Assignment4 - Tiya Chokhani

2025-03-04

Part I

Exercise 1

(a) Area between these two values of 1 standard deviation from the mean

> **Answer:** The area between these two values is 0.6827, meaning 68.27% of the data falls within one standard deviation from the mean. The empirical rule states that 68% of the data in a normal distribution lies within one standard deviation.

(b) Area between 2 and 3 standard deviations from the mean

> **Answer:** The area between ±2 SD is 0.9545 (95.45%), and between ±3 SD is 0.9973 (99.73%). The empirical rule states it should be 95% and 99.7% respectively so it confirms this.

(c) Standard deviations above and below the mean for quartiles

> **Answer:** From the applet it can be observed that at -0.686 standard deviations away from the mean, approximately 25% of the data is below that value. Since the normal distribution is symmetric, Q3 would be the positive of this at 0.686 standard deviations away from the mean, with approximately 25% of the data being above that value.

> Another way of approaching this is if the middle area is 50%, everything outside of it should be 50% as well and since it's symmetric, it's split evenly across both sides. So that means there's 25% above Q3 and 25% below Q1

Exercise 2

```
mean_height <- 69
sd_height <- 2.8

# (a) P(X < 65)
prob_less_65 <- pnorm(65, mean = mean_height, sd = sd_height)
prob_less_65
```

```
## [1] 0.07656373
```

> **Answer:** Approximately 7.65% of males are shorter than 65 inches.

```r
# (b) P(X > 75)
prob_more_75 <- 1 - pnorm(75, mean = mean_height, sd = sd_height)
prob_more_75
```

```
## [1] 0.01606229
```

> **Answer:** Approximately 1.6% of males are taller than 75 inches.

```r
# (c) P(66 < X < 72)
prob_between <- pnorm(72, mean = mean_height, sd = sd_height) - pnorm(66, mean = mean
_height, sd = sd_height)
prob_between
```

```
## [1] 0.7160232
```

> **Answer:** Approximately 71.6% of males fall between 66 and 72 inches.

Exercise 3

```r
# (a) Shortest 0.5%
shortest_height <- qnorm(0.005, mean = mean_height, sd = sd_height)
shortest_height
```

```
## [1] 61.78768
```

> **Answer:** A male must be approximately 61.8 inches tall to be in the shortest 0.5%.

```r
# (b) Tallest 0.25%
tallest_height <- qnorm(0.9975, mean = mean_height, sd = sd_height)
tallest_height
```

```
## [1] 76.85969
```

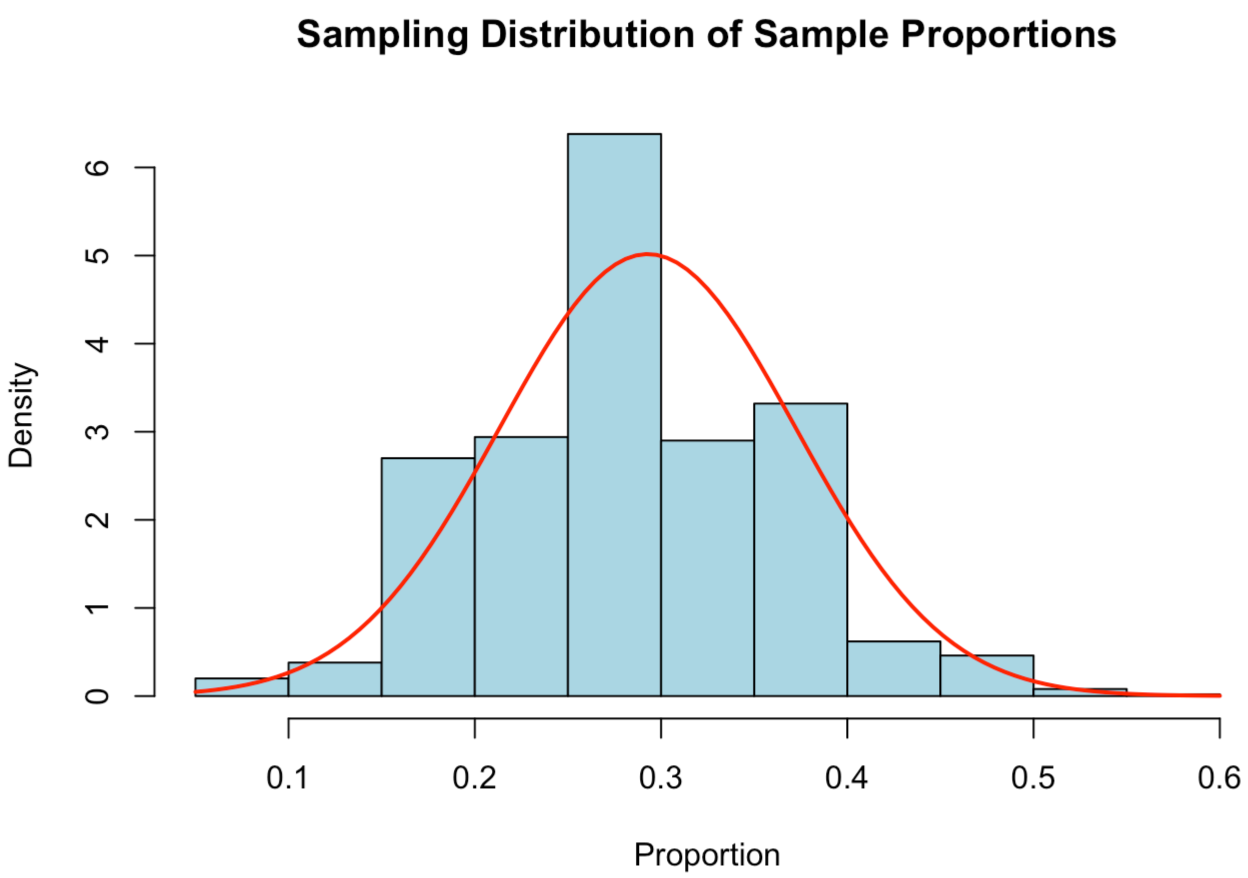
> **Answer:** A male must be approximately 76.8 inches tall to be in the tallest 0.25%.

Exercise 4

```
pawnee <- read.csv("pawnee.csv")
n <- 30 #sample siz
N <- 541
M <- 1000
phats <- numeric(M)
set.seed(123)

for(i in seq_len(M)){
  index <- sample(N, size = n)
  sample_i <- pawnee[index, ]
  phats[i] <- mean(sample_i$New_hlth_issue == "Y")
}

hist(phats, probability = TRUE, main = "Sampling Distribution of Sample Proportions",
xlab = "Proportion", col = "lightblue")
curve(dnorm(x, mean(phats), sd(phats)), add = TRUE, col = "red", lwd = 2)
```




```
# Mean and SD of simulated sample proportions
mean_phats <- mean(phats)
sd_phats <- sd(phats)
c(mean_phats, sd_phats)
```

```
## [1] 0.29280000 0.07951963
```

> **Answer:** Mean = 0.2928, SD = 0.0795.

Part II

Exercise 1

```
# (a) Probability of newborn <= 18 inches
mean_length <- 19.2
sd_length <- 0.7
prob_less_18 <- pnorm(18, mean = mean_length, sd = sd_length)
prob_less_18
```

```
## [1] 0.04323813
```

> **Answer:** Probability is 4.36%.

```
# (b) Probability of newborn > 20 inches
prob_more_20 <- 1 - pnorm(20, mean = mean_length, sd = sd_length)
prob_more_20
```

```
## [1] 0.126549
```

> **Answer:** Probability is 12.7%.

```
# (c) Percentage not fitting in 18-20.4 inches
prob_not_fit <- pnorm(18, mean = mean_length, sd = sd_length) + (1 - pnorm(20.4, mean
= mean_length, sd = sd_length))
prob_not_fit
```

```
## [1] 0.08647627
```

> **Answer:** Probability is 8.65%.

Exercise 2: Normal distribution N(400,60), top 30% admission, student scores 428, mean (μ) = 400, standard deviation (σ) = 60, X = 428

> **Answer**: Z = (X – μ)/ σ = 428 – 400/60 = 28/60 = 7/15 ≈ 0.467. Top 30% so Z-score attached to 30% is -0.524. We want top 30% so it is the Z-score of + 0.524 as the normal distribution curve tends to be symmetric. Since **0.467 < 0.524**, we can say that the student will NOT get admitted.

Exercise 3

```
# Expected sample proportion
p_hat <- 0.58
n <- 100
expected_value <- p_hat
expected_value
```

```
## [1] 0.58
```

> **Answer:** Expected proportion is 0.58.

```
# Standard error
standard_error <- sqrt((p_hat * (1 - p_hat)) / n)
standard_error
```

```
## [1] 0.04935585
```

> **Answer:** Standard error is 0.049.

c)

> **Answer:** We expect 58% of the students in the sample to have their driver's license, give or take 4.9%

```
# (d) Standard error for sample size 700
n_large <- 700
standard_error_large <- sqrt((p_hat * (1 - p_hat)) / n_large)
standard_error_large
```

```
## [1] 0.01865476
```

> **Answer:** Increasing the sample size reduces the standard error to 0.0185.

Exercise 4

a)

```
# Given proportion
p_hat <- 0.58
n <- 300

p_hat * 100
```

```
## [1] 58
```

> **Answer:** We would expect around 58% of young Americans from the sample to watch TV primarily through streaming services.

b)

```r
# Check CLT conditions
np <- n * p_hat
nq <- n * (1 - p_hat)

np
```

```
## [1] 174
```

```r
nq
```

```
## [1] 126
```

> **Answer:** Both conditions are satisfied: np = 300 * 0.58 = 174, which is greater than 10. nq = 300 * 0.42 = 126, which is also greater than 10. Thus, the conditions for the Central Limit Theorem are met and it is a normal distribution.

c) Sampling Distribution

```r
p_sample <- 181 / n  # Observed sample proportion

SE <- sqrt((p_hat * (1 - p_hat)) / n)

z_score <- (p_sample - p_hat) / SE

p_value <- 1 - pnorm(z_score)

cat("Z-score:", round(z_score, 3), "\n")
```

```
## Z-score: 0.819
```

```r
cat("P-value:", round(p_value, 5), "\n")
```

```
## P-value: 0.20644
```

> **Answer:** The probability (p-value) is 0.2 (which is > 0.05), this outcome would be not be unusual or suprising.

d) Probability of Finding More than 65% Watching Through Streaming Services

```
# Z-score for p-hat = 0.65
p_sample <- 0.65
z <- (p_sample - p_hat) / SE

# Probability corresponding to z-score
probability <- 1 - pnorm(z)
probability
```

```
## [1] 0.00701453
```

> **Answer:** The probability of more than 65% of the sample watching television primarily through streaming services is approximately 0.007.

Exercise 5

a)

```
# Given information
p_5 <- 0.82  # proportion of people who believe protecting the rights of those with u
npopular views is important
n_5 <- 800    # sample size

# Verifying the CLT conditions for Exercise 5
np_5 <- n_5 * p_5
nq_5 <- n_5 * (1 - p_5)

# Check if np >= 10 and nq >= 10
np_5 >= 10
```

```
## [1] TRUE
```

```
nq_5 >= 10
```

```
## [1] TRUE
```

> **Answer:** Both conditions are satisfied: np = 800 * 0.82 = 656, which is greater than 10. nq = 800 * 0.18 = 144, which is also greater than 10. Thus, the conditions for the Central Limit Theorem are met for this sample and dist is normal.

b)

```
# Critical value for 95% confidence
z_critical <- qnorm(0.975)

# Standard error of the sample proportion
se <- sqrt(p_5 * (1 - p_5) / n_5)

# Confidence interval
lower_bound <- p_5 - z_critical * se
upper_bound <- p_5 + z_critical * se
c(lower_bound, upper_bound)
```

```
## [1] 0.7933777 0.8466223
```

> **Answer:** The 95% confidence interval for the proportion of adults who believe protecting the rights of those with unpopular views is important is approximately (0.793, 0.847).

c) Comparing the 90% and 95% Confidence Intervals

```
# Critical value for 90% confidence
z_critical_90 <- qnorm(0.95)

# Standard error for 90% CI
se_90 <- sqrt(p_5 * (1 - p_5) / n_5)

# Confidence interval for 90% CI
lower_bound_90 <- p_5 - z_critical_90 * se_90
upper_bound_90 <- p_5 + z_critical_90 * se_90
c(lower_bound_90, upper_bound_90)
```

```
## [1] 0.7976578 0.8423422
```

> **Answer:** A 90% confidence interval is narrower than a 95% confidence interval because reducing confidence decreases the range.