Question 1

Part-a

$$\mu = \sum y \cdot p(y)$$

$$= 11\left(\frac{1}{7}\right) + 12\left(\frac{1}{7}\right) + 13\left(\frac{1}{7}\right) + 14\left(\frac{1}{7}\right) + 15\left(\frac{1}{7}\right) + 16\left(\frac{1}{7}\right) + 17\left(\frac{1}{7}\right)$$

= **14**

Part-b

$$\sigma^2 = \sum (y-\mu)^2 \cdot p(y)$$

$$= (11-14)^2\left(\frac{1}{7}\right) + (12-14)^2\left(\frac{1}{7}\right) + (13-14)^2\left(\frac{1}{7}\right) + (14-14)^2\left(\frac{1}{7}\right) +$$

$$+(15-14)^2\left(\frac{1}{7}\right) + (16-14)^2\left(\frac{1}{7}\right) + (17-14)^2\left(\frac{1}{7}\right)$$

= **4**

Part-c

```
R code to draw 2 samples:
set.seed(753) # set the seed for reproducibility
sample(11:17, 2, replace =FALSE)

# R output
[1] 16 12
```

c-1)

The sample mean is

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

= (16 + 12)/2

= **14**

c-2)

The sample variance is

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

$$= \frac{1}{2-1}[(16-14)^2 + (12-14)^2]$$

= **8**

c-3)

Calculation of $V[\bar{y}]$:

$$V[\bar{y}] = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

$$= \frac{4}{2}\left(\frac{7-2}{7-1}\right)$$

= **1.6667**

c-4)

Calculation of the estimated variance of $\bar{y}$:

$$\hat{V}[\bar{y}] = \left(1 - \frac{n}{N}\right)\frac{s^2}{n}$$

$$= \left(1 - \frac{2}{7}\right)\frac{8}{2}$$

= **2.8571**

Part-d

We already illustrated in the previous part how to calculate the sample mean and sample variance. In this part, we will show how to calculate the confidence interval for the population mean using the same sample {16, 12}

From the previous calculations, the sample mean is $\bar{y} = 14$ and the sample variance is $s^2 = 8$.

We also calculated from previous part the estimated variance

$$\hat{V}[\bar{y}] = \left(1 - \frac{n}{N}\right)\frac{s^2}{n} = 2.8571$$

Calculation of the "Bound on the error of estimation":

$$2\sqrt{\hat{V}[\bar{y}]} = 2\sqrt{\left(1 - \frac{n}{N}\right)\frac{s^2}{n}}$$

$$=2\sqrt{2.8571}$$

$$= 3.38$$

Hence, the approximate 95% confidence interval for the population mean is

$$\bar{y} \pm 2\sqrt{\hat{V}[\bar{y}]} = 14 \pm 3.38$$

$$= (10.62, 17.38)$$

We estimate that the population mean is somewhere between 10.62 and 17.38 with 95% confidence.

See table below on next page

Sample #	Sample Values		Sample Mean	Sample Variance	Estimated Variance of Sample Mean	Error Bound	95% Confidence Interval		Is the population mean 14 inside the interval?
1	11	12	11.5	0.5	0.178571	0.85	10.65	12.35	No
2	11	13	12	2	0.714286	1.69	10.31	13.69	No
3	11	14	12.5	4.5	1.607143	2.54	9.96	15.04	Yes
4	11	15	13	8	2.857143	3.38	9.62	16.38	Yes
5	11	16	13.5	12.5	4.464286	4.23	9.27	17.73	Yes
6	11	17	14	18	6.428571	5.07	8.93	19.07	Yes
7	12	13	12.5	0.5	0.178571	0.85	11.65	13.35	No
8	12	14	13	2	0.714286	1.69	11.31	14.69	Yes
9	12	15	13.5	4.5	1.607143	2.54	10.96	16.04	Yes
10	12	16	14	8	2.857143	3.38	10.62	17.38	Yes
11	12	17	14.5	12.5	4.464286	4.23	10.27	18.73	Yes
12	13	14	13.5	0.5	0.178571	0.85	12.65	14.35	Yes
13	13	15	14	2	0.714286	1.69	12.31	15.69	Yes
14	13	16	14.5	4.5	1.607143	2.54	11.96	17.04	Yes
15	13	17	15	8	2.857143	3.38	11.62	18.38	Yes
16	14	15	14.5	0.5	0.178571	0.85	13.65	15.35	Yes
17	14	16	15	2	0.714286	1.69	13.31	16.69	Yes
18	14	17	15.5	4.5	1.607143	2.54	12.96	18.04	Yes
19	15	16	15.5	0.5	0.178571	0.85	14.65	16.35	No
20	15	17	16	2	0.714286	1.69	14.31	17.69	No
21	16	17	16.5	0.5	0.178571	0.85	15.65	17.35	No

Only 15 out of 21 intervals or **71.43%** capture the population mean of 14 .
The expected coverage is 95%. The reason why the coverage is only 71.43% is that the sample mean is not approximately normal because the population data is not normally distributed and the sample size of 2 is too small.

Part-e
e-1)
The mean of the sample variances is calculated by adding the values in the “Sample Variance” column above and divide by 21. We got 98/21 = 4.6667 and it is higher than the actual population variance of 4.
e-2)
Comment: The sample variance overestimates the population variance, on the average when we sample without replacement. That’s why, we need a correction factor (N-1)/N = 6/7