Exercise 01

Exercise 1.1

In this analysis, our group decided to analyze the data of Italy.

After data preparation, the original sample size is 1215 observations, while the final sample size is 651, deleting all cases with missing values on the key variables of interest. Looking into the final dataset, 37.6% of the population is 60+ years old. The average household size is 2.95.

Exercise 1.2

a)

Defining as 60 % of the median net equalized household income, 22% of the population is at risk of poverty (see Table 1)

Table 1. The Distribution of the Population at Risk of Poverty

	n	**%**	**val%**
Not poor	508	78	78
Poor	143	22	22

b)

To evaluate the extent of poverty in Italy, one must consider the trade-offs between income-oriented and relative poverty measures. The income-oriented measure, which uses a fixed income level as a benchmark, is easy to apply and compare across countries. However, it may not capture the variations in living expenses and social norms that affect the perception of poverty (Ringen, 1988).

The relative poverty measure, which depends on the income distribution in society, accounts for these variations and adapts to changes in the economy. However, it also faces challenges with defining poverty objectively and dealing with inflationary effects. The issue is that direct welfare measurements are not as developed as indirect ones, most notably the income measure, which forces one to rely more on indirect measurement (Ringen, 1988).

Italy has strong social policies, but also income disparities. The choice to prioritize relative thresholds over absolute or fixed ones in measuring poverty has been relatively consensual, with a low income threshold fixed at a certain point in time being used as a Secondary indicator (Atkinson et al., 2004). Moreover, a balanced approach that combines both income-oriented and relative measures is necessary for a comprehensive analysis (Jenkins & Kerm, 2014).

Exercise 1.3

a)

We assume age matters for poverty in Italy. So here is the testable hypothesis:

H0: Age does not matter for poverty in Italy.
H1: Age matters for poverty in Italy.

Then a chi-square test was conducted to verify the hypothesis. With a p-value of 0.9782, the null hypothesis cannot be rejected. In other words, there is no evidence to suggest that age matters for poverty in Italy.

b)

Distinguish the sample into three age groups: 20-39, 40-59, 60 and older. Table 2 shows the result of the at-risk-of-poverty ratio by age.

Table 2. The at-risk-of-poverty Ratio by Age

	DATA03$AGE_C			
DATA03$POOR	18-39	40-59	60+	Total
Not poor	143	173	192	508
	78.1%	77.6%	78.4%	
Poor	40	50	53	143
	21.9%	22.4%	21.6%	
Total	183	223	245	651
	28.1%	34.3%	37.6%	

Exercise 1.4

a)

Distinguish those without children, with one or two, and those with three or more children as a factor variable, then calculate the at-risk-of-poverty ratio by the children categories, the result is shown below(see Table 3):

Table 3. The at-risk-of-poverty Ratio by Number of Children in Households

DATA03$POOR	DATA03$KIDS_C No kids	1-2 kids	3+ kids	Total
Not poor	372 79.3%	127 78.9%	9 42.9%	508
Poor	97 20.7%	34 21.1%	12 57.1%	143
Total	469 72.0%	161 24.7%	21 3.2%	651

With a p-value much less than 0.05, the null hypothesis can be rejected. Therefore, the number of children in the household matters for poverty in Italy.

b)

We assume rural and urban area matters for poverty in Italy. So here is the testable hypothesis:

H0: Rural or Urban factor does not matter for poverty in Italy.
H1: Rural or Urban factor matters for poverty in Italy.

Merging those “A big city”, “The suburbs or outskirts of a big city”, and “A town or a small city” as “Urban”, and put “A country village” and “A farm or home in the country” into “Rural”, then calculate the at-risk-of-poverty ratio by this factor variable, the result is shown below (see Table 4):

Table 4. The at-risk-of-poverty Ratio by Rural or Urban Areas

DATA03$POOR	DATA03$URBRURAL_C Rural	Urban	Total
Not poor	190 74.8%	318 80.1%	508
Poor	64 25.2%	79 19.9%	143
Total	254 39%	397 61%	651

With a p-value of 0.1348, the null hypothesis cannot be rejected.

Alternative levels of separation also failed to achieve statistical significance. Overall, in Italy, there is no evident statistical power to suggest that the level of urbanization matters for poverty.

References:

Atkinson, A. B., Marlier, E., & Nolan, B. (2004). Indicators and Targets for Social Inclusion in the European Union. *JCMS: Journal of Common Market Studies*, *42*(1), 47–75. https://doi.org/10.1111/j.0021-9886.2004.00476.x

Jenkins, S. P., & Kerm, P. V. (2014). *The Relationship Between EU Indicators of Persistent and Current Poverty*.

Ringen, S. (1988). Direct and Indirect Measures of Poverty. *Journal of Social Policy*, *17*(3), 351–365. https://doi.org/10.1017/S0047279400016858

```r
library(haven)
library(descr)
library(questionr)

#### Exercise 1.1: data preparation ####
rm(list=ls())
DATA00 <- read_dta("assets/ZA7600_v3-0-0.dta", encoding='latin1')

# Select the data from Italy(380)
DATA01 <- subset(DATA00, DATA00$country == 380)

# Select the variables of interest
DATA02 <- subset(DATA01, select = c(IT_INC, HHADULT, HOMPOP, AGE,
URBRURAL))

# IT_INC: Country specific household income
# Select the data with positive regular income
DATA03 <- subset(DATA02, DATA02$IT_INC >= 0 & IT_INC < 5001)

# HOMPOP: How many persons in household
# HHADULT: How many adults in household
# AGE: Age of the respondent
# Delete all cases with missing values on the key variables of interest
DATA03 <- subset(DATA03, DATA03$HOMPOP > 0 & DATA03$HHADULT > 0)
DATA03 <- subset(DATA03, DATA03$HOMPOP >= DATA03$HHADULT)
DATA03 <- subset(DATA03, DATA03$AGE >= 18)

# Generate the new variables that required for the investigation
# Var: Household Size
DATA03$TOTAL <- DATA03$HOMPOP
DATA03$TOTAL <- as.numeric(DATA03$HOMPOP)

DATA03$ADULT <- DATA03$HHADULT
DATA03$ADULT <- as.numeric(DATA03$ADULT)

DATA03$OTHER <- DATA03$ADULT-1
DATA03$OTHER <- as.numeric(DATA03$OTHER)

DATA03$KIDS <- DATA03$TOTAL - DATA03$ADULT
DATA03$KIDS <- as.numeric(DATA03$KIDS)

# Var: Age(categorical)
DATA03$AGE_C <- "NA"
DATA03$AGE_C[DATA03$AGE>=18] <- "18-39"
DATA03$AGE_C[DATA03$AGE>=40] <- "40-59"
DATA03$AGE_C[DATA03$AGE>=60] <- "60+"
DATA03$AGE_C <- as.factor(DATA03$AGE_C)

# Var: Household Income
DATA03$INCOME01 = DATA03$IT_INC
DATA03$INCOME02 = DATA03$IT_INC/(1+DATA03$OTHER*0.5+DATA03$KIDS*0.3)

# Var: AROP
DATA03$POOR <- "Not poor"
```

```r
DATA03$POOR[DATA03$INCOME02 < (median(DATA03$INCOME02)*0.6)] <- "Poor"
DATA03$POOR <- as.factor(DATA03$POOR)

# the share of older people in the population
crosstab(DATA03$AGE_C, DATA03$POOR, prop.r = T)
# 37.6% of the population is 60+ years old
# Poor: 21.6%, nor poor: 78.4%

# how large is the household size on average
mean(DATA03$TOTAL)
# the average household size is 2.95

#### Exercise 1.2 ####

# Calculate the equalized median income (based on modified OECD-scale).
median(DATA03$INCOME02) * 0.6

# How large is the share of persons at risk of poverty
# (defined as 60 % of the median net equalized household income)?

# Poverty Ratio
freq(DATA03$POOR)
# 22% of the population is at risk of poverty

#### Exercise 1.3 ####

# Do you think that age matters for poverty in your country of choice?
# Formulate a testable hypothesis.

# Testable hypothesis:
# H0: Age does not matter for poverty in Italy
# H1: Age matters for poverty in Italy

# Conduct a chi-square test to test the hypothesis
chisq.test(DATA03$POOR, DATA03$AGE_C)
# With a p-value of 0.9782, the null hypothesis cannot be rejected.
# In other words, there is no evidence to suggest that age matters for
poverty in Italy.

# Calculate the at-risk-of-poverty ratio by age.
# Distinguish the following age groups: 20-39, 40-59, 60 and older.
crosstab(DATA03$POOR, DATA03$AGE_C, prop.c = T)

#### Exercise 1.4 ####

# distinguish those without children, with one or two and those with three
and more children as a factor variable
DATA03$KIDS_C <- "NA"
DATA03$KIDS_C[DATA03$KIDS==0] <- "No kids"
DATA03$KIDS_C[DATA03$KIDS==1 | DATA03$KIDS==2] <- "1-2 kids"
DATA03$KIDS_C[DATA03$KIDS>=3] <- "3+ kids"
DATA03$KIDS_C <- as.factor(DATA03$KIDS_C)

# change the order of the factor levels as the number of children
increases
DATA03$KIDS_C <- factor(DATA03$KIDS_C, levels = c("No kids", "1-2 kids",
"3+ kids"))
```

```r
# at-risk-of-poverty ratio by number of children in households
crosstab(DATA03$POOR, DATA03$KIDS_C, prop.c = T)

# Chi-square test
chisq.test(DATA03$POOR, DATA03$KIDS_C)
# With a p-value much less than 0.05, the null hypothesis can be rejected.
# In other words, the number of children in the household matters for
poverty in Italy.


# (b)
# There is a debate of whether national or sub-national units should be
used to define poverty.
# Is this a relevant consideration for your country of choice?
# How do patterns differ if you use different levels for rural and urban
areas?

# distinguish only urban and rural places as a factor variable
DATA03$URBRURAL_C <- "NA"
DATA03$URBRURAL_C[DATA03$URBRURAL <= 3] <- "Urban"
DATA03$URBRURAL_C[DATA03$URBRURAL > 3] <- "Rural"
DATA03$URBRURAL_C <- as.factor(DATA03$URBRURAL_C)

# At-risk-of-poverty ratio by urban and rural areas
crosstab(DATA03$POOR, DATA03$URBRURAL_C, prop.c = T)

# Chi-square test
chisq.test(DATA03$POOR, DATA03$URBRURAL_C)
# With a p-value of 0.1348, the null hypothesis cannot be rejected.
# In other words, there is no evidence to suggest that the level of
urbanization matters for poverty in Italy.
```