Test development

Nils Myszkowski, PhD

Outline

Defining objectives

Some statistics involved

The test construction process

Review

Defining objectives

Some statistics involved

The test construction process

Review

Defining objectives
Some statistics involved
The test construction process
Review

Aims of this session

- At the end of this session, you should be able to:
  - Know the main qualities a psychometric test should have
  - Recognize and use the statistical methods that are involved in assessing psychometric qualities
  - Know the process of test construction
  - Understand the importance of choosing a response scale and using it properly

Outline

Defining objectives
The basic qualities
Some statistics involved
The test construction process
Review

3

Defining objectives

Some statistics involved

The test construction process

Review

Defining objectives

Defining objectives
The basic qualities
Some statistics involved
The test construction process
Review

Measurement

- "*Measurement is the numerical estimation and expression of the magnitude of one quantity* ***relative to another***." (Michell, 1997)
- Measuring is always **comparing**.
  - A weight of 80 kg is 80 times the reference weight of 1 dm^3 of pure water.
  - Performance in an exam is based on the meaning of three scores: 0% (null performance), 100% (perfect performance), and 50% (minimal required performance).

Defining objectives
The basic qualities
Some statistics involved
The test construction process
Review

Scientific measurement

- "Measurement involves the use of certain **devices** or **rules** for assigning numbers to objects or events." (Stevens, 1946)
- Again, constructs may not be "numbers"...
- Because the aim is to make the numbers reflect reality as objectively as possible, these devices and rules must be systematic.
- The devices and rules are in our case test administration procedures, and scoring systems.
- The process of making test administration and scoring procedures systematic and stable is called **standardization**.
- Standardization is essential in testing, and we will cover this topic more extensively in the course on test construction.

Defining objectives

Defining objectives 6
The basic qualities
Some statistics involved
The test construction process
Review

A "good" measure?

- A good measure is an **accurate** (or "robust") measure.
- It's an abuse of language, but robust measures are also called "valid" measures.

Defining objectives

Defining objectives 7
The basic qualities
Some statistics involved
The test construction process
Review

How to ensure a robust measure?

- Ensuring a good measure will involve using best practices in:
  - Test construction
  - Test robustness verification
- We will cover these in the future courses.

Defining objectives
The basic qualities
8
Some statistics involved
The test construction process
Review

Defining objectives

The basic qualities

Defining objectives
The basic qualities
9
Some statistics involved
The test construction process
Review

Operationalism

- "Measurement involves the use of certain **devices** or **rules** for assigning numbers to objects or events." (Stevens, 1946)
- Back to this definition: This definition is a bit problematic, in that it implies that measurement is only a procedure: It is (properly) achieved by (just) doing it.
- Opearationalism is the view that all theoretical terms in science must be defined only by their procedures or operations.
- "We built an extroversion measure, so we measured extroversion."

Defining objectives
The basic qualities
10
Some statistics involved
The test construction process
Review

Operationalism

- Borboom (2006) describes operationalism in measurement as a major issue in Psychology:
- "The dominant idea is that one has to find an “operationalization” (read: observed score) for a construct, after which one carries out all statistical analyses under the false pretense that this observed score is actually identical to the attribute itself."
- It is important to realize that intepretations that are made from psychological measurement assume that the measurement device has accurately estimated the construct.
- In other words, not because you just built an extroversion measure, does it mean that it has any accuracy. It is an assumption that must be challenged...
- ...but how?

Defining objectives
The basic qualities 11
Some statistics involved
The test construction process
Review

Psychometrical qualities

- Ensuring and investigating the accuracy of a measure requires investing different properties.
- Such properties are **metrological qualities** or **psychometrical qualities**.

Defining objectives
The basic qualities
Some statistics involved
The test construction process
Review

Metrological qualities

- The main metrological qualities of a measure are:
  - **Discrimination power**
    - Ability to reveal differences between different statistical units.
    - ≈ Variability between units
    - *Ex : If you weight 100 different individuals, the scale indicates different weights*

Defining objectives
The basic qualities
Some statistics involved
The test construction process
Review

Metrological qualities

- The main metrological qualities of a measure are:
  - **Discrimination power**
    - Ability to reveal differences between different statistical units.
    - ≈ Variability between units
    - *Ex : If you weight 100 different individuals, the scale indicates different weights*
  - **Reliability**
    - Ability to **not** reveal differences within one statistical unit.
    - ≈ Consistency within units
    - *Ex : If individuals weight themselves twice, the scale indicates the same weight*

Defining objectives
The basic qualities
Some statistics involved
The test construction process
Review

Metrological qualities

- The main metrological qualities of a measure are:
  - **Discrimination power**
    - Ability to reveal differences between different statistical units.
    - ≈ Variability between units
    - *Ex : If you weight 100 different individuals, the scale indicates different weights*
  - **Reliability**
    - Ability to **not** reveal differences within one statistical unit.
    - ≈ Consistency within units
    - *Ex : If individuals weight themselves twice, the scale indicates the same weight*
  - **Validity**
    - Ability to measure what it intends to measure.
    - ≈ Adequacy with "the real world"
    - *Ex: The scale attributes individuals their real weight.*

Defining objectives
The basic qualities

Defining objectives
The basic qualities
13
Some statistics involved
The test construction process
Review

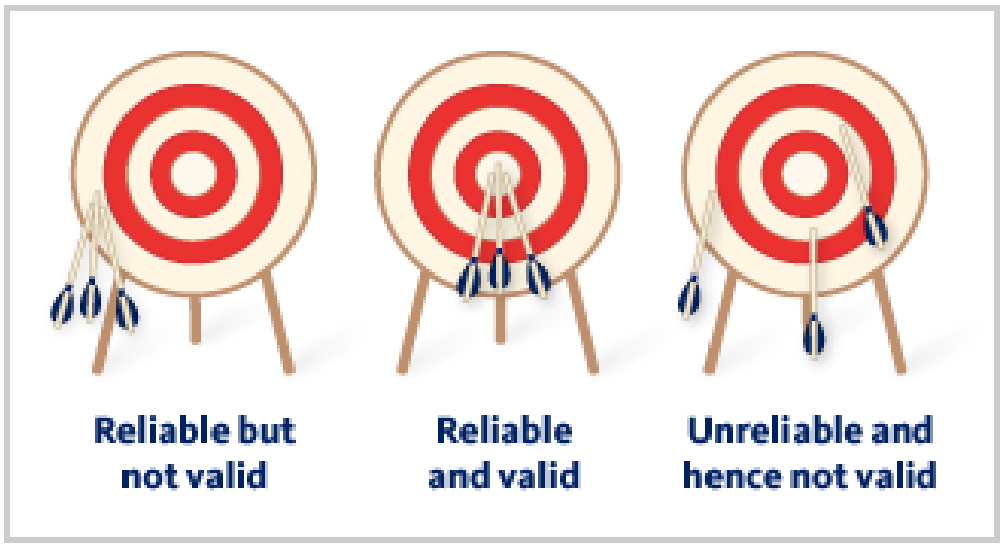


Figure: Reliability and validity are often represented using the shots of a perfect archer on a target. The reliability and the validity of using this target as a measure of archers' performance are evaluated.

Outline

Defining objectives
Some statistics involved 14
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Defining objectives

Some statistics involved

The test construction process

Review

Some statistics involved

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Some statistics involved

- Units and variables
- Sampling
- Variable Types
- Variable Roles
- Univariate statistics
- Bivariate statistics
- Multivariate statistics
- A note on inference

Defining objectives
Some statistics involved
Units and variables 16
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Units

- When data is collected in a quantitative research, it is collected from someone or something. The people or things that we collect data from are referred to as statistical **units**, also called **cases**.

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Variables

- The data we are collecting from these units concern information on specific **variables**. The label "variable", refers to the fact that this gathered information may differ between units.
- For example, we collect a course's grades (the variable) among students (units).
- *Note : Practically speaking, the data is collected, as it is input in the database. But it does not mean that all variables are necessarily observed, as some can be manipulated.*

Defining objectives
Some statistics involved
Units and variables 18
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Observations

- The information that we gather on a unit for a variable is called an **observation**.
- For example, if you record favorite color in a sample of individuals, "Green" is the observation of the variable "Favorite Color" for person A (unit).

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Modalities

- By definition, different observations can be made for one variable.
- The set of observation that *can* be made for a variable is called the set of **observable modalities**.
  - For example, A, A-, B+, B, B-,...etc. are the observable modalities of the variable "Grade in this class".

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Modalities

- By definition, different observations can be made for one variable.
- The set of observation that *can* be made for a variable is called the set of **observable modalities**.
  - For example, A, A-, B+, B, B-,...etc. are the observable modalities of the variable "Grade in this class".
- However, it happens that the possible observations are not all observed. The set of observations that are observed for a variable in a dataset is called the set of **observed modalities**.
  - If I am a very nice professor, the set of observed modalities for the variable "Grade in this class" will only include A and A-.

Some statistics involved

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Some statistics involved

- Units and variables
- Sampling
- Variable Types
- Variable Roles
- Univariate statistics
- Bivariate statistics
- Multivariate statistics
- A note on inference

Defining objectives
Some statistics involved
Units and variables
Sampling
21
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Samples

- In many situations, it can be complicated to gather observations for an entire population of units.
  - For example, all the population of "NYC residents" or "students".
- Therefore, researchers use **sampling**. They only gather a sample of units, drawn from the population.

Some statistics involved
Sampling

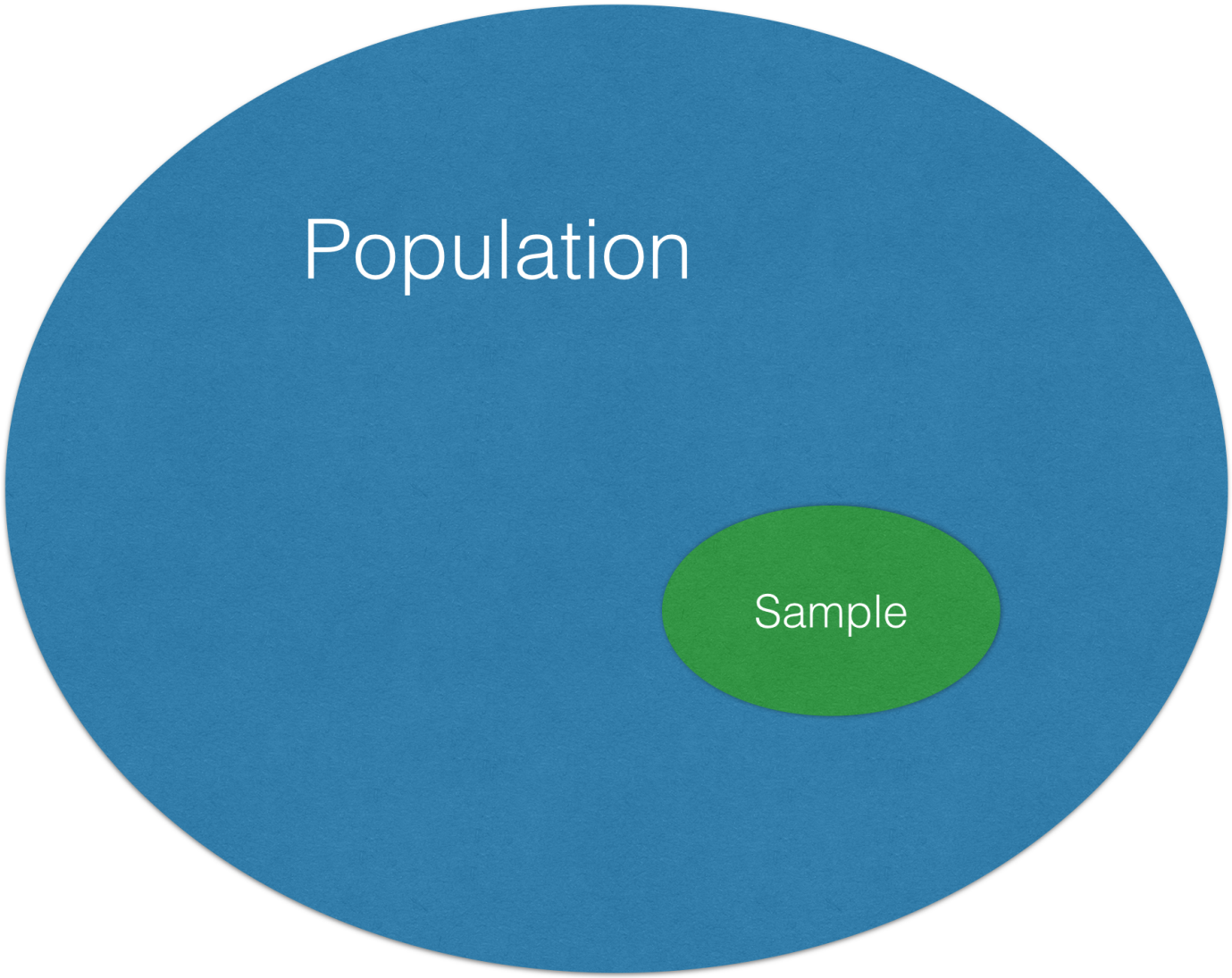
Defining objectives

Some statistics involved

- Units and variables
- Sampling
- Variable Types
- Variable Roles
- Univariate statistics
- Bivariate statistics
- Multivariate statistics
- A note on inference

The test construction process

Review



Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Sample size

- The number of units in a sample is referred to as the **sample size**.
- In articles, it is often written under the letter N.
- $N = 34$ means that the sample is composed of 34 units.
- Sample size will be involved in many computations.

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Samples

- As the point of psychometric research is still to make conclusions on the quality of a test in the population, researchers need to gather samples that are meant to represent the population.
- But samples can be more or less **representative** of the population.

Some statistics involved

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Some statistics involved

Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Variable types

- Variables can have different types. Their types will determine the statistical analyses that can be conducted.
- Variable types are also often referred to as **levels of measurement**, or **scale of measurement**.
- Once again, this simple notion is actually very important, as specific levels of measurement are a strong assumption to many techniques.

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types 27
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

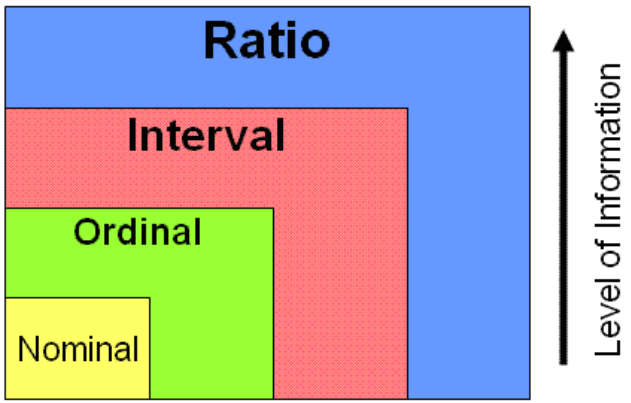
Variable types

- **Nominal** or **categorical** variables are variables that have two or more categories, but which do not have an intrinsic order. When they have only two categories, they are said to be dichotomous. (e.g. : "Psychology Student" vs. "Other Student")
- **Ordinal** variables are variables that have two or more categories that can be ordered. (e.g. : Grade Letters)
- **Interval** variables is an ordinal variables with the additional property that the magnitudes of the differences between values are meaningful. (e.g. : Temperature in Fahrenheit degrees)
- **Ratio** variables are interval variables whose any two values have meaningful ratio, making the operations of multiplication and division meaningful. Such is the consequence of the presence of an absolute zero. (e.g. : Height).

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types 28
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Variable types

- "When we move from nominal to ratio scales, we go from numbers that carry less information to numbers that carry more. As a consequence of this, going from one level of measurement to another requires us to be aware of whether the information that the numbers entail is preserved through whatever transformations or manipulations we apply to them." (Urbina, 2004)



Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Variable types

- Note that, for specific statistical techniques to be used, sometimes variable types have to be assumed.
- For example, to compute scores through additions, and make them meaningful, we need (at least) interval variables.
  - E.g. : The equality of scores of $3 + 3 = 6$ and $2 + 4 = 6$ are only permitted if the magnitude between 1, 2, 3 and 4 is (assumed to be) the same.
- Therefore, when computing scores out of Likert scales, we make the assumption that the outcome of a Likert scale is an interval variable.

Mobile phone use in on-campus computer labs should be prohibited.				
1 strongly disagree	2 mildly disagree	3 neutral	4 mildly agree	5 strongly agree

Some statistics involved

Variable Types

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types 30
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

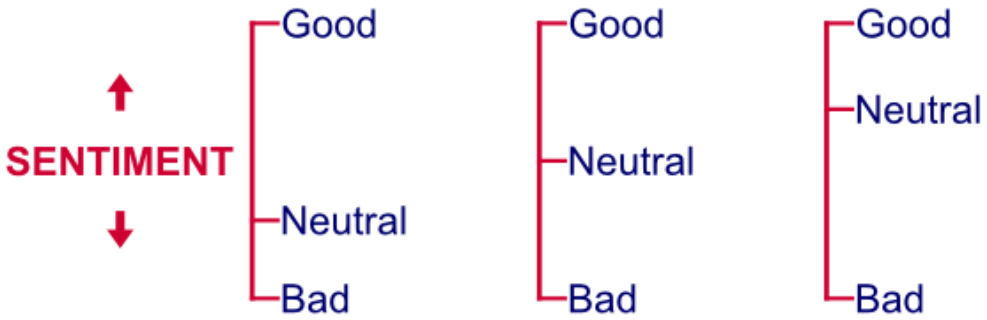


Figure: Likert scales are often used in psychology testing, and assumed to produce interval variables as outcomes. Above is a graphical representation to explain why this is just an assumption : Is the magnitude between neutral and good the same as the one between neutral and bad?

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types 31
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Psychometrics and variables types

- Psychometrics is meant to quantify phenomena. Therefore, testing mainly involve item scores, and test scores that are based on the addition of items scores.
- As a consequence, in psychometrics, we will mostly use **interval variables** (or at least assumed to be).
- However, in the case where tests yield a categorical decision (e.g. Disorder vs. No Disorder), we are also interested in the psychometric qualities of such a decision, rather than on the score itself.

Some statistics involved

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Some statistics involved

Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Independent and dependent

- Variables can have different roles. Their role will determine the statistical analyses that are conducted.
- An **independent** variable (sometimes called a predictor or predicting variable, is a variable that is being manipulated or observed in order to observe the effect on a **dependent** variable, also called an outcome variable.
- Simply put, dependent variables are (supposed to be, or not to be) dependent on independent variables.

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Variable roles

- The statistical link between a set of dependent variables and a set of independent variables is called a statistical **effect**.
- Be careful, links between dependent variables and independent variables are statistical effects, but not necessarily causal links.
- However, statistical effects are often used as an argument in favor of the demonstration of a causal link (usually along with the theoretical framework).

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Psychometrics and variable roles

- In psychometric examinations, researchers mostly measure variables, and verify that the statistical links between them are in line with theoretical expectations.
  - For example, psychometricians might want to verify that the results on an mental ability test are related to school success. To do so, they measure it with their test and measure school success. In this case, mental ability is theoretically considered as a cause (or at least an antecedent) of school success, therefore mental ability is here considered as the independent variable, while school success is here the dependent variable.

Some statistics involved

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Some statistics involved

Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Univariate statistics

- Univariate statistics are the simplest form of statistical analysis.
- Univariate statistics focus on a **single** variable.
- Univariate statistics contrast with bivariate statistics, which focus on two variables simultaneously.

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Psychometrics and univariate statistics

- In psychometrical analyses, univariate statistics are mostly used to verify the discrimination power (i.e., variability) and difficulty of the test or the items.
- Notably, psychometricians will focus on:
  - Centrality indices (mean, median, mode, success rate)
  - Dispersion indices (variance, standard deviation)
  - Distribution histograms

Some statistics involved

Univariate statistics

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Table 1. Descriptive statistics and scale score reliability.

Scale/subscale	Mean	SD	Median	Minimum	Maximum	Cronbach's *α*
IMRSS – Total	32.0	6.9	33	9	45	.86
IMRSS – Work Overload	16.4	4.8	17	4	25	.86
IMRSS – Medical Accountability	15.6	3.1	16	4	20	.78
SWLS – Satisfaction With Life	25.0	6.4	26	8	35	.89
HADS – Anxiety	7.8	3.8	7	1	20	.78
HADS – Depression	4.0	3.1	3	0	18	.77
WCC – Emotion-focused coping	21.5	6.2	22	9	36	.87
WCC – Problem-focused coping	27	5.1	27	10	40	.82
WCC – Seeking social support	21.6	4.5	22	8	32	.73
MBI – Emotional exhaustion	24.6	11.7	24	0	53	.91
MBI – Depersonalization	12.5	6.5	12	0	30	.76
MBI – Inefficacy	35.8	7.1	37	8	48	.83

Figure: An example use of univariate statistics in Psychometrics (note: Cronbach's α is actually multivariate)

Some statistics involved

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Some statistics involved

Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Bivariate statistics

- Bivariate statistics focus on the relationship between two variables at the same time.
- While univariate statistics focus on variation, bivariate statistics focus on covariation.
- In psychometrics, bivariate statistics are used to test hypotheses of association between variables.

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Psychometrics and bivariate statistics

- As psychometrics mostly involve scores (and therefore numerical variables), the main form of bivariate statistical analysis used in psychometrics is bivariate linear correlational analysis.
- Bivariate linear correlational analysis is based on descriptive and inferential statistics:
  - Pearson correlation coefficients (r)
  - Correlation diagrams
  - Inference on r in the population (confidence intervals, t-test)

Some statistics involved

Bivariate statistics

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Table 3. Correlation coefficients between the various measures.

	IMRSS	IMRSS-WO	IMRSS-MA
IMRSS – Total	/	/	/
IMRSS – Work Overload	.92***	/	/
IMRSS – Medical Accountability	.80***	.50***	/
SWLS – Satisfaction With Life	−.36***	−.35***	−.25***
HADS – Anxiety	.57***	.52***	.46***
HADS – Depression	.46***	.46***	.30***
WCC – Emotion-focused coping	.53***	.44***	.49***
WCC – Problem-focused coping	.09	.09	.07
WCC – Social support	.27***	.19**	.30***
MBI – Emotional exhaustion	.64***	.63***	.45***
MBI – Depersonalization	.35***	.31***	.30***
MBI – Inefficacy	−.11	−.14*	−.03

***$p < .001$; **$p < .01$; *$p < .05$.

Figure: An example use of bivariate statistics in Psychometrics

Some statistics involved

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Some statistics involved

- Units and variables
- Sampling
- Variable Types
- Variable Roles
- Univariate statistics
- Bivariate statistics
- Multivariate statistics
- A note on inference

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Multivariate statistics

- Multivariate statistics focus on the relationship between more than two variables at the same time.
- Like bivariate statistics, in psychometrics, multivariate statistics are used to test hypotheses of association between variables.

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Psychometrics and multivariate statistics

- In psychometrics, multivariate statistics are notably used to test hypotheses on the **factor structure** (dimensionality) of tests. Such procedures are called Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA).
- Hypotheses on the factor structure of tests have for a long time been tested through Exploratory Factory Analyses (EFA), until the recent popularization of Confirmatory Factor Analysis (CFA) and Item Response Theory (IRT) modeling.
- Variants of these analyses exist for constructs that are not assumed continuous, and these concerned models are generally called Latent Class models.

Some statistics involved
Multivariate statistics

Defining objectives

Some statistics involved

- Units and variables
- Sampling
- Variable Types
- Variable Roles
- Univariate statistics
- Bivariate statistics
- Multivariate statistics
- A note on inference

The test construction process

Review

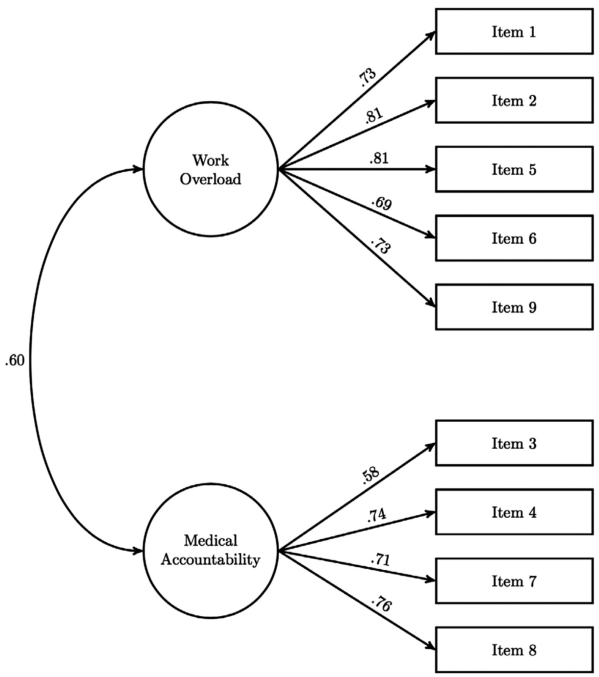


Figure 1. Path diagram of the 2 correlated factors model with standardized coefficient estimates.

Figure: An example representation of a Confirmatory Factor Analysis (Myszkowski et al., 2017)

Some statistics involved

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
48
The test construction process
Review

Some statistics involved

Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
49
The test construction process
Review

Inferential statistics

- Inferential statistics involve **making conclusions on the population based on the results observed in the sample**.
- Inferential statistics are meant to estimate how well a researcher knows the value of something in the population.
- What is at stake is the **generalizibility** on the population of the results observed in the sample.

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference
The test construction process
Review

Psychometrics and inference

- In psychometrics, most of the time we want to know about the **strength** and the **direction** of the relationships between measures. Binary outcomes about the presence or absence of a relationship is usually not enough.
- Indeed, you will see through the next courses that psychometrical qualities are usually based on effect sizes (notably correlation coefficients) themselves.

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference 51
The test construction process
Review

Psychometrics and inference

- In other words, we're more interested by effect sizes themselves, than concluding that effects are they are non-zero.
  - For example, knowing that the correlation between two mental ability tests is significantly different from zero is relevant, but insufficient. Here what we really care for is the value of the correlation coefficient, so we will be more interested in getting confidence intervals of the value of the correlation coefficient in the population than in concluding that this value is different from zero.
  - Likewise, knowing that the correlation between a test and a retest one week later is significantly different from zero is ok, but again it's obvious. Instead, we want inference on the value of the correlation in the population.

Defining objectives
Some statistics involved
Units and variables
Sampling
Variable Types
Variable Roles
Univariate statistics
Bivariate statistics
Multivariate statistics
A note on inference 52
The test construction process
Review

Psychometrics and inference

- As a conclusion, we could say that psychometricians care for inference, because it implies the potential generalization on the population of the psychometrical qualities that they observed in the sample.
- But psychometricians most of the time want to conclude about effect sizes (and other reliability/validity statistics) in the population, not only about the fact that such effect sizes are non-zero.
- Therefore, **in psychometrics, Confidence Intervals play a bigger role than Null Hypothesis Significance Testing**.
- An exception of this is perhaps the comparison between measurement models (e.g. two nested Confirmatory Factor Analysis models).

Some statistics involved

A note on inference

Defining objectives

Some statistics involved

- Units and variables
- Sampling
- Variable Types
- Variable Roles
- Univariate statistics
- Bivariate statistics
- Multivariate statistics
- A note on inference

The test construction process

Review

As hypothesized, the SPM-LS with showed satisfactory empirical reliability (empirical $r_{xx',3PL} = 0.843$, Bootstrapped 95% CI [0.828, 0.861]). The marginal reliability estimate for the 3PL model indicates that observations from an assumed normal prior distribution produce overall reliable estimates (marginal $r_{xx',3PL} = 0.829$, Bootstrapped 95% CI [0.812, 0.841]).

Figure: An example: Psychometricians frequently use confidence intervals instead of significance testing (Myszkowski & Storme, 2018)

Outline

Defining objectives

Some statistics involved

The test construction process 54

Review

Defining objectives

Some statistics involved

The test construction process

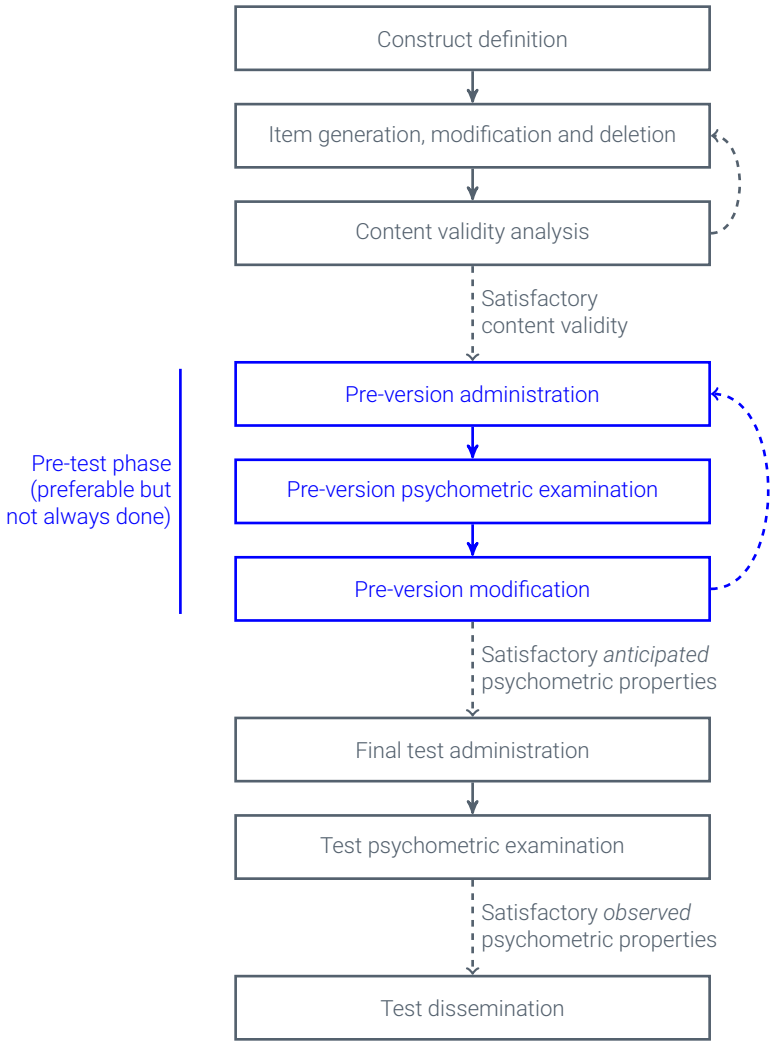
Review

The test construction process

Defining objectives
Some statistics involved
The test construction process
Review

55

The steps of the process



The test construction process

Defining objectives
Some statistics involved
The test construction process
Review

Different steps

- It is important to remember that test construction is based notably on :
  - Generating items wisely
  - Making pre versions and evaluating their psychometrical qualities through empirical evidence (data gathered on samples)

Outline

Defining objectives

Some statistics involved

The test construction process

Review

Defining objectives

Some statistics involved

The test construction process

Review

Review

Defining objectives
Some statistics involved
The test construction process
Review

What you should remember

- Measurement accuracy is ensured by implementing best practices in test construction and psychometric qualities evaluation.
- Best practices in psychometric qualities evaluation imply ensuring appropriate **discrimination power, reliability and validity**.
- Ensuring such psychometric qualities will be based on **empirical results** and specific statistical analyzes.
- In psychometrics, as we are more focused on effect sizes than on binary hypothesis acceptance/rejection, inference is preferably done with Confidence Intervals rather than with Null Hypothesis Significance Testing.

Review

Defining objectives
Some statistics involved
The test construction process
Review

What you should remember

- Measurement accuracy is ensured by implementing best practices in test construction and psychometric qualities evaluation.
- Best practices in psychometric qualities evaluation imply ensuring appropriate **discrimination power, reliability and validity**.
- Ensuring such psychometric qualities will be based on **empirical results** and specific statistical analyzes.
- In psychometrics, as we are more focused on effect sizes than on binary hypothesis acceptance/rejection, inference is preferably done with Confidence Intervals rather than with Null Hypothesis Significance Testing.