Harald Schmidbauer
Sema Bayraktar
Barış Yilmazsoy
Yelda Yücel

İstanbul Bilgi Üniversitesi

Bus 273: Statistical Analysis for Business

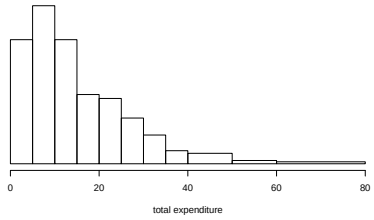
Fall 2009

PROBLEM SHEET # 4

Problem: The data in file `supermarket.xls` show 508 observations of three variables concerning customers of a supermarket: total expenditure (in euros), day of the week, sex. In the following, we investigate the total expenditure data in this file, that is: the distribution of the variable

$$X = \text{ total expenditure (in euros) of a customer at the supermarket.}$$

a) Draw a histogram of the distribution of the variable *total expenditure*, based on the breakpoints 0, 5, 10, 15, 20, 25, 30, 35, 40, 50, 60, 80. Hint: Your histogram should look more or less like this:



b) What can you say about the shape of the distribution?

c) Compute the arithmetic mean and the median. Compare their relative positions with the modal group in the histogram of X.

d) Compute the variance, the standard deviation and the coefficient of variation. (The latter is defined as the standard deviation, divided by the arithmetic mean.)

e) Compute the skewness and the kurtosis and interprete the numbers you computed.

f) Draw a boxplot of the distribution of X.

g) All observations are given in euros. Convert X to TL. Which of the measures you computed in (c)–(e) change? How do they change? Why?

h) Suppose we had similar data from supermarket in Turkey. How could we meaningfully compare the variation in both datasets?

i) Now omit all those values which are above 70 euros. (Four of the 508 observations are above 70.) Compute the measures in (c)–(e) again. Which of them change their values the most?

j) Check if the sigma-rules hold for X in this dataset.