Formative Assessment 5

SINOCRUZ, A. & TAGAYTAY, G.

2025-10-06

GITHUB LINK: https://github.com/eivra-sm/APM1111/blob/main/FA5.md

Problem 8.18

Population values (from Problem 8.17)

```
population <- c(9, 12, 15)
```

Generate all possible samples of size 2 with replacement

```
samples <- expand.grid(A = population, B = population)
```

Compute the sample mean for each combination

```
samples$mean <- rowMeans(samples)
```

Calculate probabilities (since population is uniform, each has equal prob = 1/9)

```
samples$p_xbar <- 1 / nrow(samples)
```

Compute xbar * p(xbar) and xbar^2 * p(xbar)

```
samples$xbar_p <- samples$mean * samples$p_xbar
samples$xbar2_p <- (samples$mean^2) * samples$p_xbar
```

Display table

```r
samples
```

```
##    A  B mean    p_xbar   xbar_p xbar2_p
## 1  9  9  9.0 0.1111111 1.000000    9.00
## 2 12  9 10.5 0.1111111 1.166667   12.25
## 3 15  9 12.0 0.1111111 1.333333   16.00
## 4  9 12 10.5 0.1111111 1.166667   12.25
## 5 12 12 12.0 0.1111111 1.333333   16.00
## 6 15 12 13.5 0.1111111 1.500000   20.25
## 7  9 15 12.0 0.1111111 1.333333   16.00
## 8 12 15 13.5 0.1111111 1.500000   20.25
## 9 15 15 15.0 0.1111111 1.666667   25.00
```

Compute mean of sampling distribution

```r
mu_xbar <- sum(samples$xbar_p)
```

Compute variance of sampling distribution

```r
sigma2_xbar <- sum(samples$xbar2_p) - mu_xbar^2
```

Compare with population mean and variance

```r
mu <- mean(population)
sigma2 <- var(population)
```

Show results

```r
cat("Population mean (mu):", mu, "\n")
```

```
## Population mean (mu): 12
```

```r
cat("Sampling mean (muxbar):", mu_xbar, "\n")
```

```
## Sampling mean (muxbar): 12
```

```r
cat("Population variance (sigma^2):", sigma2, "\n")
```

```
## Population variance (sigma^2): 9
```

```r
cat("Sampling variance (sigmaxbar^2):", sigma2_xbar, "\n")
```

```
## Sampling variance (sigmaxbar^2): 3
```

```r
cat("sigmaxbar^2  sigma^2 / 2:", sigma2 / 2, "\n")
```

```
## sigmaxbar^2  sigma^2 / 2: 4.5
```

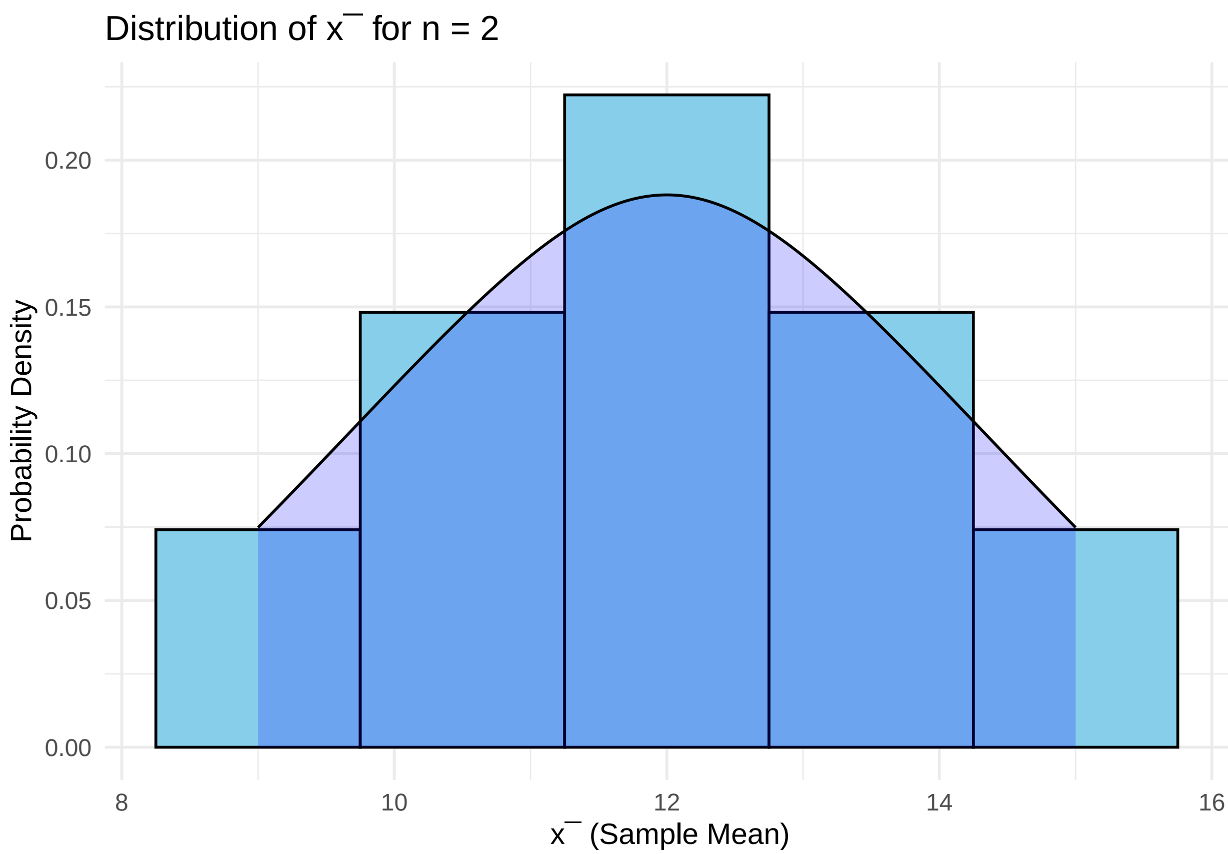
Plot sampling distribution

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.5.1
```

```r
ggplot(samples, aes(x = mean)) +
  geom_histogram(aes(y = ..density..), bins = 5, fill = "skyblue", color = "black") +
  geom_density(alpha = 0.2, fill = "blue") +
  labs(title = "Distribution of x̄ for n = 2",
       x = "x̄ (Sample Mean)",
       y = "Probability Density") +
  theme_minimal()
```

```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Problem 8.21

(a). Population mean and variance

```
# Population
population <- c(3, 7, 11, 15)

# Population mean ( )
mu <- mean(population)

# Population variance ( ²) and standard deviation ( )
sigma2 <- var(population)  # sample variance, same as population variance formula for this exercise
sigma <- sqrt(sigma2)

cat("Population mean ( ):", mu, "\n")
```

```
## Population mean ( ): 9
```

```r
cat("Population variance ( ²):", sigma2, "\n")
```

```
## Population variance ( ²): 26.66667
```

```r
cat("Population standard deviation ( ):", sigma, "\n")
```

```
## Population standard deviation ( ): 5.163978
```

(b). Generate all possible samples of size 2 (with replacement)

```r
# All combinations of 3, 7, 11, 15 with replacement
samples <- expand.grid(A = population, B = population)

# Compute sample means
samples$mean <- rowMeans(samples)

# Each sample equally likely → probability = 1 / (number of samples)
samples$p_xbar <- 1 / nrow(samples)

# Compute x̄ * p(x̄) and x̄² * p(x̄)
samples$xbar_p <- samples$mean * samples$p_xbar
samples$xbar2_p <- (samples$mean^2) * samples$p_xbar

# Show table
samples
```

```
##     A  B mean p_xbar xbar_p xbar2_p
## 1   3  3    3 0.0625 0.1875  0.5625
## 2   7  3    5 0.0625 0.3125  1.5625
## 3  11  3    7 0.0625 0.4375  3.0625
## 4  15  3    9 0.0625 0.5625  5.0625
## 5   3  7    5 0.0625 0.3125  1.5625
## 6   7  7    7 0.0625 0.4375  3.0625
## 7  11  7    9 0.0625 0.5625  5.0625
## 8  15  7   11 0.0625 0.6875  7.5625
## 9   3 11    7 0.0625 0.4375  3.0625
## 10  7 11    9 0.0625 0.5625  5.0625
## 11 11 11   11 0.0625 0.6875  7.5625
## 12 15 11   13 0.0625 0.8125 10.5625
## 13  3 15    9 0.0625 0.5625  5.0625
## 14  7 15   11 0.0625 0.6875  7.5625
## 15 11 15   13 0.0625 0.8125 10.5625
## 16 15 15   15 0.0625 0.9375 14.0625
```

(c). Mean of the sampling distribution of means (̄)

```r
mu_xbar <- sum(samples$xbar_p)
cat("Mean of sampling distribution ( ̄):", mu_xbar, "\n")
```

```
## Mean of sampling distribution ( ̄): 9
```

(d). Standard deviation of the sampling distribution of means (¯)

```
# Variance of sampling distribution
sigma2_xbar <- sum(samples$xbar2_p) - mu_xbar^2
sigma_xbar <- sqrt(sigma2_xbar)

cat("Variance of sampling distribution (¯²):", sigma2_xbar, "\n")
```

```
## Variance of sampling distribution (¯²): 10
```

```
cat("Standard deviation of sampling distribution (¯):", sigma_xbar, "\n")
```

```
## Standard deviation of sampling distribution (¯): 3.162278
```

(e). Verification: ¯ = / √n

```
n <- 2
sigma_xbar_formula <- sigma / sqrt(n)

cat("¯ (computed):", sigma_xbar, "\n")
```

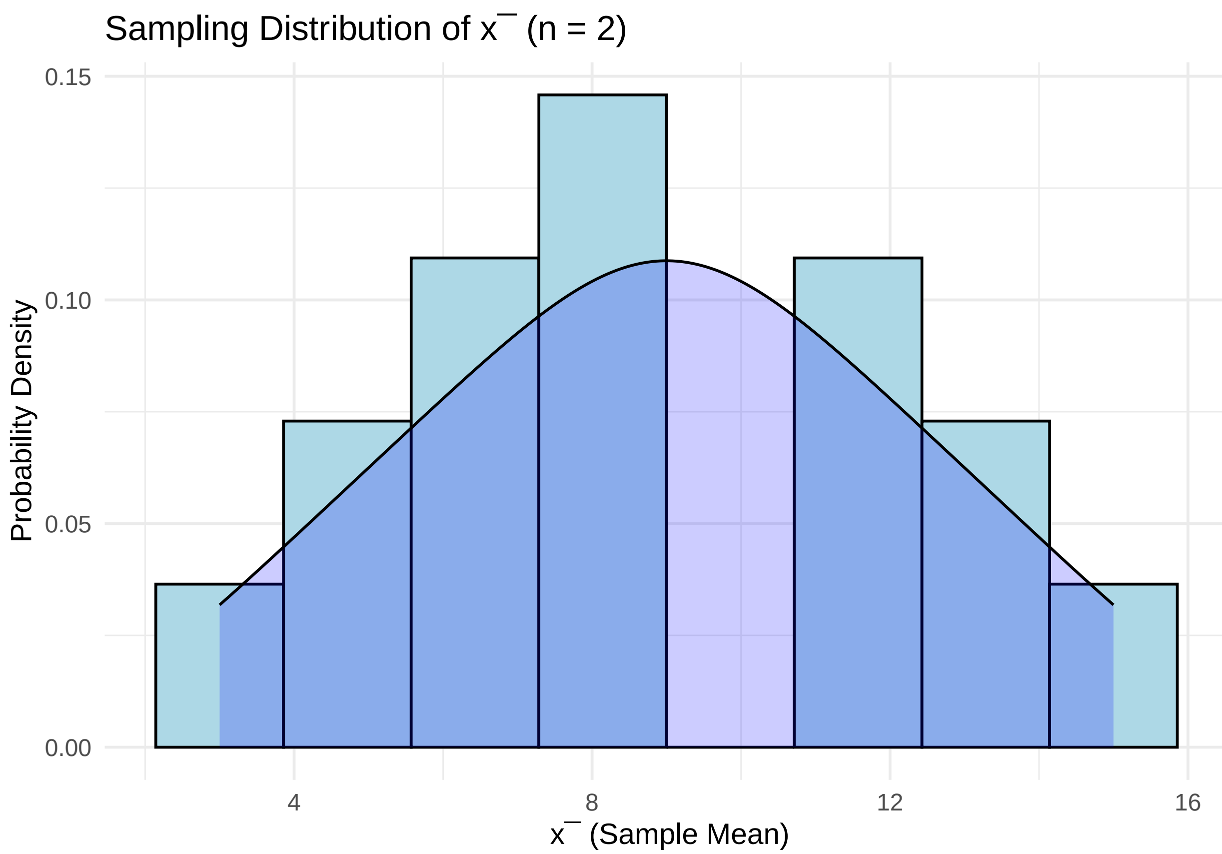
```
## ¯ (computed): 3.162278
```

```
cat(" / √n (formula):", sigma_xbar_formula, "\n")
```

```
##  / √n (formula): 3.651484
```

(f). Plot the sampling distribution of the mean

```
library(ggplot2)
ggplot(samples, aes(x = mean)) +
  geom_histogram(aes(y = ..density..), bins = 8, fill = "lightblue", color = "black") +
  geom_density(alpha = 0.2, fill = "blue") +
  labs(title = "Sampling Distribution of x̄ (n = 2)",
       x = "x̄ (Sample Mean)",
       y = "Probability Density") +
  theme_minimal()
```



Problem 8.34

```
# Given values
n <- 200
p <- 0.5
q <- 1 - p

# Standard deviation
sd_p <- sqrt(p * q / n)

# Probabilities
prob_a <- pnorm(0.40, mean = p, sd = sd_p)
prob_b <- pnorm(0.57, mean = p, sd = sd_p) - pnorm(0.43, mean = p, sd = sd_p)
prob_c <- 1 - pnorm(0.54, mean = p, sd = sd_p)

cat("$$ \\text{Standard deviation } (\\sigma_{\\hat{p}}) = ", round(sd_p, 4), " $$\n\n")
```

$$\text{Standard deviation } (\sigma_{\hat{p}}) = 0.0354$$

```
cat("$$ P(\\hat{p} < 0.40) = ", round(prob_a, 4), " $$\n")
```

$$P(\hat{p} < 0.40) = 0.0023$$

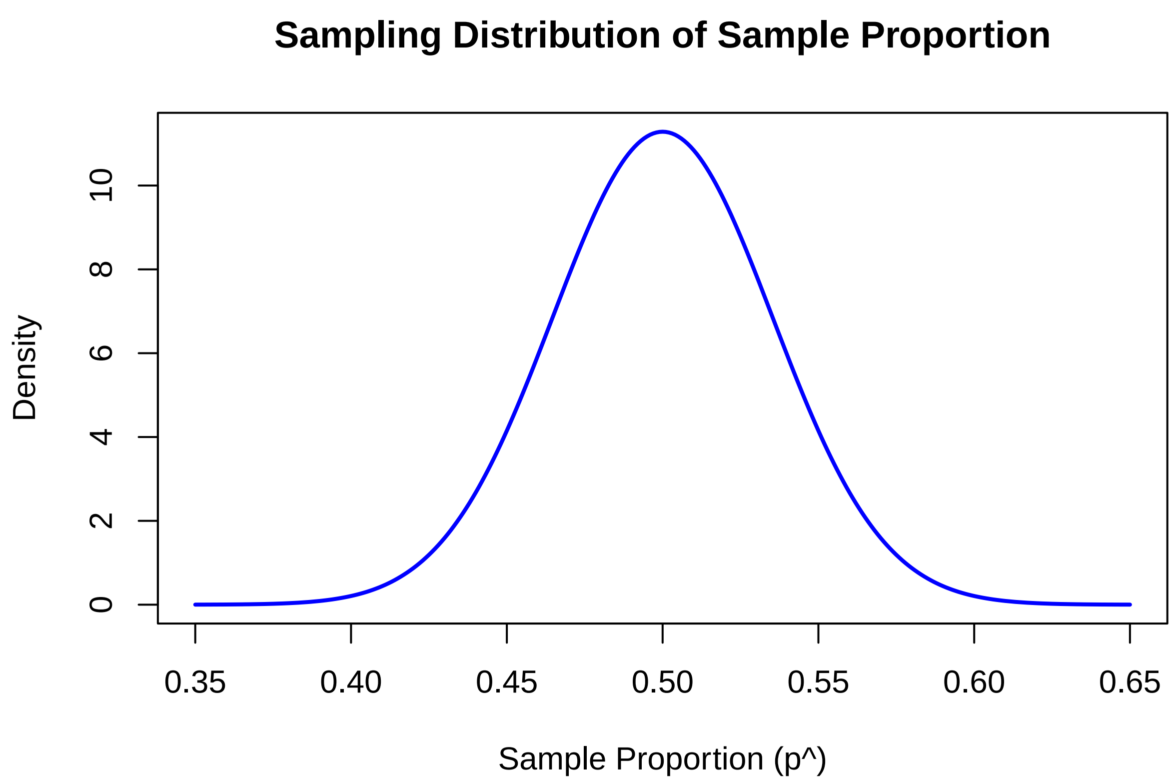
```
cat("$$ P(0.43 < \\hat{p} < 0.57) = ", round(prob_b, 4), " $$\n")
```

$$P(0.43 < \hat{p} < 0.57) = 0.9523$$

```
cat("$$ P(\\hat{p} > 0.54) = ", round(prob_c, 4), " $$\n")
```

$$P(\hat{p} > 0.54) = 0.1289$$

```
# Plot the sampling distribution
x <- seq(0.35, 0.65, by = 0.001)
y <- dnorm(x, mean = p, sd = sd_p)

plot(x, y, type = "l", lwd = 2, col = "blue",
     main = "Sampling Distribution of Sample Proportion",
     xlab = "Sample Proportion (p̂)",
     ylab = "Density")
```



Problem 8.34

```r
# Given values
n <- 200
p <- 0.5
q <- 1 - p

# Standard deviation
sd_p <- sqrt(p * q / n)

# Probabilities
prob_a <- pnorm(0.40, mean = p, sd = sd_p)
prob_b <- pnorm(0.57, mean = p, sd = sd_p) - pnorm(0.43, mean = p, sd = sd_p)
prob_c <- 1 - pnorm(0.54, mean = p, sd = sd_p)

cat("$$ \\text{Standard deviation } (\\sigma_{\\hat{p}}) = ", round(sd_p, 4), " $$\n\n")
```

$$\text{Standard deviation } (\sigma_{\hat{p}}) = 0.0354$$

```r
cat("$$ P(\\hat{p} < 0.40) = ", round(prob_a, 4), " $$\n")
```

$$P(\hat{p} < 0.40) = 0.0023$$

```r
cat("$$ P(0.43 < \\hat{p} < 0.57) = ", round(prob_b, 4), " $$\n")
```

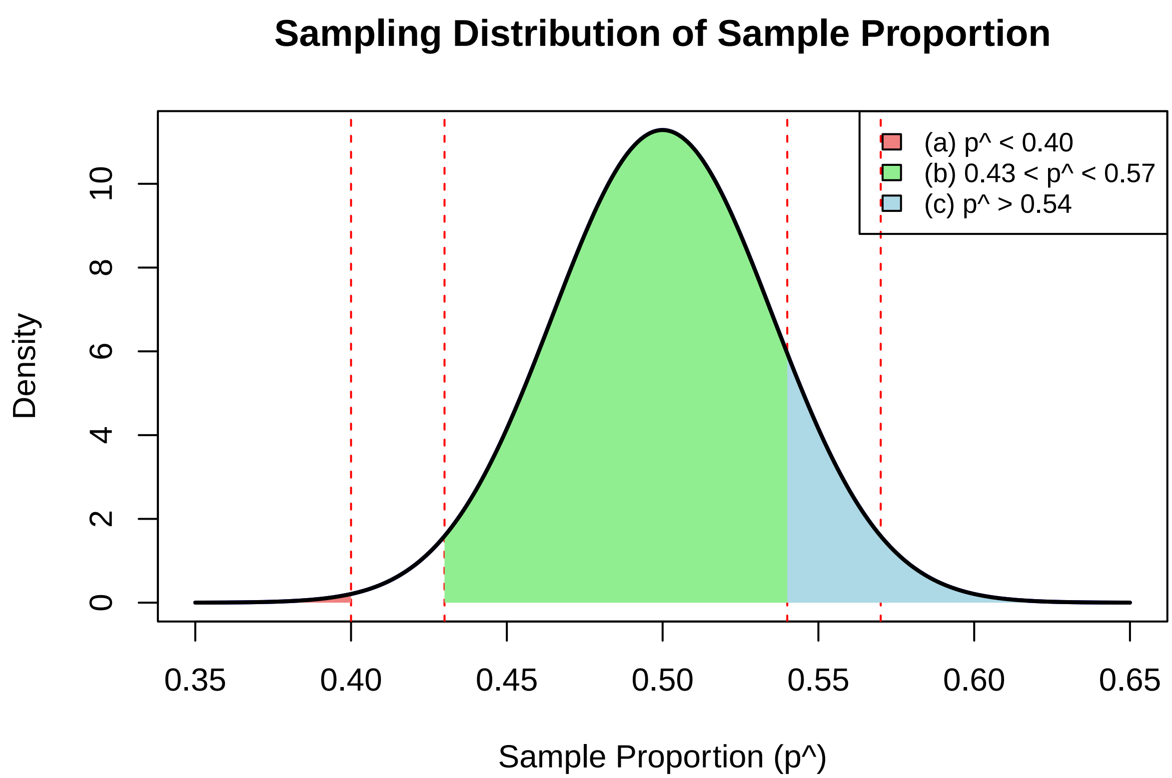
$$P(0.43 < \hat{p} < 0.57) = 0.9523$$

```r
cat("$$ P(\\hat{p} > 0.54) = ", round(prob_c, 4), " $$\n")
```

$$P(\hat{p} > 0.54) = 0.1289$$

```r
# Plot the sampling distribution
x <- seq(0.35, 0.65, by = 0.001)
y <- dnorm(x, mean = p, sd = sd_p)

plot(x, y, type = "l", lwd = 2, col = "blue",
     main = "Sampling Distribution of Sample Proportion",
     xlab = "Sample Proportion (p̂)",
     ylab = "Density")

# Reference lines
abline(v = c(0.40, 0.43, 0.54, 0.57), col = "red", lty = 2)

# --- (a) Shade p̂ < 0.40 ---
x_a <- seq(0.35, 0.40, by = 0.001)
y_a <- dnorm(x_a, mean = p, sd = sd_p)
polygon(c(x_a, rev(x_a)), c(y_a, rep(0, length(y_a))), col = "lightcoral", border = NA)

# --- (b) Shade 0.43 < p̂ < 0.57 ---
```

```
x_b <- seq(0.43, 0.57, by = 0.001)
y_b <- dnorm(x_b, mean = p, sd = sd_p)
polygon(c(x_b, rev(x_b)), c(y_b, rep(0, length(y_b))), col = "lightgreen", border = NA)

# --- (c) Shade p̂ > 0.54 ---
x_c <- seq(0.54, 0.65, by = 0.001)
y_c <- dnorm(x_c, mean = p, sd = sd_p)
polygon(c(x_c, rev(x_c)), c(y_c, rep(0, length(y_c))), col = "lightblue", border = NA)

# Redraw curve
lines(x, y, lwd = 2, col = "black")

# Legend
legend("topright",
       legend = c("(a) p̂ < 0.40", "(b) 0.43 < p̂ < 0.57", "(c) p̂ > 0.54"),
       fill = c("lightcoral", "lightgreen", "lightblue"),
       border = "black", cex = 0.8)
```



Standard deviation $(\sigma_{\hat{p}}) = 0.0354$

$$P(\hat{p} < 0.40) = 0.0082$$

$$P(0.43 < \hat{p} < 0.57) = 0.9802$$

$$P(\hat{p} > 0.54) = 0.1141$$

Interpretation

1. There's only about a **0.82%** chance that less than 40% will be boys.
2. There's a **98%** chance that between 43% and 57% will be girls.
3. There's an **11.5%** chance that more than 54% will be boys.

Problem 8.49

```r
x <- c(6, 9, 12, 15, 18)
probab <- c(0.1, 0.2, 0.4, 0.2, 0.1)
n <- 2

# --- Population mean and variance ---
mu <- sum(x * probab)
sigma2 <- sum(probab * (x - mu)^2)

cat("$$ \\text{Population mean } (\\mu) = ", mu, " $$\n")
```

$$\text{Population mean } (\mu) = 12$$

```r
cat("$$ \\text{Population variance } (\\sigma^2) = ", sigma2, " $$\n\n")
```

$$\text{Population variance } (\sigma^2) = 10.8$$

```r
samples <- expand.grid(x1 = x, x2 = x, KEEP.OUT.ATTRS = FALSE)
samples$mean <- rowMeans(samples)

# Probability lookup for each element
p_lookup <- setNames(probab, x)
samples$prob <- p_lookup[as.character(samples$x1)] * p_lookup[as.character(samples$x2)]

# --- Table 1: All possible samples ---
cat("**Table 1. All Possible Samples (n = 2)**\n\n")
```

Table 1. All Possible Samples (n = 2)

```r
knitr::kable(samples,
             col.names = c("x1", "x2", "Sample Mean (x̄)", "Probability"),
             caption = "All ordered samples with their probabilities.",
             align = "c")
```

Table 1: All ordered samples with their probabilities.

x1	x2	Sample Mean (x̄)	Probability
6	6	6.0	0.01
9	6	7.5	0.02
12	6	9.0	0.04
15	6	10.5	0.02
18	6	12.0	0.01
6	9	7.5	0.02
9	9	9.0	0.04
12	9	10.5	0.08
15	9	12.0	0.04
18	9	13.5	0.02
6	12	9.0	0.04
9	12	10.5	0.08
12	12	12.0	0.16
15	12	13.5	0.08
18	12	15.0	0.04
6	15	10.5	0.02
9	15	12.0	0.04
12	15	13.5	0.08
15	15	15.0	0.04
18	15	16.5	0.02
6	18	12.0	0.01
9	18	13.5	0.02
12	18	15.0	0.04
15	18	16.5	0.02
18	18	18.0	0.01

```r
# --- Sampling distribution of the sample mean ---
dist_mean <- aggregate(prob ~ mean, data = samples, FUN = sum)
dist_mean <- dist_mean[order(dist_mean$mean), ]

# --- Table 2: Sampling distribution ---
cat("\n**Table 2. Sampling Distribution of the Sample Mean**\n\n")
```

Table 2. Sampling Distribution of the Sample Mean

```r
knitr::kable(dist_mean,
             col.names = c("Sample Mean (x̄)", "Probability"),
             caption = "Distribution of sample means and their probabilities.",
             align = "c")
```

Table 2: Distribution of sample means and their probabilities.

Sample Mean (x̄)	Probability
6.0	0.01
7.5	0.04
9.0	0.12
10.5	0.20
12.0	0.26

Sample Mean ($\bar{x}$)	Probability
13.5	0.20
15.0	0.12
16.5	0.04
18.0	0.01

```r
# --- Mean and variance of sampling distribution ---
mu_xbar <- sum(dist_mean$mean * dist_mean$prob)
sigma2_xbar <- sum(dist_mean$prob * (dist_mean$mean - mu_xbar)^2)

# --- Display summary in LaTeX ---
cat("\n\n$$ \\text{Mean of sampling distribution } (\\mu_{\\bar{X}}) = ", mu_xbar, " $$\n")
```

$$\text{Mean of sampling distribution } (\mu_{\bar{X}}) = 12$$

```r
cat("$$ \\text{Variance of sampling distribution } (\\sigma^2_{\\bar{X}}) = ", sigma2_xbar, " $$\n")
```

$$\text{Variance of sampling distribution } (\sigma^2_{\bar{X}}) = 5.4$$

```r
cat("$$ \\frac{\\sigma^2}{n} = ", sigma2 / n, " $$\n")
```

$$\frac{\sigma^2}{n} = 5.4$$

$$\text{Population mean } (\mu) = 12$$

$$\text{Population variance } (\sigma^2) = 10.8$$

Table 1. All Possible Samples (n = 2)

x1	x2	Sample Mean ($\bar{x}$)	Probability
6	6	6.00	0.0100
9	6	7.50	0.0200
12	6	9.00	0.0400
15	6	10.50	0.0200
18	6	12.00	0.0100
6	9	7.50	0.0200
9	9	9.00	0.0400
12	9	10.50	0.0800
15	9	12.00	0.0400
18	9	13.50	0.0200
6	12	9.00	0.0400
9	12	10.50	0.0800
12	12	12.00	0.1600
15	12	13.50	0.0800
18	12	15.00	0.0400

x1	x2	Sample Mean ($\bar{x}$)	Probability
6	15	10.50	0.0200
9	15	12.00	0.0400
12	15	13.50	0.0800
15	15	15.00	0.0400
18	15	16.50	0.0200
6	18	12.00	0.0100
9	18	13.50	0.0200
12	18	15.00	0.0400
15	18	16.50	0.0200
18	18	18.00	0.0100

Table 2. Sampling Distribution of the Sample Mean

Sample Mean ($\bar{x}$)	Probability
6.00	0.0100
7.50	0.0400
9.00	0.1200
10.50	0.2000
12.00	0.2400
13.50	0.2000
15.00	0.1200
16.50	0.0400
18.00	0.0100

$$\text{Mean of sampling distribution } (\mu_{\bar{X}}) = 12$$

$$\text{Variance of sampling distribution } (\sigma^2_{\bar{X}}) = 5.4$$

$$\frac{\sigma^2}{n} = 5.4$$

Interpretation

The population of student credit hours at Metropolitan Technological College has an average of **12 hours** with moderate variation.
When we take samples of size 2 (with replacement), the mean of all possible sample means remains **12**, confirming that the sample mean is an **unbiased estimator**.
However, the spread of these sample means is smaller (**variance = 5.4**) than the original population (**variance = 10.8**), showing that averaging reduces variability and makes sample means more reliable indicators of the true population mean.