Tutorial 2 (chapter 2)

Exercise 2.4:

What is a parameter? Work out mean and variance for the following given population values:

44, 56, 60, 48, 55, 50, 58, 62, 60, 40.

Solution

- Any real valued function of variable values for all the population is known as a population parameter or simply a parameter.

Mean: $\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{44+56+60+48+55+50+58+62+60+40}{10} = 53.3$

$$\sum_{i=1}^{N} Y_i^2 = 2890.9$$

Population Variance: $\sigma^2 = \frac{1}{N}\sum_{i}^{n} Y_i^2 - \bar{Y}^2 = 2890.9 - 53.3^2 = 50.01$

By R:

```
> Y <- c(44,56,60,48,55,50,58,62,60,40)
> (Ybar <- mean(Y))
[1] 53.3
> (Sigma <- var(Y)*(9/10))
[1] 50.01
```

Exercise 2.9:

Assuming that 20, 12, 15, 16, 18, 14,22,28,24, and 26 are the observations for a sample of 10 units, calculate sample mean and sample mean square.

Solution

Sample mean: $\bar{y} = \frac{\sum_{i=1}^{N} y_i}{n} = \frac{20+12+15+16+18+14+22+28+24+26}{10} = 19.5$

$$\sum_{i=1}^{n} y_i^2 = 4065$$

Sample mean square: $s^2 = \frac{1}{n-1}(\sum_i^n y_i^2 - n\bar{y}^2) = \frac{1}{9}(4065 - 10 * 19.5^2)$ =29.16667

By R:

```
> y <- c(20,12,15,16,18,14,22,28,24,26)
> (ybar <- mean(y))
[1] 19.5
> (Ssquare <- var(y))
[1] 29.16667
```

Exercise 2.12:

Five babies were born in a particular year in village Beonhin of Mathura district. The age (in years) of mothers at the time of child birth were 29, 32, 26, 28, and 36. Enumerate all possible **WR** equal probability **samples of size 2**, and show numerically that the sample mean age is an unbiased estimator of population mean age of the mothers.

Solution

Let's name the mothers (A, B, C, D, E) with weights (29, 32, 26, 28, 36) respectively.

$$\bar{Y} = \frac{29 + 32 + 26 + 28 + 36}{5} = \mathbf{30.2}$$

N= 5 n=2

The number of possible samples is $5^2 = 25$

Sample	Mothers in the sample	Sample mean $\bar{y}$
1	AA	29
2	AB	30.5
3	BA	30.5
4	AC	27.5
5	CA	27.5
6	AD	28.5
7	DA	28.5
8	AE	32.5
9	EA	32.5
10	BB	32
11	BC	29
12	CB	29
13	BD	30

Sample	Mothers in the sample	Sample mean $\bar{y}$
14	DB	30
15	BE	34
16	EB	34
17	CC	26
18	CD	27
19	DC	27
20	CE	31
21	EC	31
22	DD	28
23	DE	32
24	ED	32
25	EE	36

Distribution of the sample mean:

Sample mean $\bar{y}$	Frequency (f)	Probability (p)
29	3	0.12
30.5	2	0.08
27.5	2	0.08
28.5	2	0.08
32.5	2	0.08
32	3	0.12
30	2	0.08
34	2	0.08
26	1	0.04
27	2	0.08
31	2	0.08
28	1	0.04
36	1	0.04
Total	25	1

$$E(\bar{y}) = \frac{1}{25}(29 + 30.5 + 30.5 + \cdots + 32 + 36) = 30.2$$

$$Since\ E(\bar{y}) = \bar{Y}\ , then\ the\ sample\ mean\ \bar{y}\ is\ unbaised\ of\ the\ population\ mean\ \bar{Y}$$

By R:

```
> install.packages("tidyverse")
> library(tidyverse)
> Y <- c(29,32,26,28,36)
> Ybar <- mean(Y)
[1] 30.2
> sample <- crossing (Var1=Y, Var2=Y)
> samples<- sample (Y, 2, replace= TRUE)
> sample
# A tibble: 25 x 2
   Var1  Var2
  <dbl> <dbl>
 1    26    26
 2    26    28
 3    26    29
 4    26    32
 5    26    36
 6    28    26
 7    28    28
 8    28    29
 9    28    32
10    28    36
# ... with 15 more rows

#to view as a table
> view(sample)

> (ybars= apply(sample,1 , mean))
```

```
[1] 26.0 27.0 27.5 29.0 31.0 27.0 28.0 28.5 30.0
[10] 32.0 27.5 28.5 29.0 30.5 32.5 29.0 30.0 30.5
[19] 32.0 34.0 31.0 32.0 32.5 34.0 36.0
> (unbiased=mean(ybars))
[1] 30.2
```

Let's take an example of size three:

By R:

```
> Y <- c(29,32,26,28,36)
> Ybar <- mean(Y)
[1] 30.2
> sample <- crossing (Var1=Y, Var2=Y, Var3=Y)
> samples<- sample (Y, 3, replace= TRUE)
> sample     #No. of observation = 5^3 = 125
# A tibble: 125 x 3
   Var1  Var2  Var3
  <dbl> <dbl> <dbl>
 1   26    26    26
 2   26    26    28
 3   26    26    29
 4   26    26    32
 5   26    26    36
 6   26    28    26
 7   26    28    28
 8   26    28    29
 9   26    28    32
10   26    28    36
# ... with 115 more rows

#to view as a table
> view(sample)
```

```
> (ybars= apply(sample,1 , mean))
  [1] 26.00000 26.66667 27.00000 28.00000 29.33333
  [6] 26.66667 27.33333 27.66667 28.66667 30.00000
 [11] 27.00000 27.66667 28.00000 29.00000 30.33333
 [16] 28.00000 28.66667 29.00000 30.00000 31.33333
 [21] 29.33333 30.00000 30.33333 31.33333 32.66667
 [26] 26.66667 27.33333 27.66667 28.66667 30.00000
 [31] 27.33333 28.00000 28.33333 29.33333 30.66667
 [36] 27.66667 28.33333 28.66667 29.66667 31.00000
 [41] 28.66667 29.33333 29.66667 30.66667 32.00000
 [46] 30.00000 30.66667 31.00000 32.00000 33.33333
 [51] 27.00000 27.66667 28.00000 29.00000 30.33333
 [56] 27.66667 28.33333 28.66667 29.66667 31.00000
 [61] 28.00000 28.66667 29.00000 30.00000 31.33333
 [66] 29.00000 29.66667 30.00000 31.00000 32.33333
 [71] 30.33333 31.00000 31.33333 32.33333 33.66667
 [76] 28.00000 28.66667 29.00000 30.00000 31.33333
 [81] 28.66667 29.33333 29.66667 30.66667 32.00000
 [86] 29.00000 29.66667 30.00000 31.00000 32.33333
 [91] 30.00000 30.66667 31.00000 32.00000 33.33333
 [96] 31.33333 32.00000 32.33333 33.33333 34.66667
[101] 29.33333 30.00000 30.33333 31.33333 32.66667
[106] 30.00000 30.66667 31.00000 32.00000 33.33333
[111] 30.33333 31.00000 31.33333 32.33333 33.66667
[116] 31.33333 32.00000 32.33333 33.33333 34.66667
[121] 32.66667 33.33333 33.66667 34.66667 36.00000
> (unbiased=mean(ybars))
[1] 30.2
```

Exercise 2.22:

Distinguish between sampling and nonsampling errors. Which of these errors are more likely to be present in a census or a sample survey?

Solution

- **Sampling error** is the resultant discrepancy between the sample estimate and the population parameter value is the error of the estimate. Conversely, **non-sampling error** is arising due to defective sampling procedures, ambiguity in definitions, faulty measurement techniques, mistakes in recording, errors in coding-decoding, tabulation and analysis, etc…

- The sampling errors usually decreases with increase in sample size. In contrary, the nonsampling errors are likely to increase with increase in sample size. It is quite possible that nonsampling errors in a complete enumeration survey are greater than both the sampling and nonsampling errors taken together in a sample survey.