

FEC 522: Financial Econometrics II

Spring 2011

Example: Supermarket Customers

The data in file `supermarket_data.2011-03-21.csv` show 508 observations of the following variables concerning customers of a supermarket:

symbol	description	values
day	day of the week	Mon,Tue,...,Sat
time	time of day	e.g. 09:23
sex	gender	f,m
items	number of items	1,2,...
total	total expenditure	e.g. 10.07
total.so	expenditure on special offers	e.g. 0
	in particular:	
meat	meat etc. (deli counter)	0
cheese	cheese (counter)	0
bread	bread, pastries (counter)	0
fruit	fruit, vegetables	3.31
dairy	dairy products	0
snacks	snacks	0
eggs	eggs	0
tea.coff	tea, coffee	0
drinks	beverages, drinks	1.78
other	others	4.68
suppl	bottle deposit, bags	0.30

Define X = total expenditure.

1. Plot a histogram of X , based on intervals $[5(i-1), 5i)$, $i = 1, \dots, 8$, $[40, 50)$, $[50, 60)$, $[60, 80)$.
2. Draw a boxplot for X .
3. Draw boxplots for X , differentiating by gender.
4. Draw boxplots for X , differentiating by day of the week.
5. Compute the median and the arithmetic mean of X . What does the relative position of modal interval, median, and arithmetic mean tell us?
6. Add vertical lines to the histogram to indicate median and arithmetic mean.
7. Compute the skewness and the kurtosis of X .
8. Using bootstrap, compute standard errors skewness and kurtosis of X . Why should we not use the approximate formulas derived for the normal distribution, as some software packages do? Why is it useful to know the standard errors? What can we do with them?
9. How many customers buy
 - (i) cheese,
 - (ii) cheese and/or bread?
10. Are buying cheese / buying bread independent? What could our finding imply for managing a supermarket?