Question 1: (Ex 1)

Solution:

a. (2)

The value of $\bar{y}_U = \dfrac{(y_1 + y_2 + y_3 + y_4 + y_5 + y_6)}{6} = \dfrac{(98+102+154+133+190+175)}{6} = 142$

b. (8)

For plan 1, first compute the estimate from each of the 8 samples:

Sample Number	Sampled Units	Prob	Estimator	estimate
1	1,3,5	1/8	$(y_1 + y_3 + y_5)/3$	442/3
2	1,3,6	1/8	$(y_1 + y_3 + y_6)/3$	427/3
3	1,4,5	1/8	$(y_1 + y_4 + y_5)/3$	421/3
4	1,4,6	1/8	$(y_1 + y_4 + y_6)/3$	406/3
5	2,3,5	1/8	$(y_2 + y_3 + y_5)/3$	446/3
6	2,3,6	1/8	$(y_2 + y_3 + y_6)/3$	431/3
7	2,4,5	1/8	$(y_2 + y_4 + y_5)/3$	425/3
8	2,4,6	1/8	$(y_2 + y_4 + y_6)/3$	410/3

Then, we can compute the following values:

$$E(\bar{y}) = \frac{1}{8}\frac{442}{3} + \frac{1}{8}\frac{427}{3} + \frac{1}{8}\frac{421}{3} + \frac{1}{8}\frac{406}{3} + \frac{1}{8}\frac{446}{3} + \frac{1}{8}\frac{431}{3} + \frac{1}{8}\frac{425}{3} + \frac{1}{8}\frac{410}{3} = 142$$

$$\begin{aligned} V(\bar{y}) = {} & \frac{1}{8}\left(\frac{442}{3} - 142\right)^2 + \frac{1}{8}\left(\frac{427}{3} - 142\right)^2 + \frac{1}{8}\left(\frac{421}{3} - 142\right)^2 + \frac{1}{8}\left(\frac{406}{3} - 142\right)^2 \\ & + \frac{1}{8}\left(\frac{446}{3} - 142\right)^2 + \frac{1}{8}\left(\frac{431}{3} - 142\right)^2 + \frac{1}{8}\left(\frac{425}{3} - 142\right)^2 + \frac{1}{8}\left(\frac{410}{3} - 142\right)^2 \\ & = 18.94 \end{aligned}$$

$$Bias(\bar{y}) = E(\bar{y}) - \bar{y}_U = 142 - 142 = 0$$

$$MSE(\bar{y}) = V(\bar{y}) = 18.94$$

For plan 2, the estimates from the 3 samples are:

Sample Number	Sampled Units	Prob	Estimator	estimate
1	1,4,6	1/4	$(y_1 + y_4 + y_5)/3$	406/3
2	2,3,6	1/2	$(y_2 + y_3 + y_6)/3$	431/3
3	1,3,5	1/4	$(y_1 + y_3 + y_5)/3$	442/3

Then we have:

$$E(\bar{y}) = \frac{1}{4}\frac{406}{3} + \frac{1}{2}\frac{431}{3} + \frac{1}{4}\frac{442}{3} = 142.5$$

$$V(\bar{y}) = \frac{1}{4}\left(\frac{406}{3} - 142.5\right)^2 + \frac{1}{2}\left(\frac{431}{3} - 142.5\right)^2 + \frac{1}{4}\left(\frac{442}{3} - 142.5\right)^2 = 19.36$$

$$Bias(\bar{y}) = E(\bar{y}) - \bar{y}_U = 142.5 - 142 = 0.5$$

$$MSE(\bar{y}) = \frac{1}{4}\left(\frac{406}{3} - 142\right)^2 + \frac{1}{2}\left(\frac{431}{3} - 142\right)^2 + \frac{1}{4}\left(\frac{442}{3} - 142\right)^2 = 19.61$$

c. (2)

Plan 1 is better than Plan 2, since the estimator of Plan 1 is unbiased and the variance and mean square error are smaller than Plan 2.

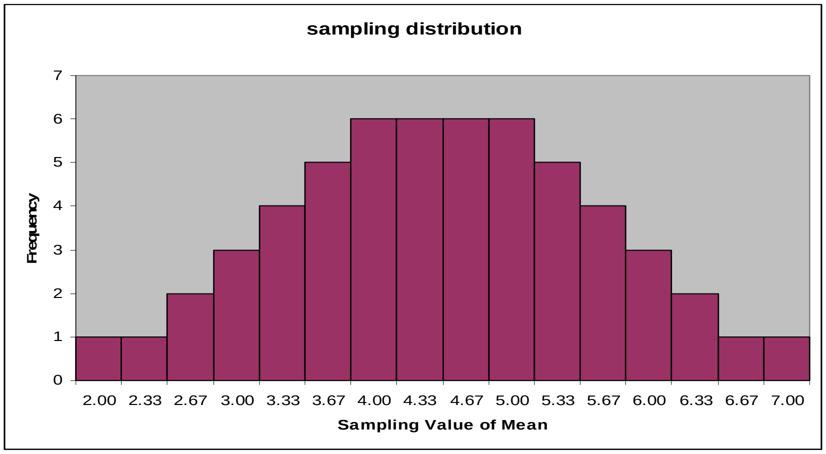
Question 2 (Ex 3)

Solution:

(a) (4)

For an SRS of size 3 without replacement, we will have totally $\binom{8}{3} = 56$ different samples. Their distribution is listed below

Sampling Mean	6/3	7/3	8/3	9/3	10/3	11/3	12/3	13/3	14/3	15/3	16/3	17/3	18/3	19/3	20/3	21/3
Frequency	1	1	2	3	4	5	6	6	6	6	5	4	3	2	1	1

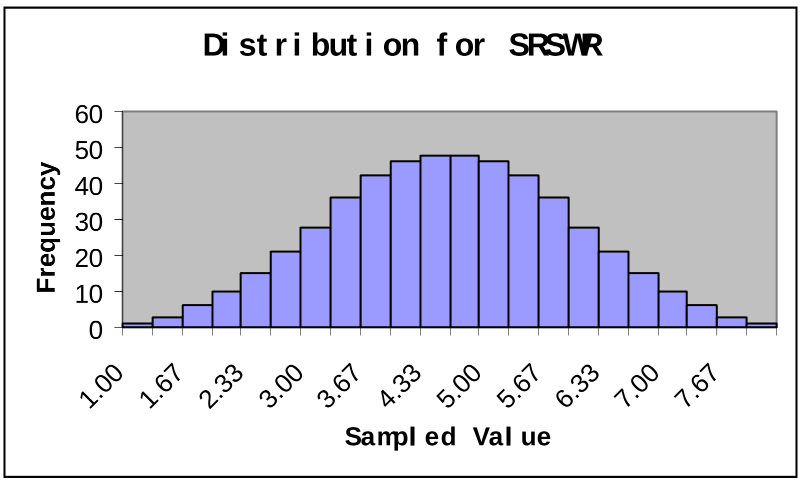


(b) (4)

For an SRS of size 3 with replacement, the total number of sample is: $8 \cdot 8 \cdot 8 = 512$. The frequency of all the possible value of total is listed below:

Sampling mean	3/3	4/3	5/3	6/3	7/3	8/3	9/3	10/3	11/3	12/3	13/3	14/3	15/3	16/3	17/3	18/3	19/3	20/3	21/3	22/3	23/3	24/3
Frequency	1	3	6	10	15	21	28	36	42	46	48	48	46	42	36	28	21	15	10	6	3	1

The histogram is displayed below:

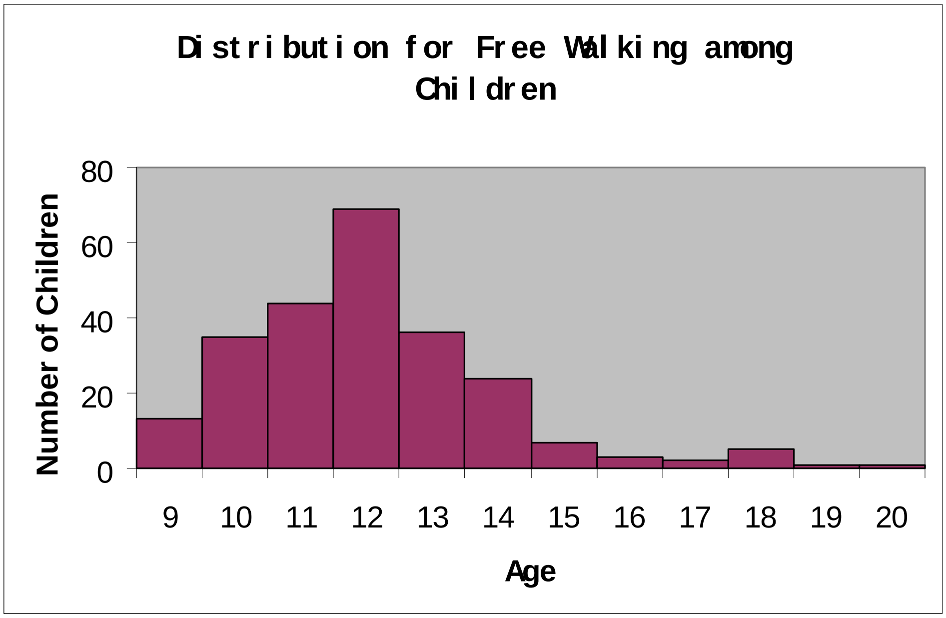


The sampling distribution of SRS without replacement has smaller variance. From the two histograms, we can see that the histogram for SRS with replacement is a little bit flatter than the histogram for SRS without replacement. (2)

Question 3 (Ex 5)

(a) (4)

The histogram is displayed below. We can use Excel to produce this graph.



The histogram shows the data is not normally distributed: it is not symmetry, it is skewed to the right. Yes the sampling distribution of the sample average will be normally distributed, since the sample size 240 is quite large, so by Central Limit Theorem, the sampled mean will be normally distributed.

(b) (4)

$$mean = \frac{9 \cdot 13 + 10 \cdot 35 + ... + 19 \cdot 1 + 20 \cdot 1}{13 + 35 + ... + 1 + 1} = \frac{2899}{240} = 12.08$$

The sample variance is:

$$s^2 = \frac{(9 - 12.08)^2 \cdot 13 + (10 - 12.08)^2 \cdot 35 + ... + (19 - 12.08)^2 \cdot 1 + (20 - 12.08)^2 \cdot 1}{13 + 35 + ... + 1 + 1 - 1}$$
$$= \frac{885.50}{239} = 3.705$$

Then, the variance of the mean estimate is estimated by:

$$\text{var} = \frac{s^2}{n} = \frac{3.705}{240} = 0.0154$$

Thus the estimate of the standard error of the sample mean is:

$$sd = \sqrt{\text{var}} = \sqrt{0.0154} = 0.1242$$

and the 95% Confidence Interval is:

$$mean \pm z_{0.025} \cdot sd = 12.08 \pm 1.96 \cdot 0.1242 = 12.08 \pm 0.24$$

or: $(11.84, \;\; 12.32)$

(c) (2)
From:

$$0.5 = z_{0.025} \cdot \frac{sd}{\sqrt{n}}$$

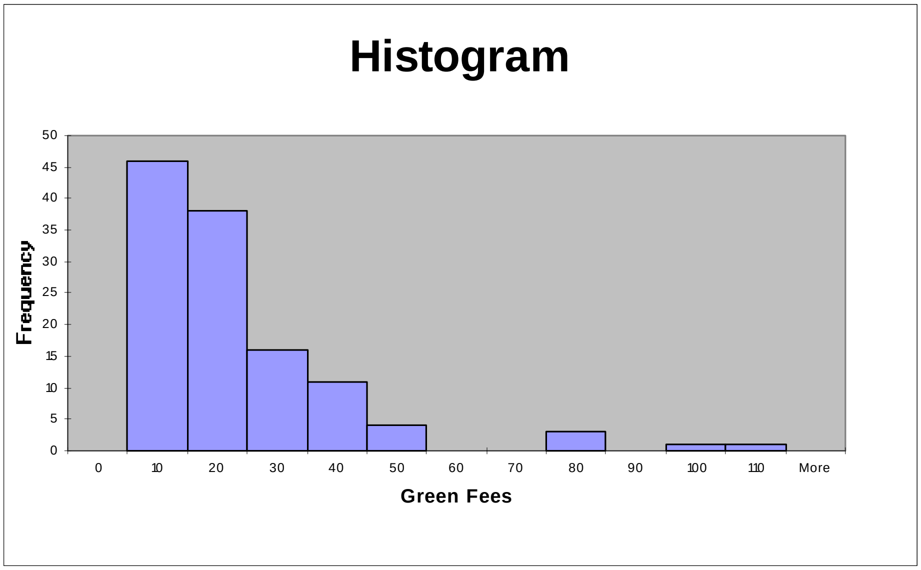
we have:

$$n = \left(\frac{z_{0.025} \cdot SD}{0.5}\right)^2 = \left(\frac{1.96 \cdot 1.92}{0.5}\right)^2 = 56.69 \approx 57$$

Thus the sample size should be 57

Question 4 (Ex 15)

(a) (5)
The histogram is displayed below:



The histogram shows that the data is not normally distributed, it is skewed to the right, with few golf courses have much more green fees.
(b) (5)
The average weekday greens fee to play nine holes of golf is:

$$\bar{y} = \frac{\sum_{i \in s} y_i}{n} = \frac{2418.4}{120} = 20.15$$

The standard error of the estimate can be computed as:

$$S_y^2 = \frac{\sum_{i \in s} (y_i - \bar{y})^2}{n-1} = 321.36$$

and the standard error is:

$$SE[\bar{y}] = \sqrt{\frac{S_y^2}{n}} = \sqrt{\frac{321.36}{120}} = 1.63$$

(Since the sampling rate is very small, we can simple ignore the term $1 - \frac{n}{N}$).

Question 5 (Ex 17)

Solution: (8)

There are totally 120 golf courses in the sample, of which, 85 are 18 holes, thus we have the estimate of the proportion as:

$$\hat{p} = \frac{85}{120} \approx 0.71$$

and the variance of this estimate can be computed as:

$$V(\hat{p}) = \left(1 - \frac{n}{N}\right) \frac{p(1-p)}{n-1} \xrightarrow{Estimated\ \ by} v(\hat{p}) = \left(1 - \frac{n}{N}\right) \frac{\hat{p}(1-\hat{p})}{n-1} = 0.001722$$

Thus the 95% CI is given by:

$$\hat{p} \pm 1.96\sqrt{v(\hat{p})} = 0.71 \pm 0.081$$

or:

$$p \in [0.63, 0.79]$$