University of Waterloo

STAT 332 – W. H. Cherry

Tutorial 2 Outline Solution

T2 – 1. (a) We first list the six study population responses in order of *in*creasing magnitude:

5, 7, 9, 15, 17, 19;

the population average is $\bar{\mathbf{Y}} = 12$ and the square of the population (data) standard deviation is $\mathbf{S}^2 = 33.2 \simeq 5.762^2$.

We then tabulate the information required for part (a) of the question, and also for parts (b) and (e), for the $\binom{6}{2} = 15$ possible samples of size 2, under equiprobable selecting (EPS), as follows:

Sample	(5,7)	(5,9)	(5,15)	(5,17)	(5,19)	(7,9)	(7,15)	(7,17)	(7,19)	(9,15)	(9,17)	(9,19)	(15,17)	(15,19)	(17,19)
Average ($\bar{y}$)	6	7	10	11	12	8	11	12	13	12	13	14	16	17	18
Probability	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$
s^2	2	8	50	72	98	2	32	50	72	18	32	50	2	8	2 .

(b) The value ($\bar{y}$) of each sample average in the second line of the table above and the corresponding probabilities in the third line [which are all equal (of course) under *equi*probable selecting] are a (discrete) probabiliuty function.

(c) Using the probability function in the table above, the mean of $\bar{Y}$ is $E(\bar{Y}) = \sum_{\bar{y}} \bar{y} \bullet f(\bar{y}) = 12$,
and the standard deviation of $\bar{Y}$ is $s.d.(\bar{Y}) = \sqrt{E(\bar{Y}^2) - [E(\bar{Y})]^2} \simeq 3.32666$ $\quad [E(\bar{Y}^2) = \sum_{\bar{y}} \bar{y}^2 \bullet f(\bar{y}) = 155.0\dot{6}]$.

(d) It is easily verified that the values obtained in (c) *are*, respectively, $\bar{\mathbf{Y}}$, the population average and $\mathbf{S}\sqrt{\frac{1}{\mathrm{n}} - \frac{1}{\mathbf{N}}}$,
the standard deviation of the sample average under EPS, $s.d.(\bar{Y})$.

(e) For each sample, the square of the sample (data) standard deviation (s), as defined at the right, is shown in the fourth line of the table above in (a). Then, using the probability function in the fourth and third lines of the table, we find that the mean of S^2 is:

$$s^2 = \frac{1}{\mathrm{n}-1}\sum_{j=1}^{\mathrm{n}}(y_j - \bar{y})^2$$

$E(S^2) = \sum_{s^2} s^2 \bullet f(s^2) = 33.2 \simeq 5.762^2$, which *is* the value of $\mathbf{S}^2$ for this study population.

NOTE: Calculating the mean and (probabilistic) standard deviation of a random variable in part (c) *differs* from calculating an average and (data) standard deviation in part (e). [Similarities and differences are discussed in Section 5 on pages HL94.6 to HL94.8 in Statistical Highlight #94.]

T2 – 2. (a) One way to select the 50 buses from the 4,000 is as follows:

- label the buses 0000 through 3999;
- divide the digits in the table of equiprobable digits into groups of 4 and, with a suitable haphazard point of entry into the table, use these successively as follows, until 50 different bus labels have been obtained:
  - if the 4 digits yield a number from 0000 to 3999 inclusive, select the bus with this label;
  - if the 4 digits yield a number from 4000 to 7999 inclusive, subtract 4000 and select the bus with this label;
  - if the 4 digits yield a number from 8000 to 9999 inclusive, discard the number.

(b) We have five 8-hour days, which is 40 hours or 2,400 minutes, in the frame.

As in (a), we could use successive groups of 4 digits from the table until 100 different minutes have been selected:

- label the minutes 0000 through 2399;
  - if the 4 digits are a number k from 0000 to 9599 inclusive, select the minute with label k (mod 2,400);
  - if the 4 digits yield a number from 9600 to 9999 inclusive, discard this number.

***T2 – 3.** The study population consists of $\mathbf{N} = 9$ elements with the following values for a quantitative characteristic of interest:

9, 17, 5, 10, 8, 18, 15, 19, 7 [5, 7, 8, 9, 10, 15, 17, 18, 19 in order of increasing magnitude];

the population average is $\bar{\mathbf{Y}} = 12$ and the square of the population (data) standard deviation is $\mathbf{S}^2 = 27.75 \simeq 5.268^2$.

(a) Based on two units obtained by EPS from the population:
the mean of $\bar{Y}$ is $E(\bar{Y}) = \bar{\mathbf{Y}} = 12$; the standard deviation is $s.d.(\bar{Y}) = \mathbf{S}\sqrt{\frac{1}{\mathrm{n}} - \frac{1}{\mathbf{N}}} \simeq 3.285$.

(b) The sample average is $\bar{y} = 14$.

The value of this sample average is *not* equal to $\bar{\mathbf{Y}}$, the population average; in fact, for only three of the $\binom{9}{2} = 36$ possible samples of size 2 is $\bar{y}$ actually equal to 12 in this situation. However, the *average* of $\bar{y}$ over all (equiprobable) samples *is* $\bar{\mathbf{Y}}$, and this is what is meant by the statement that $\bar{Y}$, the random variable representing the sample average, is an unbiased estimator of the population average under equiprobable selecting.

(c) The square of the sample standard deviation is $s^2 = 50 \simeq 7.071^2$.

The value of this sample (data) standard deviation squared is *not* equal to $\mathbf{S}^2$, the square of the population (data) standard deviation; in fact, *none* of the 36 possible samples has exactly this value for its squared (data) standard deviation – the closest is (8, 15) for which $s^2 = 24.5$. However, the *average* of s^2 over all samples under EPS *is* $\mathbf{S}^2$, and this is what is meant by the statement that S^2, the random variable representing the square of the sample (data) standard deviation, is an unbiased estimator of the square of the population (data) standard deviation under EPS.

*Bonus question

Blank page