

### **Question 1:**

Use data set "Pima.tr" in the MASS package that includes a population of women who were at least 21 years old, of Pima Indian heritage and living near Phoenix, Arizona. They were tested for diabetes according to World Health Organization criteria. These data sets contains the following variables:

npreg: number of pregnancies.

glu: plasma glucose concentration in an oral glucose tolerance test.

bp: diastolic blood pressure (mm hg).

skin: triceps skin fold thickness (mm).

bmi: body mass index ( $\text{weight in kg}/(\text{height in m})^2$ ).

ped: diabetes pedigree function.

age: age in years.

Type: Yes or No, for diabetic according to WHO criteria.

Use R commands to conduct the following tasks

- (a) Describe the shape of the distribution of the body mass index
- (b) Construct the relative frequency distribution for the diabetic and non-diabetic women, represent these data as Pie-chart.
- (c) Construct the frequency distribution of plasma glucose concentration using 6 intervals.
- (d) Represent the previous distribution using different colors for each interval, and suitable labels.
- (e) Is the skin fold thickness normally distributed
- (f) Plot the boxplot for Skin Thickness differentiated by Diabetes.
- (g) Test if the mean of the Skin Thickness has the same value between the two group diabetic and non-diabetic women.
- (h) Conclude a regression model to describe the effect of being diabetic or not on the skin fold thickness.
- (i) Add a new variable to the dataset named "obese" with the value TRUE if  $\text{bmi} > 30$
- (j) Test if the proportion of obese women is equal to lean women (hint: use `prop.test()` function)

### **Question 2:**

Using the R code:

- (a) Create a vector x of 100 random numbers from the standard normal distribution.
- (b) Perform a basic descriptive analysis on x
- (c) plot the histogram of x with suitable settings

### **Question 3:**

Using the R commands:

Install the package ISwR, and View the “Blood Pressure Versus Obesity” dataset bp.obese.

- (a) Conclude the linear regression equation
- (b) Plot the scatter plot for obesity versus blood pressure, using different plotting characters(pch) to distinguishing the data between men and women (Hint: Use the R-documentation OR internet).
- (c) Add a line representing the regression model to the previous scatter plot
- (d) Add a legend to the graph depicting the different "pch" used to represent male and female (Hint: Use the R-documentation OR internet).