CASE STUDY:
HEALTH DETERMINANTS ACROSS COUNTRIES

To: Prof. Paweł Strzelecki
Prof. Izabela Grabowska
Prof. Maja Rynko
Prof. Lucas van der Velde
From: Thi Thu Trang Do (75098)
Katarzyna Szołucha (75171)
Date: 30 May, 2017
Subject: Report on case study: Health determinants across countries

I. Introduction

The purpose of this report is to evaluate several statements on the BMI by analyzing the sample (25% of the original) from two countries which are CL-Chile and TR-Turkey. The original data is conducted in 2011 by the International Social Survey Programme.

II. Statement of findings

1. Data preparation

1.1.

The number of missing values in each category

	Height	Weight	Sex	Degree	Exercise	Drinking	Smoking
Do not know	0	0	-	-	-	-	-
No answer	7	3	0	1	10	8	4

Total number of observations	468
Number of observations containing at least 1 missing answer	24
Proportion	5.13%

Source: 1-ORIGIN Excel Sheet

If we use the database without corrections, the later results can change dramatically. There are two main reasons.
Firstly, the values for missing variables are significantly higher than those for normal variables. For example, concerning variable age, we have values (15-102) for normal variable and 999 for missing variables.
Secondly, the proportion of these missing variables is considerable (i.e. 5.13%).

The table below is an example for how these missing values affect our results.

Measures of height for all sample with corrections and without corrections

	Without corrections	With corrections
Mean	176.78	164.26
Standard deviation	101.83	9.11

Source: 1-ORIGIN Excel Sheet

1.2.

Result of Calculating Confidence Interval

Proportion of observations containing at least 1 missing answer	5.13%
Confidence coefficient	95%
Alpha	0.05
Test statistic (z)	1.96
Standard Error	0.01
Sampling Error (SE)	0.02
Confidence Interval	<3.13% ; 7.13%>

Source: 1-ORIGIN Excel Sheet

In the population, it is 95% certain that proportion of observations that contain at least one missing answer falls into this interval <3.13% ; 7.13%>.

1.3. The BMI is calculated in Excel sheet 1-*ORIGIN* by the following formula.

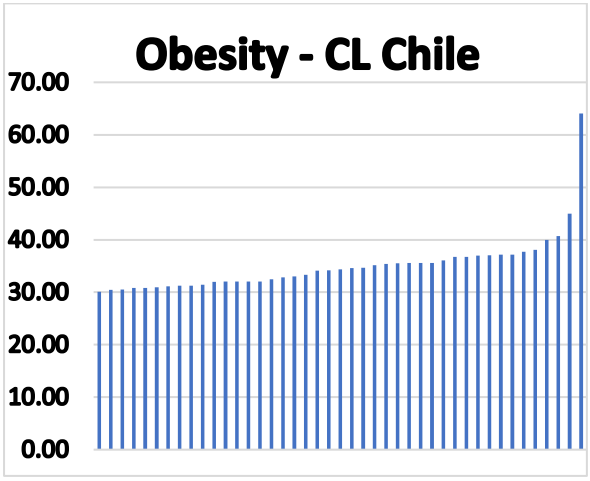
$$\text{BMI}= \frac{weight\,[kg]}{height[m]^2}$$

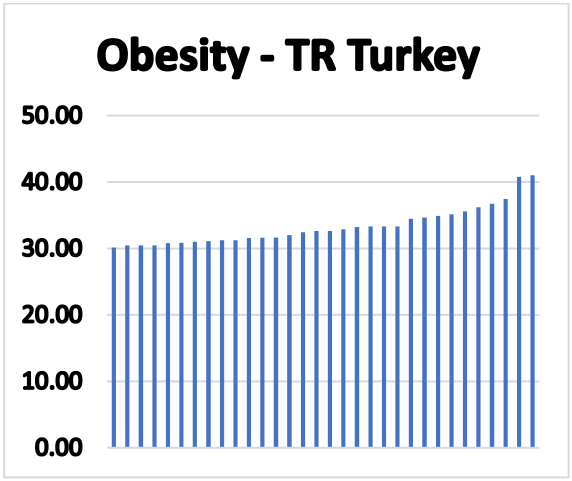
2. Initial Comparisons

2.1.

2.1.1. **Measures of BMI**

	CL CHILE	TR TURKEY
Mean	26.63	25.65
Standard deviation	5.32	4.56
Skewness	2.11 (rightward)	0.48 (rightward)
Median	25.56	25.51
Q1	23.55	22.50
Q3	28.72	28.41





Source: CL-CHILE and TR-TURKEY Excel Sheets

The graphs indicate that the number of obese observations for Chile is higher than that for Turkey. It means there are more obese people in Chile than in Turkey. Moreover, the highest value of BMI for Chile is 64.09 kg/m^2 which is more than 1.5 times as much as that value for Turkey.

2.1.2.

Outliers (measured in BMI)

	Chile	Turkey
Lower limit m-3S	10.67	11.99
Upper limit m+3S	42.58	39.33
Outliers	44.98 64.09	40.79 41.02

Source: CL-CHILE and TR-TURKEY Excel Sheets

We exclude above outliers because they are higher than Upper limit X+3S.
Because both distributions of two countries have rightward skewness, we would conclude that in each country, 50% of population has lower value of BMI than mean of BMI in that country. (Outliers are highlighted in the Excel file)

2.2.

Comparisons of the means of two countries

	Chile	Turkey
Mean (m1,m2)	26.38	25.51
Standard deviation (S1,S2)	4.57	4.33
Total number of observations (n1,n2)	229	211
Proportion of obese observations (p1,p2)	17.90%	14.22%

	Comparison of the means of BMI of two countries	Comparison of the proportion of obese observations of two countries
Null hypothesis - H0	m1=m2	p1=p2
Alternative hypothesis - Ha	m1 and m2 are different	P1 and p2 are different
Test statistic – z	2.05	1.05
P value	0.04 (<0.05)	0.29 (>0.1)
Decision	Reject H0	Do not reject H0

Source: 2-COMPARE Excel Sheet

Conclusion:
1) The means of two countries are different. The probability of the situation when both means are different is 4%. The averages BMI of residents of two countries are different.
2) The proportion of obese observations of two countries are equal. The number of obese people in two countries are equal.

3. Risk factor

3.1. Comparison of the proportions of overweight observations for male and female in two countries

Method: Test of hypotheses
H0: p1=p2
Ha: p1>p2

	Chile		Turkey	
Data	p1=0.64	81	p1=0.57	n1=83
	p2=0.48	n2=148	p2=0.53	n2=128
Test statistic - z	2.35		2.05	
P value	0.009 (<0.05)		0.31 (>0.1)	
Decision	Reject H0		Do not reject H0	
Source	*3A-CHILE Excel Sheet*		*3B-TURKEY Excel Sheet*	

Conclusion:
1) For Chile, the proportion of overweight observations for male is higher than that for female. Men are more likely to have overweight than women. 0.9% is the probability of the circumstance when the probability of being overweight of male and female are equal.
2) For Turkey, the proportions of overweight observations for men and women are equal. Men are not more likely to have overweight than women.

3.2. Comparison the means of BMI of workers of different educational levels

Method: ANOVA
H0: m0=m1=m2=m3=m4=m5=m6
Ha: Not all means are equal.

	Chile	Turkey
SSB	458.10	285.769
SSE	4296.58	3646.025
r	7	6 (no one gets Master or PhD degree)
F	3.94	3.21
P value	0.0009 (<0.05)	0.008 (<0.05)
Decision	Reject H0	Reject H0
Source	*3B-CHILE Excel Sheet*	*3B-TURKEY Excel Sheet*

Conclusion:
For both Chile and Turkey, not all means are equal. Workers of different educational levels have different average values of BMI.

3.3. Finding the relationship between age and BMI

Method: Chi square
H0: Age and BMI are independent
Ha: Age and BMI are dependent
***Note: The frequency of people that are underweight is too small so we will not take it into consideration.

	Chile	Turkey
Chi square - χ^2	21.95	10.91
P value	0.005 (<0.05)	0.21 (>0.1)
Decision	Reject H0	Do not reject H0
Source	*3C-CHILE Excel Sheet*	*3C-TURKEY Excel Sheet*

Conclusion:

1) For Chile, Age and BMI are dependent with the scales of independence which are V cramer coefficient and Pearson coefficient of correlation are 0.22 and 0.23 respectively. It would mean that the older have on average slightly higher BMI.
   Coefficient of correlation is significant.
   The proportion of the case when age and BMI are independent is 0.5%.
2) For Turkey, Age and BMI are independent. It is not true that the older have on average lower BMI.

3.4. Finding relationships between Exercise and BMI, Drinking and BMI, and Smoking and BMI respectively

Method: Chi square

Excel results

Hypotheses	Measures	Chile	Turkey
H0: Exercise and BMI are independent Ha: Exercise and BMI are dependent	Chi square - χ^2	1.22	0.08
	P value	0.54 (>0.1)	0.96 (>0.1)
	Decision	Do not reject H0	Do not reject H0
H0: Drinking and BMI are independent Ha: Drinking and BMI are dependent	Chi square - χ^2	80.03	0.08
	P value	0.00 (<0.1)	0.96 (>0.1)
	Decision	Reject H0	Do not reject H0
H0: Smoking and BMI are independent Ha: Smoking and BMI are dependent	Chi square - χ^2	4.89	1.14
	P value	0.09 (<0.1)	0.56 (>0.1)
	Decision	Reject H0	Do not reject H0
Source		*3D-CHILE Excel Sheet*	*3D-TURKEY Excel Sheet*

Conclusion:

1) For Chile, exercise and BMI are independent, while the consumption of cigarettes and alcohol and BMI are dependent. The proportion of the case when smoking and BMI are independent is 9%.
The scales of dependence including V cramer coefficient and Pearson coefficient of correlation are respectively 0.6 and 0.09 for drinking and 0.15 and 0.08 for smoking. It would mean that the less the consumption of cigarettes and alcohol smoke is, the slightly higher BMI people get.
These coefficients of correlation are significant.

2) For Turkey, all of three elements which are exercise, the consumption of alcohol and the consumption of cigarettes and BMI are independent.
It is not true that the consumption of cigarettes and alcohol is related to greater BMI, whereas people that exercise regularly have lower values of BMI.

III. Conclusions

To conclude, ANOVA is used to compare the means of more than one small sample and check whether all means are equal, whereas Correlation and Chi square is for checking the relationships between two variables and calculating the scale of dependence.