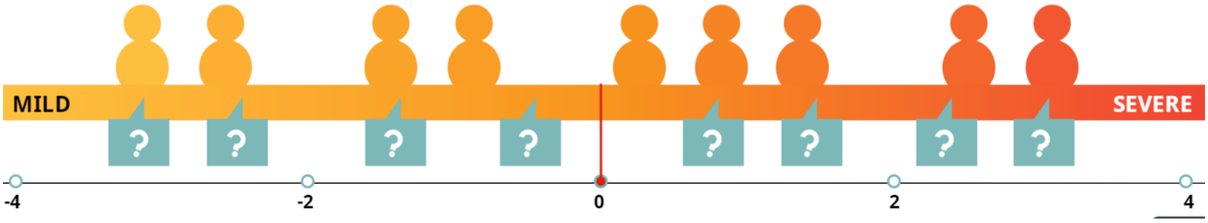
ITEM	Item severity	Standard error
WORRIED	-1.829	0.163
FEWFOOD	-1.176	0.153
HEALTHY	-0.576	0.151
ATELESS	-0.083	0.153
SKIPPED	0.101	0.155
RUNOUT	0.417	0.159
HUNGRY	0.615	0.162
WHLDAY	2.532	0.227

Note that they are in order from least to most severe, based on this particular dataset, and that this order does differ from the standard administrative order.

Characteristics of the scale

Four important characteristics to understand about the scale are:

- There is **no absolute** interpretation of the numerical values of the **parameters**, as they **only indicate the relative severity** of the items along the scale.
- The **scale is specific to a particular application** of the FIES in a sample of a population. Therefore, parameters cannot be compared with the values of the parameters on a scale coming from another dataset.
- The origin (zero) of the scale is arbitrarily set as the **mean of the item severity parameters** within each application of the Rasch model. The measures can be rescaled without altering the relative positions of items and respondents.
- The units of the scale are **logits**.



Respondent parameters

A respondent's **raw score**[1] is the basis for calculating the **respondent parameter**.

No matter how many respondents are included in a sample, there will always be up to **nine distinct values** of respondent parameters, one for **each possible raw score** (0-8).

An **essential point** to understand is that every respondent who answers "yes" to the **same number of questions** (**irrespective of which ones**) will be assigned the **same parameter**. The specific questions affirmed are not used to determine the parameter.

Expected patterns in good quality data

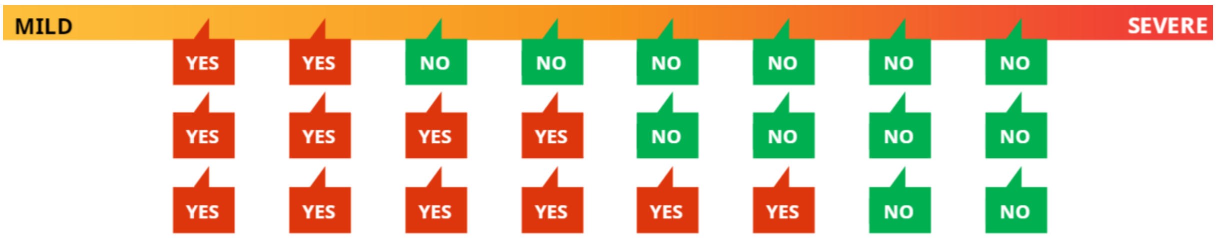
Although the raw scores and associated respondent parameters only depend upon the number of affirmative ("yes") responses, **certain response patterns** are **expected** in what is considered **good quality data**.

---

[1] The **raw score** is the **number** of affirmative responses given to the eight FIES questions - it is an **integer number** with a value between zero and 8.

Response patterns that **fit** with the theory behind the Rasch model are those that, when arranged in order of increasing severity, **start with** "**yes**" and are **followed by** "**no**" responses, **without alternating**.

We consider these to be "expected" patterns.

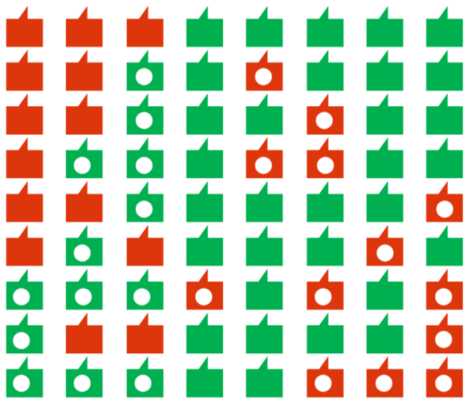


However, by **analysing** the frequency of **unexpected response patterns**, and examining which items cause them, we are able to test whether a question is functioning adequately.

Assigning a "best bet" parameter based on the raw score

Let's consider the **patterns which do not fit** the assumption in a little more detail.

Although they correspond to the **same raw score** (3) as the first pattern, there are **two or more** "**unexpected**"**answers** (the responses with white dots).



These are where the respondent either failed to affirm one of the first three items or where he or she affirmed one of the five items at the upper end of the severity scale. Each of these unexpected response patterns contains **contradictory pieces of information**:

the answer **yes** given to a more severe item suggests moving it to the **higher** end of the **severity scale**

the answer **no** given to a milder item suggests moving the respondent parameter to the **lower** end of the **severity scale**

The Rasch model recognizes that **both pieces of information are relevant**, and considers that the "**best bet**" is still to assign the respondent the **same parameter** based on the **number of affirmative answers**, in this case 3.

Standard errors of respondent parameters

The FIES methodology recognizes and takes into account the presence of possible measurement error: with only eight items we can never be sure that we have exactly measured the latent trait, so we need to **quantify** the extent of our **uncertainty**.

For **patterns closer to** the **theoretically expected** ones, we are **more certain of the precision** of the respondent parameter.

To each respondent parameter we associate a standard error, which is a **measure of the uncertainty that surrounds the severity level** associated with a given raw score.

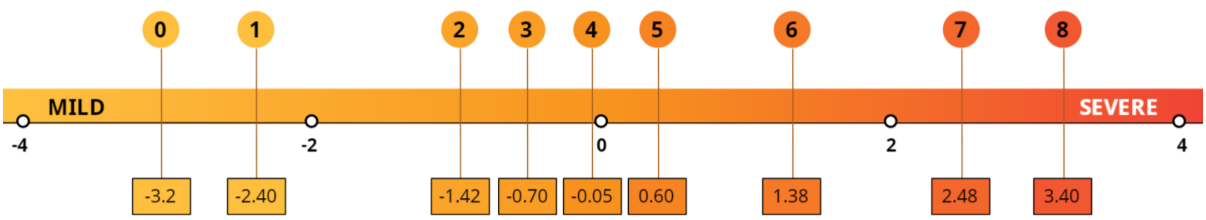
The **more cases** in a sample that have a pattern **consistent** with the assumptions of the model, the more confident we are that our **measurement is precise**, and this will be reflected in a **lower** value of the **estimated standard error** for the respondent parameter.

Difference between the raw score and the respondent parameter

Even if expressed by **numbers**, the raw score only provides an **ordinal measure** of food insecurity: the difference in severity between **adjacent raw scores** is **not constant**.

Raw score

Looking at raw scores only, you can see that a respondent with a raw score of 6 is more food **secure** than someone with score of 7 and more food **insecure** than someone with score of 5.



Respondent parameter

Using respondent parameters, however, you see that difference in severity is smaller between raw scores 5 and 6 than between 6 and 7.

Statistical validation allows us to see if the data are **consistent with assumptions** of the Rasch model. If they are, we can say that:

- the **raw score** is an **ordinal measure**[2] of the severity of food insecurity
- the **respondent parameter** is an **interval-level**[3] **measure** of severity

Respondent parameters allow us to **more precisely** evaluate the relative differences in food insecurity severity along the scale, between the respondents with each raw score.

Example of respondent parameter output

This is an example of output showing the **respondent parameters** associated with each raw score and the errors around each of them.

Raw score	Respondent severity	Standard Error
0	-3.235	1.507
1	-2.404	1.132
2	-1.420	0.895
3	-0.699	0.816
4	-0.053	0.799
5	0.604	0.832
6	1.375	0.938
7	2.479	1.203
8	3.404	1.507

Recall that these parameters indicate where a person with each raw score will be located along the scale of severity.

Example

Is it true that a respondent with a **raw score of 4** is **twice as food insecure** as a respondent with a **raw score of 2**.

It is not true. Raw score can only be used as an ordinal measure of food insecurity, meaning that we know that someone with a raw score of 4 is **more** food insecure than someone with a raw score of 2, but we do not know the exact difference in food insecurity severity between these two respondents. Only the respondent parameter can tell us this, as it is an interval measure of food insecurity.

[2] **Ordinal Scale:** In an **ordinal measure**, the order of values is meaningful, such as low, medium and high, but the exact distance between these values is unknown.

[3] **Interval Scale:** In an **interval measure**, the order of and distance between the values are known and meaningful.

Assessing data quality

You can estimate item and respondent parameters from any dataset, but this **does not guarantee** that the data are always of **acceptable quality** so that they can be used to meaningfully measure food insecurity. The **statistical validation** process is a **fundamental screening tool** for assessing data quality and for evaluating whether a given measure of food insecurity has a solid enough statistical basis for its intended policy and research uses.

Statistical validation should **always** be conducted when the food insecurity scale is applied for the first time in a given population, and **may be repeated** for numerous waves of a survey to ensure that the scale performs **consistently** well.

FIES analysis methodology

Statistical validation is an analysis that is used to check the quality of the data collected. This involves **applying** the Rasch Model to the FIES response data and **assessing** whether the data conform to the **model's assumptions**. If the data do conform to the assumptions, we can conclude that the **data can be used** to calculate a valid measure of food insecurity.

Assessing data quality

The next step is to learn **how to carry out statistical validation**, which will help you to:

Determine **how well** the **FIES works** in the population where it was administered

Identify **problematic items** where greater attention to translation and/or survey administration may be required

Identify **outliers**, or respondents with highly **unexpected** response patterns

Determine whether the FIES does **perform differently** when administered to different language groups or culturally distinct subpopulations

Interpreting Rasch output

The Rasch model allows us to produce **four results** that are **useful to test the quality of data** collected. They will help you to identify:

Infit	Outfit	Residual correlation matrix	Rasch reliability
Items that did not perform well in a particular population.	Similar to infit, but **sensitive to cases with unusual response patterns**, even among a few respondents	Items that may be slightly redundant, meaning that they **represent the same or closely related conditions** caused by food insecurity.	The proportion of total variance in the population that is accounted for by the measurement model. Provides information on the **discriminatory power** of the scale overall.

Let's consider these in more detail, learn **how to interpret them** and see what actions we can take if they fall outside acceptable ranges.

What is discriminatory power?

The scale's discriminatory power is its **ability to differentiate** among respondents with different levels of severity of food insecurity.

It is worthwhile considering the **practical consequences** of higher or lower discriminatory power.

Where can information on discriminatory power be found among the results of the Rasch analysis?

To assess the discriminatory power of **one** particular **item**, infit is the relevant statistic to consider, whereas **reliability** (see discussion in a later section) is a proxy for discriminatory power of the **overall scale**.

For a **given item**, it may be lower if respondents do not clearly understand the meaning of the item, if the item is understood differently by different types of respondents, or if the item is not contributing as much as the other items towards measuring food insecurity.

The discriminatory power of the **scale overall** may be lower if respondents are not attentive or not taking the survey seriously, if interviewers do not properly record responses or if the overall scale was poorly translated.

Fit statistics

Infit statistics are particularly useful to **identify items that did not perform well** in a given population. This may be due to problems with the item's translation, meaning that the question was not understood well by respondents, or to other problems during data collection.

LOW (Below 0.7)

A **single item** with a particularly **low infit** may be somewhat redundant with other items.
This usually **does not substantially worsen** the measure of the latent trait, but means that the item does not add important information to the measure. Particular care should be taken in the formulation of these items in **future surveys** to more clearly differentiate them.

0.7 – 1.3

An **adequate fit** to the Rasch model is indicated by **infit and outfit statistics of 0.7 – 1.3** for each item. This is obtained if all items are associated to the latent trait and discriminate equally well among respondents.

HIGH (Above 1.3)

These are items with unexpected response patterns. An item with **infit larger than 1.3** is considered to be **performing poorly**, and should be examined to decide whether to drop it from the scale for the current analysis. This also suggests that work is needed to improve this item for future surveys.

As an **additional investigation**, you may apply the Rasch model and estimate parameters separately for different sub-groups of interest. You can then **examine infits separately** for **different languages or distinct cultures** where the item may be understood differently. If an infit is high only for some sub-groups, then attention should be given to improving translation that affects those groups in future surveys, but no action needs to be taken in the current analysis of the FIES data.

Dropping items from the scale due to missing responses

Before learning more about the fit statistics, let's consider an **important decision** you may face during statistical validation **if you encounter** either a high percentage of missing responses or a high infit **affecting only one or two questions**. **These items may be dropped from the scale** and an analysis attempted on the remaining 6- or 7-item scale. Following data collection, an important initial step is to carry out a **descriptive analysis of the cases with missing responses** to any of the FIES items.

Did the missing responses come primarily from one particular item?

If an item has more than 10% missing responses, this may suggest that it was difficult for respondents to answer and the decision may be made to drop it from the scale. Doing so will affect the criteria for what is considered a "complete" case, as the scale will now have only 7 items and all respondents who answered those seven items will be retained in the analysis.

What were the characteristics of the cases with missing responses?

If cases with missing responses are concentrated in one sub-population (language, culture, or gender), attention should be given to translation, cultural adaptation, or gender matching of respondent and interviewer for future surveys.

Dropping items from the scale due to high infit

The second reason to consider dropping an item from the scale would be a high infit. If an item's infit is high, the decision may be made to drop the problematic item from the scale, but this should only be done after careful consideration of:

➲ The **standard error** around the infit statistic. A wide standard error indicates a less trustworthy result, and therefore provides weaker evidence for dropping the item.

➲ The **number of affirmative responses** to the problematic item. For example, in a very food secure country, if you find a high infit for one of the most severe items in the scale, it has probably been answered "yes" by only a small number of respondents. In this case the high infit can essentially be ignored and the item retained in the scale, as this will not significantly alter the estimates of food insecurity prevalence rates.

Although a scale with fewer than eight items may be used for the analysis, a **minimum of six items must be retained** to produce an acceptable measure of food insecurity severity.

Drop an item temporarily or permanently?

If the decision is made to drop an item from the current analysis, it can still be retained for future surveys. Once work has been done to improve translation or correct problems that occurred in the field, the item may work well in future rounds of data collection. Eliminating it permanently from the survey module should only be considered after repeated surveys demonstrate poor performance of that item.

Fit statistics - Outfit

An outfit of >2 is considered "high"

While the principle statistic to focus on is infit, interpretation of the outfit statistic is also worthwhile. Outfit statistics have a similar interpretation to the infit statistics, but are sensitive to the presence of even just a few cases with highly unexpected response patterns, and are therefore useful to flag the **presence of outliers**.

The interpretation of a high outfit entails investigation to determine whether the high outfit results from:

- A **few highly unusual respondent** response patterns - When less than 0.25% of cases, e.g. 1 case out of a sample of 400, it can be **ignored** or the decision may be made to **drop the unusual cases and refit the model**, checking that the item outfit for the remaining cases is acceptable.
- A more **pervasive set** of somewhat unusual patterns - If there is a larger number of moderately to highly discrepant responses to the item, the decision may be made to **drop the unusual cases and refit the model**, checking that the item outfit for the remaining cases is acceptable.

In both situations, the discrepant cases that have been dropped to refit the model should still be **retained** when calculating **prevalence rates**.

Interpreting Residual correlation matrix

A residual correlation between a **pair of items** is **considered high if it is >|0.4|**

Do you recall the Rasch model assumption that all correlations among items should result from their common association with the latent trait?

This assumption is assessed by comparing residual correlations among items, after having eliminated the correlation that exists because the items contribute to measure the same latent trait. One **possible cause of high residual correlation** between one pair of items is when two **items** basically **overlap in meaning**, which may be a reflection of inaccurate translation or wording. Each item is meant to capture a different aspect of food insecurity, so having **redundant questions** weakens the ability to measure food insecurity accurately.

Fit statistics - Rasch reliability

Rasch reliability provides information about the **discriminatory power of the overall scale**, measuring the proportion of variability in the data that is explained by the Rasch model.

What level of reliability is considered acceptable?

For an 8-item FIES scale, a Rasch reliability value **above 0.7** is considered acceptable.

For a 7-item scale, a Rasch reliability value **above 0.6** is considered acceptable.

The response to each **single item** as a measure of the latent trait has a **substantial measurement error**. But taken **together**, every item:

- **contributes information** to measuring food insecurity along the severity continuum;
- increases **precision**;
- **reduces** the overall impact of measurement **error**;
- **filters** out **any subjective** component that a single item may have.

Improving translation following statistical validation

You may have noticed that suggested solutions to most problems point to improving translation and/or cultural adaptation of the FIES questions.

If this is the case, you should revisit the process of cultural and linguistic adaptation **described in lesson 2 'Including the FIES in a survey' of this course**.

Once the translation has been modified and new data have been collected, statistical validation should be conducted once again to see if scale performance has improved as a result.

Cases used for statistical validation

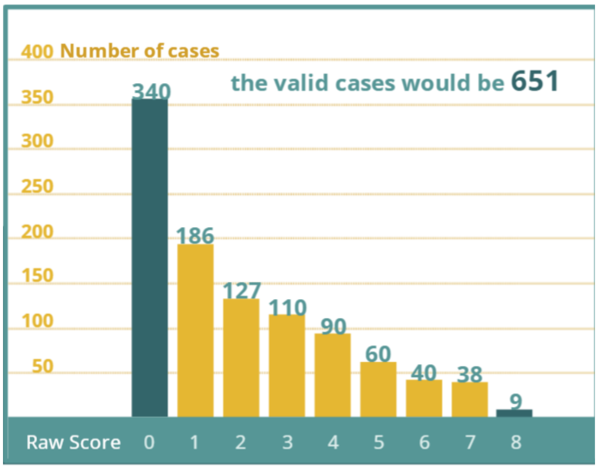
The FIES analytical tools perform statistical validation using only cases with **complete responses** (where all items have been responded with either "yes" or "no") and those with non-extreme raw scores (i.e. with raw scores other than 0 or the maximum number of scale items). This sub-sample of complete, non-extreme cases must be sufficiently large for statistical validation to be reliable.

This requirement is especially important to consider in populations where the **number of non-extreme cases** are **reduced** due to:

- **very low prevalence** of food insecurity (a high proportion of raw score 0);
- **very high prevalence** of severe food insecurity (a high proportion of raw score 8).

Example

In this example, let's assume that after eliminating the missing cases, you have a **sample of 1000** with complete FIES responses. Since it is a very food secure country, however, 340 of them have raw score 0 and cannot be used for statistical validation.



In the next section, you will learn about the number of complete, non-extreme cases needed to have confidence in the statistical validation of FIES data.

A sufficient number of complete, non-extreme cases is necessary in order to have confidence in the statistical validation of FIES data.

Less than 100	**100 - 299**	**300 - 999**	**1000 or more**
Insufficient number of non-extreme cases. Do not rely on the data for validation of the measure.	The number of non-extreme cases is too small to describe it as more than a **provisional validation**. It may be worth carrying out statistical validation to spot major violations of infit or expected item severity. Residual correlation analysis is unreliable with fewer than 300 non-extreme cases.	Item severity parameters will be quite reliable, but infit statistics may have large standard errors.	You can be confident in the results of the statistical validation with **1000 or more non-extreme cases**.

Remember that while cases with extreme raw scores cannot be used for statistical validation, they **must be included** in the **calculation of food insecurity prevalence estimates**!

Case studies

Marc lives in a country with very low rates of food insecurity. As the FIES produces reliable prevalence estimates even in countries with very low rates of food insecurity, Marc has investigated using the FIES for national monitoring purposes in his country. A pilot survey was conducted in only one province with a sample size of 5000. Marc has carried out the Rasch analysis on the pilot survey data, and is looking at the results to assess the performance of the FIES in this particular application. After eliminating the incomplete cases and those with extreme raw scores (0 or 8), there were 4,022 cases remaining. He is satisfied that this is sufficient to carry out statistical validation. The output looks like this:

	Item severity	Standard error of item severity	Infit	Standard error of infit	Outfit
Worried	-1.59	0.13	1.17	0.077	1.46
Healthy	-0.75	0.11	0.85	0.06	1.03
Fewfood	-0.36	0.11	1.07	0.05	1.28
Skipped	0.87	0.11	-0.24	0.05	0.74
Ateless	-0.43	0.11	0.91	0.05	0.80
Runout	0.24	0.11	0.86	0.05	0.72
Hungry	0.87	0.11	0.99	0.05	0.98
WHLday	2.27	0.13	**1.39**	**0.07**	1.63

Marc notices that the **infit for WHLDAY**, the most severe item, is **above 1.3** and so is outside the acceptable range. He first examines the **standard error of the infit** to understand if this result is an important problem. By subtracting the standard error from the infit (1.399-0.0721 = 1.3269), he finds that it is still outside the acceptable range. As this strongly suggests that the item did not perform well, he wonders whether it would be preferable to drop it from the scale for the analysis.

Marc **re-runs the analysis without this item**. The Rasch reliability goes down, but Marc understands that this is to be expected, as the reliability is related to the number of items in the scale. He then looks at the number of "yes" responses to each item, and finds that the high infit item was only affirmed by 48 respondents. This is understandable as it is the most severe item in the scale and the survey was conducted in a food secure context. He recognizes that since there are so few "yes" responses to this poor-performing item, retaining it in the scale should not pose major problems. For the national survey, **he decides to maintain the 8-item scale** to measure food insecurity in his country.

Case study: Carlos

Carlos has received data collected from a survey of 1200 households and will now carry out statistical validation. As a first step, he examines the data to ensure that only complete cases (those who answered either yes or no to each of the eight items) are included in the validation. He is surprised to find a very high proportion of missing responses: 280 respondents failed to answer all of the FIES questions (23%)!

He looks at a descriptive analysis of the missing responses and finds that they primarily come from one item, FEWFOOD. He plans to work on improving the translation of this item for future surveys. Because of this, he immediately drops this item, making it a 7-item scale.

This changes the criteria for what is a "complete" response to those cases who have answered the seven remaining items. He re-examines the descriptive statistics and sees that there are now only 60 incomplete cases (5%). He then eliminates the 100 cases with extreme raw scores and is left with 1040 valid cases (based on the 7-item scale).

Carlos proceeds with his analysis and calculation of food insecurity prevalence estimates, but **without the "FEWFOOD" item**. He **makes future plans** to work on improving the translation of the "FEWFOOD" item and to collect new FIES data in a future survey with the improved translation of that item included.

A National Statistics team scenario

Denis is the analyst from Grace's team who will use the FIES data to produce an estimate of the national prevalence of food insecurity for SDG monitoring.

"In the next lesson I will be tackling my next challenge - calculating the food security prevalence estimates. Now that we have gone through the process of statistical validation of our data, we can have more confidence that the FIES from this survey will give us reliable information about the prevalence of food insecurity in our country. "

Summary

In this lesson you have learnt about the basic concepts of the food insecurity scale, and validation of FIES data using the Rasch model.

You have learnt:

- the concepts underlying the FIES analysis, including the concept that people answer according to the severity level they experience, and that these respondents are on the same underlying continuum of severity as the question (items);
- the Rasch model of the Item Response Theory method and how it generates a relative scale of severity;
- the use of affirmative answers (raw score) as a base for assigning a parameter to a respondent;
- the analysis of the Rasch model-based outputs (infit, outfit, residual correlation and reliability) as a means of statistical validation to check the quality of the data collected.

In the next lesson, you will learn how to calculate a measure of severity of the food insecurity condition experienced by respondents and the prevalence of food insecurity at different levels of severity in a population.

Additional reading

Cafiero, C., Nord, M., Viviani, S., Del Grossi, M.E., Ballard, T., Kepple, A., Miller, M. and Nwosu, C. (2016). **Methods for estimating comparable prevalence rates of food insecurity experienced by adults throughout the world (www.fao.org/3/c-i4830e.pdf)**. Voices of the Hungry Technical Report Number 1.FAO, Rome, Italy.

Nord, M. (2014). **Introduction to item response theory applied to food security measurement** (**www.fao.org/3/a-i3946e.pdf**). Basic Concepts, Parameters and Statistics.

FAO, Rome, Italy.

Cafiero, C., Viviani, S. and Nord, M. (2017). **Food security measurement in a global context: The food insecurity experience scale**

https://www.sciencedirect.com/science/article/pii/S0263224117307005. FAO, Rome, Italy.