



Programa
Universitario
de **Estudios**
del **Desarrollo**

UNAM

Empirical scrutiny of reliability and validity

Hector Najera

Outline

- Quick introduction to R-software and Rstudio
 - Power point presentation
 - Live session
 - Handbook with notes
- Estimation of reliability in R
 - Power point presentation
 - Live session
 - Handbook with notes
- Estimation of validity in R
 - Power point presentation
 - Live session
 - Handbook with notes

Quick introduction to R-software and Rstudio

- Why are we using R and Rstudio?
 - R is a programming language made by statisticians for statisticians
 - has cutting-edge methods, it's almost everything you need.
 - Rstudio is just a more friendly and powerful interface of R
 - Both are free, widely used,
- For reliability has the best estimators available and can be used jointly with more powerful programmes like Mplus for latent variable modelling

R-software

- It started in the early 1990s as an open-source implementation of S-plus language
- It is still free and is one of the main platforms for operationalising the most recent developments in statistics
- Yes but R is difficult, it has a very steep learning curve!
- R works with objects and this is one of the main challenges from people migrating from point-and-click programmes
- But... working with objects is one of its main advantages
- R is much more than a software, it's a movement towards open code, free use and scientific replication.

R-software: Objects

A data set is an object made of other objects:

- Columns and rows

Each column is an object made of other objects:

- Characters -string or numeric-

But the output of a model can be an object

The contents (e.g. the coefficients) are an object too

	cluster	country	v001	id	dhs_year	gdp	parliament	labor_fem	gii	maternal_mor	adolsc_birth	country_num
1	1 1 1	Afghanistan	1	7968	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
2	1 1 1	Afghanistan	1	27901	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
3	1 1 1	Afghanistan	1	21788	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
4	1 1 1	Afghanistan	1	9122	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
5	1 1 1	Afghanistan	1	38147	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
6	1 1 1	Afghanistan	1	29721	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
7	1 1 1	Afghanistan	1	27483	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
8	1 1 1	Afghanistan	1	19997	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
9	1 1 1	Afghanistan	1	39948	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
10	1 1 1	Afghanistan	1	34092	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
11	1 1 1	Afghanistan	1	6597	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
12	1 1 1	Afghanistan	1	7610	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
13	1 1 1	Afghanistan	1	22023	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
14	1 1 1	Afghanistan	1	36438	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
15	1 1 1	Afghanistan	1	30468	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
16	1 1 1	Afghanistan	1	27024	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
17	1 1 1	Afghanistan	1	37749	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
18	1 1 1	Afghanistan	1	28135	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
19	1 1 1	Afghanistan	1	7547	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
20	1 1 1	Afghanistan	1	15171	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
21	1 1 1	Afghanistan	1	32256	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
22	1 1 1	Afghanistan	1	20376	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
23	1 1 1	Afghanistan	1	37605	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
24	1 1 1	Afghanistan	1	33911	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
25	1 1 1	Afghanistan	1	9619	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan
26	1 1 1	Afghanistan	1	26098	2015	.211176	27.70999908	18.79999924	.7020000219	701	100.8000031	Afghanistan

. regress bwght cigs male parity faminc faminc2

Source	SS	df	MS	Number of obs	=	1,387
Model	20057634.7	5	4011526.94	F(5, 1381)	=	12.54
Residual	441644355	1,381	319800.402	Prob > F	=	0.0000
				R-squared	=	0.0434
				Adj R-squared	=	0.0400
Total	461701990	1,386	333118.319	Root MSE	=	565.51

bwght	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cigs	-13.25496	2.588155	-5.12	0.000	-18.3321 -8.177817
male	91.00571	30.45073	2.99	0.003	31.27103 150.7404
parity	47.51602	17.07493	2.78	0.005	14.02041 81.01163
faminc	8.660125	3.013622	2.87	0.004	2.748353 14.5719
faminc2	-.0826576	.0414853	-1.99	0.047	-.1640386 -.0012765
_cons	3115.172	57.80237	53.89	0.000	3001.782 3228.562

Wouldn't be great if we could manipulate all these objects and do whatever we want?

R-software: Objects

- Calculate the same table across objects - Different rounds of a survey-

- Run the same model across different rounds of a survey

- Extract what I need for my documents, etc.

- Reproduce all -data cleaning, data processing, outputs, written document- in one place

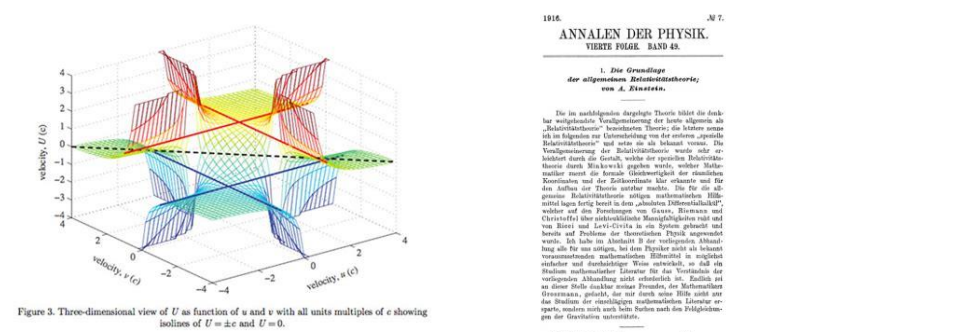
id	country	year	age	edu_year	gdp	particem1	labor_year	gpi	material_wt	material_wt2	country_year
1	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
2	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
3	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
4	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
5	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
6	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
7	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
8	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
9	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
10	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
11	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
12	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
13	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
14	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
15	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
16	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
17	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
18	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
19	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
20	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
21	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
22	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
23	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
24	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
25	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
26	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889

id	country	year	age	edu_year	gdp	particem1	labor_year	gpi	material_wt	material_wt2	country_year
1	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
2	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
3	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
4	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
5	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
6	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
7	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
8	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
9	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
10	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
11	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
12	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
13	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
14	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
15	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
16	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
17	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
18	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
19	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
20	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
21	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
22	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
23	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
24	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
25	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889
26	Algeria	1970	2001	11.1	21.769368	16.789912	761	108.888889	16.789912	761	108.888889

```
regress bwght cigs male parity faminc faminc2
```

Source	SS	df	MS	Number of obs	F(5, 1381)
Model	20057634.7	5	4011526.94		12.54
Residual	441644355	1,381	319800.402		0.0000
Total	461701990	1,386	333118.319		0.0434
					Adj R-squared = 0.0400
					Root MSE = 565.51

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
bwght					
cigs	-13.25496	2.588155	-5.12	0.000	-18.3321 -8.177817
male	91.00571	30.45073	2.99	0.003	31.21703 150.7404
parity	47.51602	17.07493	2.78	0.005	14.02041 81.01163
faminc	8.660125	3.013622	2.87	0.004	2.748353 14.5719
faminc2	-0.0826576	.0414853	-1.99	0.047	-0.1640386 -.0012765
_cons	3115.172	57.80237	53.89	0.000	3001.782 3228.562



1916. ANNALEN DER PHYSIK. VIERTE FOLGE. 8430 43.

1. Die Grundzüge der allgemeinen Relativitätstheorie; von A. Einstein.

Die in vorstehendem dargelegte Theorie bildet die äußerlich weitgehendste Verallgemeinerung der bereits allgemein als „Relativitätstheorie“ bezeichneten Theorie der letzten vierzig Jahre. Sie ist die Grundlage der Relativitätstheorie, welche die spezielle Relativitätstheorie als Grenzfall enthält. Die Theorie ist im wesentlichen die Fortsetzung der allgemeinen Relativitätstheorie, welche im Jahre 1915 von Einstein veröffentlicht wurde. Die Theorie ist im wesentlichen die Fortsetzung der allgemeinen Relativitätstheorie, welche im Jahre 1915 von Einstein veröffentlicht wurde. Die Theorie ist im wesentlichen die Fortsetzung der allgemeinen Relativitätstheorie, welche im Jahre 1915 von Einstein veröffentlicht wurde.

R-software: Languaje

- Everything you do has at least one object as target

```
s <- 1+1
```

“Desired object (name it as you want)” “equals” “1+1”

```
Mymodel <- lm(y ~ x1 + x2, data=D)
```

Your model.

Argument/options of a function

Function: The algorithm that estimates what you want to estimate

“Desired object equals” “linear model using data=D”

Estimation of reliability in R

- Steps:
 - Identify the underlying measurement model
 - Make the model explicit
 - Write the model for computation
 - Estimate the model
 - Assess the model
 - Estimate reliability

Reliability: Potential problems

- The model is not clear, research is at preliminary stages, I only want to explore potential solutions!
- Exploratory methods:
 - Alpha, beta and omega from an Exploratory Factor Analyses are alternatives
 - The relevant library is "psych" -I'll share a note-
 - Eventually this will help you to calibrate a better measurement model but it won't define it -theory defines the model-

Reliability: Potential problems (2)

- The measurement model does not work at all
 - Always fit a null -unidimensional model-
 - Compare the model with the null
- Be careful, confirmatory models cannot be used in an exploratory fashion
- Some minor deviations are acceptable, but all should be based on the theoretical measurement framework
- You need new data to put under scrutiny an improved model

Estimation of validity

- Criterion validity
 - Find a proper validator -theory and data-
 - Use several but sensible specifications
- Latent construct validity
 - Specify the model and assess it
- Criterion + latent construct
 - Find a proper validator and embed the criterion validation into a SEM model

Potential computational problems

- Computation with categorical variables is difficult and complex
- We have much better algorithms and computational power but sometimes we demand a lot
- Maximum likelihood is one of the most robust estimation methods but not always is feasible for categorical data
- Categorical data + complex model: WLSMV estimator (lavaan and Mplus) or MCMC -Bayes-.

Software alternatives

- SPSS+AMOS:
 - CFA -continuous and categorical- reliability estimation -including omega- requires and extra package
 - Alfa, Gutman variants as well
 - IRT. Requires installing an extension
- Stata
 - CFA -continuous and categorical- no reliability estimation. Stata+R
 - Alfa
 - IRT models
- Mplus
 - When combined with R is the best option we have... by far!

A workflow for empirical scrutiny in poverty measurement:

- Assess both global and item reliability
- Assess validity -different aspects of it-
- Look for consistencies and inconsistencies in both analyses
- Rely on theory to make decisions about the assumptions that do not hold given the empirical analyses
 - If a dimension seems a sound construct but its poorly measured, it is not a problem of *content validity* but *with some assumptions within the measurement model, i.e. water deprivation is important but perhaps the threshold you have leads a poor split.*
 - Amend -if possible- and/or drop indicators that add random and systematic error
- Produce deprivation scores, assess the poverty line and the rest of your assumptions.