BS2280 – Econometrics I
Homework 1: Review of Statistics - Solution

October 11, 2023

1

A survey was conducted to ask employees how long they commute to work (in minutes). The results are captured in Table 1

PID	**Commuting Time**
1	30
2	45
3	10
4	91
5	5
6	18
7	46

Table 1: Results of a Transport Survey

a. What is the arithmetic mean for the commuting time?

   Arithmetic (sample) mean:$\bar{x} = \frac{30+40+10+91+5+18+46}{7} = 35$

b. What is the variance and standard deviation of the commuting time?

   To calculate the variance it is useful to use the sorted data, but it is not strictly required. One could use the unsorted data, but it should be easier to avoid mistakes using the sorted data. The sorted data is given by the following array:
   5, 10, 18, 30, 45, 46, 91

   The formula for the (sample) variance is given by:

$$s^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{n-1}$$

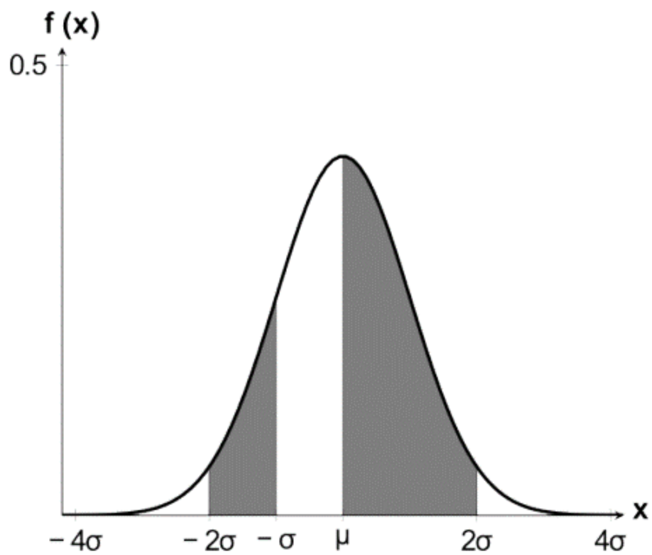
   We already calculated the mean and found the value $\bar{x} = 35$. Inserting the values, yields:

$$
\begin{aligned}
s^2 &= \frac{[(5-35)^2+(10-35)^2+(18-35)^2+(30-35)^2+(45-35)^2+(46-35)^2+(91-35)^2]}{7-1} \\
&= \frac{900+625+289+25+100+121+3136}{6} = 866 \\
s &= \sqrt{866} \\
&\approx 29.43
\end{aligned}
$$

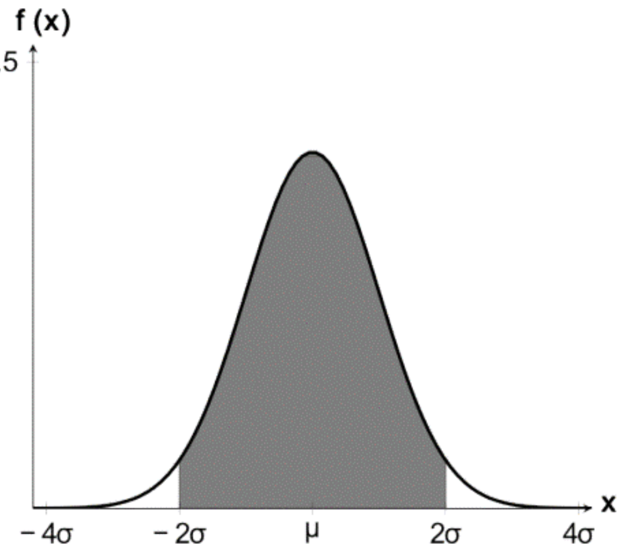
Why is the numerator of the sample variance $n-1$? If we would not do it, the estimator would not converge to the true value of the variance.

2

Calculate the probability corresponding to the shaded area



Use figure in lecture slides to derive the size of areas
Right area: 47.5, left area $47.5 - 34 = 13.5$, total $= 61$



Use figure in lecture slides to derive the size of areas
total = 95

3

R will often present its output with scientific notation, so a good understanding of scientific notation is important for interpreting regression results.

a. Write the following numbers in scientific notation:

- -0.1
  $-1e-1$ (equivalent to -1×10^{-1})

- 100
  $1e2$ (equivalent to 1×10^{2})

- 0.000000054
  $5.4e{-8}$ (equivalent to 5.4×10^{-8})

- 814,502,856,329,062,153,636
  $8.1e20$ (equivalent to 8.1×10^{20})

b. Write the following numbers derived from R in decimal form:

- 3e01
  30

- 3e-01
  0.3

- 6.5e-09
  0.0000000065

- 5.2e05
  520,000

4

A random variable X is defined to be the difference between the higher value and the lower value when two dice are thrown. If they have the same value, X is defined to be zero.

a. Find the probability distribution for X. (Hint: Use a table to identify all possible outcomes, see lecture slides on "Discrete Random Variables and Expectations ")

green \ red	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

The table above show the 36 possible outcomes. The probability distribution is derived by counting the number of times each outcome occurs and dividing by 36. The probabilities in the table below have been written as fractions, but they could equally well have been written as decimals

Value of X	0	1	2	3	4	5
Frequency	6	10	8	6	4	2
Probability	6/36	10/36	8/36	6/36	4/36	2/36

b. Find the expected value of X.

c. Calculate $E(X^2)$ for X.

X	p	Xp
0	$6/36$	$0 \times 6/36 = 0$
1	$10/36$	$1 \times 10/36 = 10/36$
2	$8/36$	$2 \times 8/36 = 16/36$
3	$6/36$	$3 \times 6/36 = 18/36$
4	$4/36$	$4 \times 4/36 = 16/36$
5	$2/36$	$5 \times 2/36 = 10/36$
Total		$E(X) = \mu$=70/36=1.9444

X	X^2	p	X^2p
0	0	$6/36$	0
1	1	$10/36$	$10/36$
2	4	$8/36$	$32/36$
3	9	$6/36$	$54/36$
4	16	$4/36$	$64/36$
5	25	$2/36$	$50/36$
Total			$E(X^2)$=210/36=5.8333

d. Calculate the population variance and the standard deviation of X. (Hint: use $\sigma^2 = \sum (X - E(X))^2 p$ to calculate the variance)

X	p	$X - \mu$	$(X - \mu)^2$	$(X - \mu)^2 p$
0	$6/36$	$(0 - 1.9444) = -1.9444$	$(-1.9444)^2 = 3.7087$	$3.7807 \times 6/36 = 0.6301$
1	$10/36$	$(1 - 1.9444) = -0.9444$	$(-0.9444)^2 = 0.8919$	$3.7807 \times 10/36 = 0.2477$
2	$8/36$	$(2 - 1.9444) = 0.0556$	$(0.0556)^2 = 0.0031$	$3.7807 \times 8/36 = 0.0007$
3	$6/36$	$(3 - 1.9444) = 1.0556$	$(1.0556)^2 = 1.1143$	$3.7807 \times 6/36 = 0.1857$
4	$4/36$	$(4 - 1.9444) = 2.0556$	$(2.0556)^2 = 4.2255$	$3.7807 \times 4/36 = 0.4695$
5	$2/36$	$(5 - 1.9444) = 3.0556$	$(3.0556)^2 = 9.3367$	$3.7807 \times 2/36 = 0.5187$
Total				Population variance = 2.0525

The population variance is 2.05 and the standard deviation is $\sqrt{2.05} = 1.43$

5 *

The formula for the population variance is:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(Y_i - \mu)^2}{N}$$

Show with using Summation Operator rules (see lecture slides) how this formula can be simplified to: (Hint: $\sum_{i=1}^{N} Y_i = \sum Y = N\mu$)

$$\sigma^2 = \frac{\sum_{i=1}^{N} Y_i^2}{N} - \mu^2$$

$$
\begin{aligned}
\sigma^2 &= \frac{\sum_{i=1}^{N}(Y_i-\mu)^2}{N} \\
&= \frac{\sum_{i=1}^{N}\left(Y_i^2-2Y_i\mu+\mu^2\right)}{N} && \textit{we expand} \\
&= \frac{\sum_{i=1}^{N}Y_i^2-\sum_{i=1}^{N}2Y_i\mu+\sum_{i=1}^{N}\mu^2}{N} \\
&= \frac{\sum_{i=1}^{N}Y_i^2-2\mu\sum_{i=1}^{N}Y_i+\sum_{i=1}^{N}\mu^2}{N} \\
&= \frac{\sum_{i=1}^{N}Y_i^2-2\mu N\mu+N\mu^2}{N} && \textit{use } \textstyle\sum_{i=1}^{N}Y_i=\sum Y=N\mu \\
&= \frac{\sum_{i=1}^{N}Y_i^2-2N\mu^2+N\mu^2}{N} \\
&= \frac{\sum_{i=1}^{N}Y_i^2-N\mu^2}{N} \\
&= \frac{\sum_{i=1}^{N}Y_i^2}{N}-\mu^2
\end{aligned}
$$