

# Tests Before Factor Analysis

Luis Carlos Castillo

University of Urbino and University of Bremen

2024-07-25

## Important considerations

In the process of selecting variables consistent across all six years of the ICT usage in enterprises survey (2014-2019), conducted by ISTAT, several challenges were encountered due to variations in variable codification and measurement across different years. These discrepancies suggest that the survey's methodology evolved over time, possibly to adapt to changes in technology or business practices, which occasionally led to alterations in how certain variables were defined or measured.

It is important for readers to be aware that, although the variable names and codes may vary from year to year, careful review has been conducted to ensure that each selected variable consistently measures the intended construct over the entire period. This preparatory work is critical for the next step of the analysis, which will involve data imputation to address any missing or inconsistent data points, thereby preserving the integrity of the analysis.

# List of selected variables

name_EN	ict2014	ict2015	ict2016	ict2017	ict2018	ict2019
company code	Codice	Codice	codice	codice_	codice_	codice_
class of revenues from the sale of goods and services	Ricavi	Ricavi	ricavi	ricavi_cl	ricavi_cl	ricavi_cl
Percentage of employees using the computer out of the total employees	A2	A2	A2	A2_	A3_	A3_
employment of specialists in computer subjects	B1	B1	B1	B1	B1	B1
IT training courses for employees with specialist ict skills	B2a	B2a	B2a	B2a	B2a	B2a
IT training courses for employees without specialist ict skills	B2b	B2b	B2b	B2b	B2b	B2b
Use of internal personnel for ICT infrastructure maintenance	B5a	B5a	B5a	B5a	B5a	B5a
Percentage of employees using computers connected to the internet	C2	C2	C2	C2_	C2_	C2_
Internet download speed low	C4	C4	C4	C4	C4	C4
Enterprise provides mobile devices with mobile connection	C5a	C5	C5	C6	C5	C5
Website	C7	C7	C8	C8	C8	C7
possibility to place orders or reservations online eg online shopping cart	C8a	C8a	C9a	C9a	C9a	C8a
access to product catalogs or price lists	C8c	C8c	C9c	C9c	C9c	C8c
Social network	C9a	C9a	C10a	C10a	C10a	C10a
Social media and multimedia	C9c	C9c	C10c	C10c	C10c	C10c
announcement of vacancies or possibility to apply for employment online	C8g	C8g	C9f	C9f	C9g	C8f
links or references to company profiles on social media	C8h	C8h	C9g	C9g	C9f	C8g
using erp software	E1	E1	E1	D1	D1	D1
use operational crm software	E2b	E2b	E2b	D2b	D2b	D2a
use analytical crm software	E2a	E2a	E2a	D2a	D2a	D2b
web sales through intermediary websites or ecommerce sites marketplaces or apps	J7	I8	H9	G11	I1b	F1B
Size by number of employees	clad4	clad4	clad4	clad3	clad3	clad3
Classification of ICT companies	dom4	dom4	dom4	dom4	dom4	dom4
Region NUTS-1	rip	rip	rip	rip	rip	rip
Groups of economic sectors	ateco_1	ateco_1	ateco_1	Ateco_1	Ateco_1	Ateco_1

## Missing values for 2014

In the analysis of the ICT usage in enterprises survey data from 2014 to 2019, notable gaps in data availability were identified for several key variables, notably variable B5A. This variable presents a unique challenge, with 100% missing data in 2014, reappearing in subsequent years but missing again in 2017 and 2019. These gaps are attributed to changes in the survey's measurement approach and question modifications over the specified years.

To address these inconsistencies and ensure the continuity of our dataset for longitudinal analysis, a strategic imputation approach will be employed. For variable B5A, missing values will be imputed by interpolating or extrapolating from available data points in the adjacent years where the variable is present. This method aims to reconstruct the missing data based on observable trends and values from 2015 and 2018.

Additionally, other variables with approximately 26% missing data will undergo imputation using data from 2014. This approach leverages the more complete dataset from the earliest year in the survey series to fill in gaps in subsequent years, ensuring a robust dataset for comprehensive analysis. This paragraph aims to transparently communicate the methodological adjustments made to compensate for data irregularities across different survey years.

# Missing chart 2014













