Introductory Statistics for Economics

ECON1013: TUTORIAL 2

Duong Trinh

University of Glasgow

Spring 2024

Intro

- Duong Trinh
  - PhD Student in Economics (Bayesian Microeconometrics)
  - Email: Duong.Trinh@glasgow.ac.uk
- ECON1013-TU04
  - Monday 12-1 pm
  - 4 sessions (22-Jan, 5-Feb, 19-Feb, 4-March)
- ECON1013-TU05
  - Tuesday 12-1 pm
  - 4 sessions (23-Jan, 6-Feb, 20-Feb, 5-March)
- ECON1013-TU07
  - Tuesday 2-3 pm
  - 4 sessions (23-Jan, 6-Feb, 20-Feb, 5-March)

Record Attendance

Exercise 1

- Let X be the amount of exported grain next month (in tonnes) of a grain exporting company.
- Based on previous exporting records, the company estimates that the exporting probabilities and amounts.

x	P(x)
100	0.05
150	0.20
200	0.50
250	0.20
300	0.05

Exercise 1

(a) What is the probability that the company does not export more than 250 tonnes?

$P(X \leq 250) =?$

(b) What is the expected amount of tonnes that the company will export?

$E[X] =?$

(c) What is the standard deviation?

$\sqrt{Var[X]} =?$

(a) What is the probability that the company doesn't export more than 250 tonnes?

- From the cumulative probability distribution we see that the probability that the company does not export more than 250 tonnes is

$$P(X \leq 250) = F(250) = 0.95$$

x	P(x)	F(x)
100	0.05	0.05
150	0.20	0.25
200	0.50	0.75
250	0.20	0.95
300	0.05	1.00

(b) What is the expected amount of tonnes that the company will export?

$$
\begin{aligned}
E[X] = \mu &= \sum_x xP(x) \\
&= 100 \times 0.05 + 150 \times 0.20 + 200 \times 0.50 + 250 \times 0.20 + 300 \times 0.05 \\
&= 200
\end{aligned}
$$

(c) What is the standard deviation?

$$
\begin{aligned}
\sigma &= \sqrt{\sum_x (x-\mu)^2 P(x)} \\
&= \sqrt{(100-200)^2 \times 0.05 + (150-200)^2 \times 0.20 + \ldots + (300-200)^2 \times 0.05} \\
&= \sqrt{2000} \\
&\approx 44.72
\end{aligned}
$$

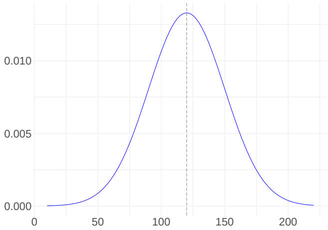
Exercise 1

- Let Y be the amount of exported salt (in tonnes) next month of the company.

- The contracts for salt allow for salt being sold in much smaller quantities, making it a continuous random variable.

$$Y \sim \mathcal{N}(120, 30^2)$$

$$\mu_Y = 120; \quad \sigma_Y = 30$$



Exercise 1

(d) What is the probability that the amount of exported salt next month will not be bigger than 160 tonnes?

$P(Y \leq 160) = ?$

(e) What is the probability that the amount of exported salt next month will be between 80 and 160 tonnes?

$P(80 \leq Y \leq 160) = ?$

(d) What is the probability that the amount of exported salt next month will not be bigger than 160 tonnes?

S1. Transform Y into the standard normal random variable Z

$$Z = \frac{Y - \mu}{\sigma} = \frac{Y - 120}{30}; \quad Z \sim \mathcal{N}(0, 1^2)$$

S2. Rewrite the probability by transforming 160 tonnes to Z units

$$P(Y \leq 160) = P\left(\frac{Y - 120}{30} \leq \frac{160 - 120}{30}\right) \approx P\left(Z \leq 1.33\right)$$

S3. Looked up from a standard normal distribution statistical table

$$P\left(Z \leq 1.33\right) = F(1.33) \approx 0.908$$

(d) What is the probability that the amount of exported salt next month will not be bigger than 160 tonnes?

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

(e) What is the probability that the amount of exported salt next month will be between 80 and 160 tonnes?

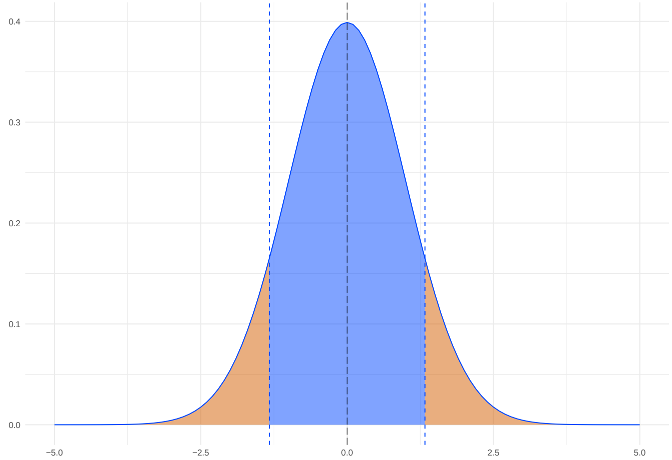
S1. Transform Y into the standard normal random variable Z

$$Z = \frac{Y - \mu}{\sigma} = \frac{Y - 120}{30}$$

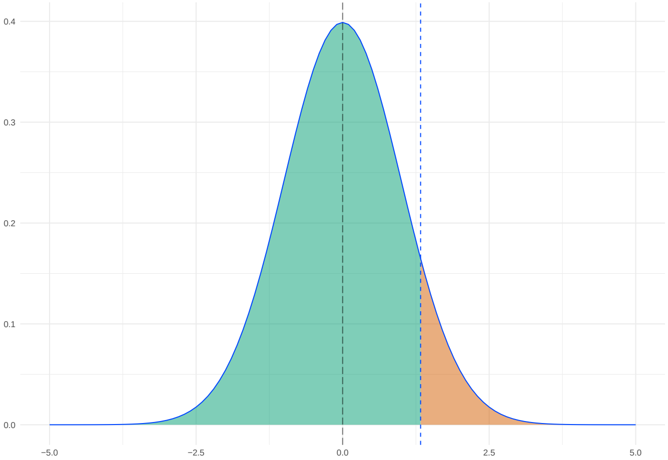
S2. Rewrite the probability

$$\begin{aligned}
P(80 \leq Y \leq 160) &= P\left(\frac{80 - 120}{30} \leq \frac{Y - 120}{30} \leq \frac{160 - 120}{30}\right) \\
&\approx P\left(-1.33 \leq Z \leq 1.33\right) \\
&= 1 - \left[P\left(Z \geq 1.33\right) + P\left(Z \leq -1.33\right)\right] \\
&= 1 - 2P\left(Z \geq 1.33\right) \\
&= 1 - 2[1 - P\left(Z \leq 1.33\right)] \\
&= 2F(1.33) \\
&\approx 2 \times 0.908 - 1 = 0.816
\end{aligned}$$

(e) What is the probability that the amount of exported salt next month will be between 80 and 160 tonnes?



(e) What is the probability that the amount of exported salt next month will be between 80 and 160 tonnes?



Exercise 2

- A population consists of all first-year primary school students in a specific neighborhood ($N = 500$).
- Administrative data on the number of years spent in nursery of the full population of students:
  - Population mean: $\mu = 3$
  - Population standard deviation: $\sigma = 1.6$
- Pick a random sample of $n = 47$ students.
  - Denote by $\bar{X}$ the random variable "sample mean" and by $\bar{x}$ its realization.

Exercise 2

(a) Explain what do we mean by "*the sampling distribution of the sample mean* ($\bar{X}$)"?

(b) Using your knowledge about sampling (without making any calculations), characterize the sampling distribution of $\bar{X}$.

(c) Calculate the *mean* and *standard deviation* of the sampling distribution.

(d) What values of $\bar{X}$ are likely? Would, for example, a sample mean $\bar{x} = 3.5$ be likely or unlikely?

(a) Explain what do we mean by "*the sampling distribution of the sample mean* ($\bar{X}$)"?

- It tells us how sample means are distributed across samples
- It tells us how likely are different possible values of $\bar{x}$ given the population parameters and the sample size.

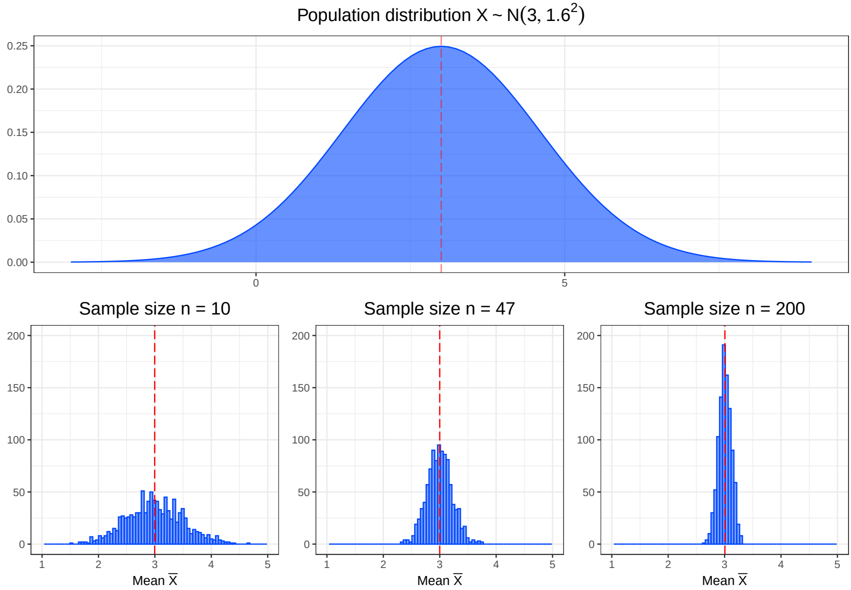
(b) Characterize the sampling distribution of $\bar{X}$.

Central Limit Theorem

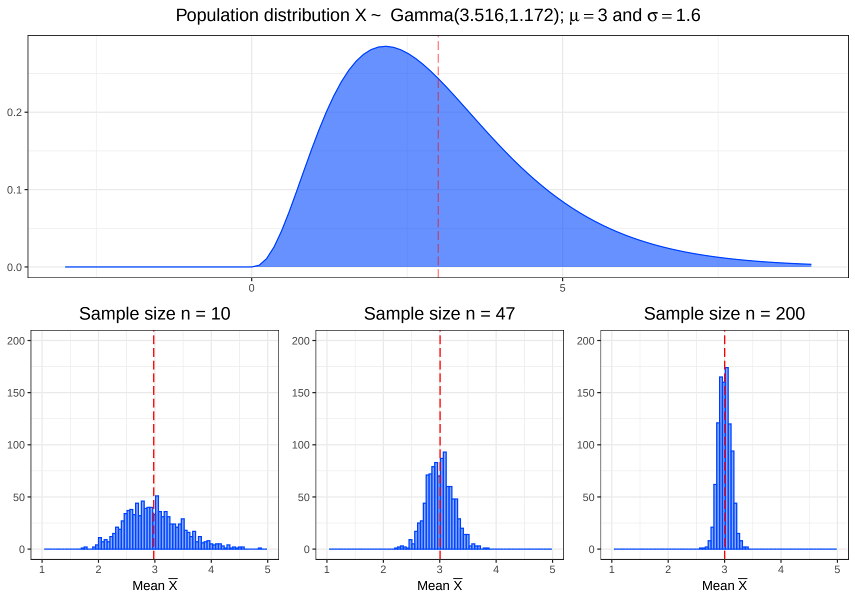
- In word: the sampling distribution of the sample mean is approximately normal when the sample size is large enough.
- In math:

$$\bar{X} \sim \mathcal{N}\left(\mu_{\bar{X}}, \sigma^2_{\bar{X}}\right)$$

Sample size $n = 47$ is sufficiently large (i.e. ≥ 30) to have an approximately normally distributed sampling distribution of the sample mean. This is true even if the population distribution of X is not normal.



Population distribution X ~ N(3, 1.6^2)
0.25
0.20
0.15
0.10
0.05
0.00
0
5
Sample size n = 10
Sample size n = 47
Sample size n = 200
200
150
100
50
0
1
2
3
4
5
Mean $\overline{X}$



Population distribution X ~ Gamma(3.516,1.172); μ = 3 and σ = 1.6
0.0
0.1
0.2
0
5
Sample size n = 10
Sample size n = 47
Sample size n = 200
0
50
100
150
200
1
2
3
4
5
Mean $\overline{X}$

(c) Calculate the *mean* and *standard deviation* of the sampling distribution.

- The mean of the sampling distribution for the sample mean will be the same as the population mean

$$\mu_{\bar{X}} = \mu = 3$$

- The standard deviation of the sampling distribution for the sample mean is given by

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1.6}{\sqrt{47}} \approx 0.23$$

(d) What values of $\bar{X}$ are likely? Would, for example, a sample mean $\bar{x} = 3.5$ be likely or unlikely?

- For random variable which follows a normal distribution, more than 95% of the probability mass is located with a distance of at most 2 standard deviations from the mean.
- In the case of the sample mean, we know that $\bar{X} \sim \mathcal{N}\left(\mu_{\bar{X}}, \sigma^2_{\bar{X}}\right)$, hence

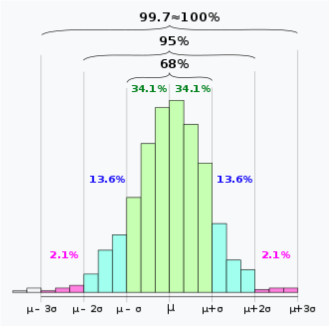
$$
\begin{aligned}
& P(\mu_{\bar{X}} - 2\sigma_{\bar{X}} \leq \bar{X} \leq \mu_{\bar{X}} + 2\sigma_{\bar{X}}) > 95\% \\
\Rightarrow \quad & P(3 - 2 \cdot 0.23 \leq \bar{X} \leq 3 + 2 \cdot 0.23) > 95\% \\
\Rightarrow \quad & P(2.54 \leq \bar{X} \leq 3.46) > 95\%
\end{aligned}
$$

- In other words, values of the sample mean that are below 2.54 or above 3.46 happen relatively rarely. On the other hand, a sample mean like 3.5 is not impossible.

68–95–99.7 rule

When $X \sim \mathcal{N}(\mu, \sigma^2)$,

$$\Pr(\mu - 1\sigma \leq X \leq \mu + 1\sigma) \approx 68.27\%$$
$$\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 95.45\%$$
$$\Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 99.73\%$$



Exercise 3

A population contains two million **zeros** an nine million **ones**.

- *Hint*: Make a table which contains all the possible values of the sample mean when $n = 4$ and their probabilities.
- *Hint*: Remember that when the data is binary (0/1), there is a link between the sample mean and the sample proportion.

(b) What is the approximate sampling distribution of the sample mean, when the sample size is $n = 50$?

(a) What is the sampling distribution of the sample mean, when the sample size is $n = 4$?

Drawing 4 observations from a population of zeros and ones, we have 5 different possible samples

- $\{0, 0, 0, 0\} \rightarrow$ sample mean $= 0$
- $\{1, 0, 0, 0\} \rightarrow$ sample mean $= 1/4$
- $\{1, 1, 0, 0\} \rightarrow$ sample mean $= 1/2$
- $\{1, 1, 1, 0\} \rightarrow$ sample mean $= 3/4$
- $\{1, 1, 1, 1\} \rightarrow$ sample mean $= 1$

(a) What is the sampling distribution of the sample mean, when the sample size is $n = 4$?

Drawing 4 observations from a population of zeros and ones, we have 5 different possible samples

- $\{0, 0, 0, 0\} \rightarrow$ sample mean $= 0$
- $\{1, 0, 0, 0\} \rightarrow$ sample mean $= 1/4$
- $\{1, 1, 0, 0\} \rightarrow$ sample mean $= 1/2$
- $\{1, 1, 1, 0\} \rightarrow$ sample mean $= 3/4$
- $\{1, 1, 1, 1\} \rightarrow$ sample mean $= 1$

- $\{Y_1, Y_2, Y_3, Y_4\} \mid Y_i = 0$ or $Y_i = 1$

  sample mean $= \frac{\sum_{i=1}^{4} Y_i}{4} = \frac{\text{number of 1s}}{\text{sample size}} =$ sample proportion $= \hat{p}$.

(a) What is the sampling distribution of the sample mean, when the sample size is $n = 4$?

Formula for the probability of x successes in n trials [Binomial distribution]

$$P(x) = \frac{n!}{x!\,(n-x)!}\, p^x (1-p)^{n-x}, \qquad \text{for } x = 1, \ldots, n$$

- ◇ Probability that the first x trials to be successes and the remaining $n - x$ to be failures

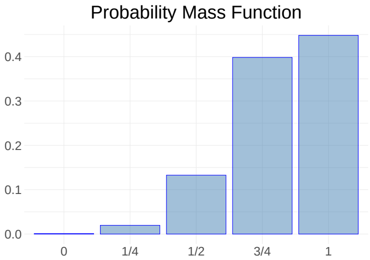
$$p \times p \times \ldots \times p \times (1-p) \times (1-p) \times \ldots \times (1-p) = p^x (1-p)^{n-x}$$

- ◇ The number of ways of arranging x successes in n trials is equal to the number of ways of choosing x objects from n objects

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

(a) When $n = 4$; $x = 0, 1, 2, 3, 4$; $p = 9/11$

x	P(x)	p_hat
0	0.0011	0
1	0.0197	1/4
2	0.1328	1/2
3	0.3983	3/4
4	0.4481	1



(b) What is the sampling distribution of the sample mean, when the sample size is $n = 50$?

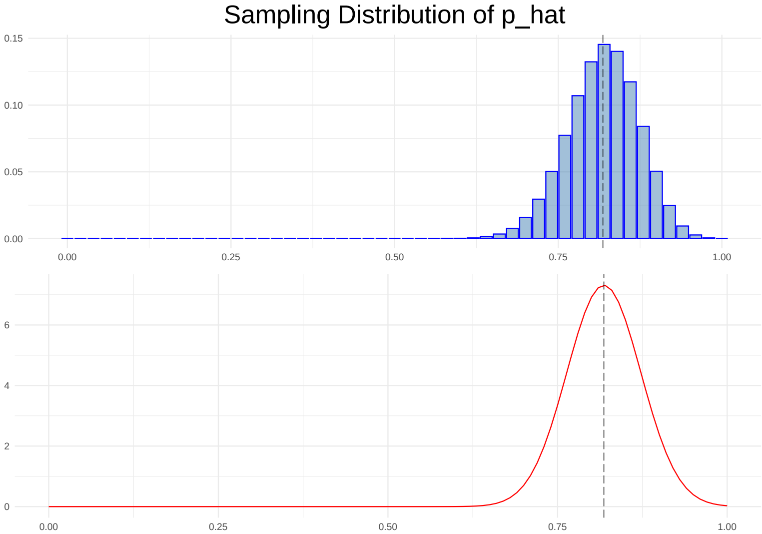
- Formula for the probability of x successes in n trials

$$P(x) = \frac{n!}{x!\,(n-x)!}\, p^x (1-p)^{n-x}, \qquad \text{for } x = 1, \ldots, n$$

- Making table is complicated!

x	P(x)	p_hat
0	...	0
1	...	1/50
2	...	2/50
⋮	⋮	⋮
50	...	50/50

(b) When $n = 50$; $x = 0, 1, \ldots, 50$; $p = 9/11$



Remark: Normal approximation for Binomial distribution

- Verify that n is large is enough

$$n \cdot p \cdot (1-p) > 5,$$

which is correct as $50 \cdot (9/11) \cdot (2/11) \approx 7.44$.

- Hence, asymptotically,

$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

where $\mu_{\hat{p}} = p = 9/11$,
and $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(9/11)\cdot(2/11)}{50}} \approx 0.055$.

Exercise 4

A small company has six employees, whose years of experience are

$$\{2 \quad 4 \quad 6 \quad 6 \quad 7 \quad 8\}.$$

Two of these employees are to be chosen randomly for a particular work group.

Let us consider the average number of years of experience of the two employees $\{X_1, X_2\}$ chosen randomly from the population of six. We denote by $\bar{X}_2$ the random variable sample means and by $\bar{x}_2$ its realization.

Exercise 4

(a) Compute the population mean μ_X and the population variance σ_X^2.

(b) Compute the sampling distribution of $\bar{X}_2$.

(c) Compute the expectation of $\bar{X}_2$ using the above sampling distribution and compare it with μ_X.

(d) Compute the variance of $\bar{X}_2$ using the sampling distribution computed above. Compare your result with $\sigma_X^2/2$, which would compute the variance of $\bar{X}_2$ if $\{X_1, X_2\}$ was a simple random sample.

Exercise 4

The variance σ_X^2 is unknown. Let us consider the sample variance of the $\{X_1, X_2\}$, denoted by s_2^2.

(e) Compute the sampling distribution of s_2^2.

(f) Compute the expectation of s_2^2 using the above sampling distribution and compare it with σ_X^2.

(g) Verify that $E(s_2^2) = N\sigma^2/(N-1)$, where $N = 6$.

See also Tables 6.1 and 6.2 in the Textbook, and Exercise 6.56

Exercise 4

The problem suggests that the sampling is without replacement (the work group should be formed by two persons). There are different 15 work groups, with ages

$$
\begin{array}{c|cccccc}
 & 2 & 4 & 6 & 6 & 7 & 8 \\
\hline
2 & -- & (2,4) & (2,6) & (2,6) & (2,7) & (2,8) \\
4 & -- & -- & (4,6) & (4,6) & (4,7) & (4,8) \\
\vdots & -- & -- & -- & \ddots & \dots & \vdots
\end{array}
$$

We consider only the element above the main diagonal, because the matrix is symmetric and on the main diagonal we would have work groups made of one person.

(a) Compute the population mean μ_X and the population variance σ_X^2.

$$\mu_X = \frac{2+4+6+6+7+8}{6} = 5.5$$

and

$$\sigma_X^2 = \frac{(2-\mu_X)^2 + \cdots (8-\mu_x)^2}{6} = 23.5/6 = 3.916667.$$

.

(b) Compute the sampling distribution of $\bar{X}_2$.

$\bar{x}_2$	3	4	4.5	5	5.5	6	6.5	7	7.5
$p(\bar{x}_2)$	1/15	2/15	1/15	3/15	1/15	2/15	2/15	2/15	1/15

(c) Compute the expectation of $\bar{X}_2$ using the above sampling distribution and compare it with μ_X.

$$\sum_{j=1}^{9} \bar{x}_{2,j}\, p(\bar{x}_{2,j}) = 5.5$$

◇ The estimator is unbiased.

(d) Compute the variance of $\bar{X}_2$ using the sampling distribution computed above. Compare your result with $\sigma_X^2/2$, which would compute the variance of $\bar{X}_2$ if $\{X_1, X_2\}$ was a simple random sample.

$$\sum_{j=1}^{9}(\bar{x}_{2,j} - \mu_X)^2 p(\bar{x}_{2,j}) = \frac{23.5}{15} = 1.566667$$

- ◇ Note that we should not apply the formula $\sigma_X^2/2 = 23.5/12$ but we should apply the finite population correction factor.

$$\frac{\sigma_X^2}{2}\frac{N-n}{N-1} = \frac{23.5}{12}\frac{4}{5} = \frac{23.5}{15}$$

(e) Compute the sampling distribution of s_2^2.

s_2^2	0.0	0.5	2.0	4.5	8	12.5	18
$p(s_2^2)$	1/15	3/15	5/15	1/15	3/15	1/15	1/15

(f) Compute the expectation of s_2^2 using the above sampling distribution and compare it with σ_X^2.

$$\sum_{j=1}^{7} s_{2,j}^2 p(s_{2,j}^2) = 5.5$$

- The estimator is biased.

(g) Verify that $E(s_2^2) = N\sigma^2/(N-1)$, where $N = 6$.

Multiplying the number above by $5/6$ we get σ_X^2.