

# Economics 257 – Assignment 1

At least one student sensibly asked what degree of precision is wanted for the answers to this assignment. It is of course silly to give ten digits after the decimal point if only the first two are meaningful. That is the case here, and so there is no need to give any higher degree of (spurious) precision.

Here is the first table of numbers in Chapter 1 of the textbook.

(omitted)

You are asked to compute, for this table of 200 numbers, the various descriptive statistics presented in Chapter 3:

**D3.0** The sample mean

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N x_i.$$

Note that here  $N = 200$ .

$$\bar{X} = 2.58.$$

**D3.1** A set of trimmed means:

$$\bar{X}_{[k,k]} = \frac{1}{N - 2k} \sum_{i=k+1}^{N-k} x_{(i)},$$

for  $k = 1, 5, 10, 20$ . Recall that the notation  $x_{(i)}$  refers to the  $i^{\text{th}}$  order statistic.

The trimmed means are as follows:

$k$	trimmed mean
1	2.55
5	2.45
10	2.37
20	2.23

**D3.2** The sample median, with this definition:

$$\hat{q}_{0.5} = \begin{cases} x_{(j)} & j = (N + 1)/2, \text{ if } N \text{ is odd;} \\ \frac{x_{(j)} + x_{(j+1)}}{2} & j = N/2, \text{ if } N \text{ is even.} \end{cases}$$

Since  $N$  is even, we use the second formula above, and find that the median is 1.88.

**D3.3** Some quantiles, with the following definition:

$$\hat{q}_\alpha = \min_j x_{(j)} \text{ such that } j/N \geq \alpha.$$

Obtain the median ( $\alpha = 0.5$ ), and the quintiles.

With only two digits after the decimal point, we find that the median is once more equal to 1.88. If we went to the third decimal place, we would find that the first median is actually 1.885, while the second is 1.880.

The quintiles are given in the following table:

$\alpha$	quintile
0.2	0.76
0.4	1.37
0.6	2.30
0.8	4.20

**D3.4** The sample variance

$$s^2 = (N - 1)^{-1} \sum_{i=1}^N (x_i - \bar{X})^2.$$

Would your answer be different if we had  $x_{(i)}$  instead of  $x_i$  in the above formula?

The variance is 4.90.

No, the answer would not be different if the order statistics were used. The calculation of the mean,  $\bar{X}$ , does not depend on the order of the data, and the same is true of the sum used in the definition of the variance.

**D3.5** The standard error:

$$s = +[(N - 1)^{-1} \sum_{i=1}^N (x_i - \bar{X})^2]^{1/2}.$$

The standard error is 2.21.

**D3.6** The range:  $x_{(N)} - x_{(1)}$ .

The largest number in the table is 10.68, and the smallest is 0.09, so that the range is 10.59.

**D3.7** The interquartile range:  $\hat{q}_{0.75} - \hat{q}_{0.25}$ .

The interquartile range is 2.62.

**D3.8** The coefficient of skewness:

$$\frac{(N-1)^{-1} \sum_{i=1}^N (x_i - \bar{X})^3}{s^3}.$$

This measure gives 1.36.

**D3.9** The alternative skewness measure:  $(\bar{X} - \hat{q}_{0.5})/s$ .

Are the results of this and the preceding computation directly comparable? If so, why; if not, why not?

This second measure gives a very different answer, namely 0.31.

The measures are directly comparable only in the sense that they are both dimensionless quantities, but they clearly provide measures of a different sort.

**D3.10** The coefficient of kurtosis:

$$\frac{(N-1)^{-1} \sum_{i=1}^N (x_i - \bar{X})^4}{s^4} - 3.$$

Are the results for skewness and kurtosis compatible with the idea that these numbers were drawn from a normal distribution?

The coefficient of kurtosis is 1.50.

The normal distribution has the coefficients of both skewness and kurtosis equal to zero. The values found here, 1.36 and 1.50, appear to be rather different. This suggests (nothing more!) that the numbers were *not* drawn from a normal distribution.

For information, here is the *Ects* code I used to obtain the answers for this assignment.

```
# Read in the data as a 20x10 matrix
readmatrix ascii e257_as1.dat X 20 10

# Only 2 digits past the decimal point
setprecision 2

# Sample size
set N = 200
sample 1 N

# Group all 200 elements in one column vector
mat X = reorder(X,N,1)

# Formula for the sample mean
sets mean = sum(X)/N

# Sort data in increasing order to get order statistics
gen X = sort(X)

# Values of k for trimming
mat k = rowcat(1,5,10,20)

# make space for results
mat trimmed = emptymatrix(4)

#loop over values of k
set i = 0
while i < 4
  set i = i+1
  del x
  sample k(i)+1 N-k(i)
  gen x = X
  sample 1 N
  set trimmed(i) = sum(x)/(N-2*k(i))
end

# Create matrix with k in 1st column, trimmed mean in 2nd
mat trim = colcat(k,trimmed)

# Median by first formula
set j = N/2
sets median1 = (X(j)+X(j+1))/2

# Median by second formula
set a = 0.5
set j = ceil(a*N)
sets median2 = X(j)

# Create vector of fractions with 5 in denominator
sample 1 4
gen a = time(0)/5
```

```

# Compute quintiles
gen j = ceil(a*N)
gen quintile = X(j)

# Create matrix with alpha in 1st column, quintile in 2nd
mat quintile = colcat(a,quintile)

# Reset sample size
sample 1 N

# Centre data
gen Xcen = X-mean

# Compute variance and standard error
gen X2 = Xcen^2
sets variance = sum(X2)/(N-1)
sets stderror = sqrt(variance)

# Formula for range
sets range = X(N)-X(1)

# Formula for interquartile range
sets interquartile = X(ceil(0.75*N))-X(ceil(0.25*N))

# Two measures of skewness
gen X3 = X2*Xcen
sets skewness = sum(X3)/(N-1)/stderror^3
sets altskew = (mean-median2)/stderror

# Formula for kurtosis
gen X4 = X3*Xcen
sets kurtosis = sum(X4)/(N-1)/variance^2-3

# Print results
print mean
print trim
setprecision 3
print median1
print median2
setprecision 2
print quintile
print variance
print stderror
print range
print interquartile
print skewness
print altskew
print kurtosis

```