

Case study: Health determinants across countries

May 2017

The file contains the text of the exercise that is to be solved using the data included in the Excel file DataXXXX. The XXXX gives the number of the dataset which was assigned to you and which you are obliged to use.

Using the proper dataset please solve the Case study and prepare a short report (up to five pages, including tables and graphs) which describes the obtained results, provides answers to the questions and interpretations. Please send the report (possible formats: doc, docx, pdf) together with the solution in the Excel file by 31st May 2017 (midnight) to: case_study_2017@yahoo.com

Both files should be named with students' surname and the number of the dataset in the following way: surname1_surname2_XXXX.

We will mainly verify the file with the text of the report (Word or pdf), but we may refer to the Excel file to check the computations or solve any possible doubts. Making the excel file readable will help us on correcting.

Max number of points: 10.

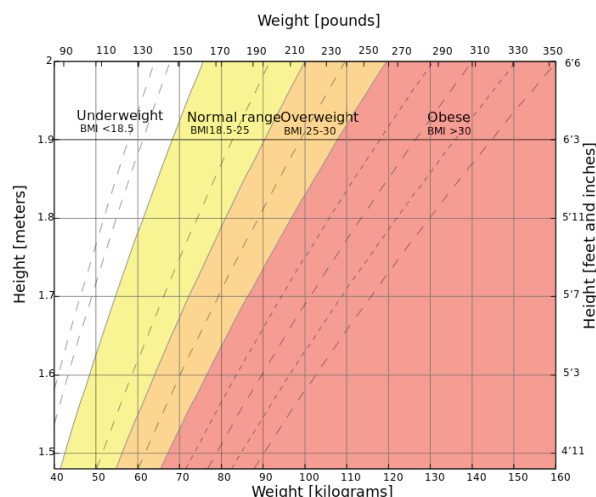


Figure 1: The BMI index - see more at Wikipedia

1 Introduction

Since the late 1980's, the body mass index has been used as a measure of overall adiposity and physical condition. While the measure is not perfect, it is based on information that

is easy to collect, both because it does not require particularly precise instruments nor any invasive techniques. In 2011, the International Social Survey Programme compiled data on participants from around the world. In this exercise, you will have to use your knowledge of statistics to evaluate several statements on the BMI. In order to do so, you are given a random sample (25% of the original) from two countries.

2 Exercises

1. Data preparation:

- Identify those observations that refuse to participate in the survey or did not provide information on some key variables (sex = 9, age=999, height and weight = 998 or 999 or 0, smoking = 98 or 99, drinking and exercise= 8 or 9, degree = 9). How would these observations affect your later results if you use the database without any corrections?
- Calculate the proportion of observations that contain at least one missing answer and estimate a 95% confidence interval for that proportion. Before next steps discard all observations that contain at least one missing value.
- Use information on height (cm) and weight (kg) to calculate the body mass index (BMI) for each person (The BMI is defined as the body mass divided by the square of the body height, and is universally expressed in units of kg/m², resulting from mass in kilograms and height in metres).

2. Initial comparisons

- Obtain measures of BMI for both countries (mean, standard deviations, Skewness, quantile). Interpret them in terms of obesity in both countries. Be creative. The use of graphical tools for data presentation and its correct interpretation is an asset. Exclude any possible outliers (please mention in the report which observations you excluded and why).
- Compare the BMI of both countries by means of a test.
- Interpret the results

3. Risk factors: Looking at each country separately test whether various hypotheses are true using t-test, correlation, chi-square test or ANOVA. State clearly the null and the alternative hypotheses.

- a) Men are more likely to have overweight (BMI>25) than women. Use variable sex (=1 for men and =2 for women)
- b) Workers of different educational levels have different average values of the BMI. Use variable degree.
- c) Older people have on average lower BMI. Use variable age.
- d) The consumption of cigarretes and alcohol is related to greater BMI, whereas people that exercise regularly have lower values of BMI. Use variables smoking, drinking, exercise respectively.

4. Write your conclusions concerning hypotheses tested in previous point.