STAT3355(HW-6)

Tulasi Janjanam

2024-11-11

Problem 1

```r
# Given parameters
alpha <- 256000  # shape parameter
beta <- 16000    # rate parameter

# Mean and variance of the Gamma distribution
mean_pop <- alpha / beta
var_pop <- alpha / (beta^2)

# Standard deviation of the population
sd_pop <- sqrt(var_pop)
# Sample size
n <- 34
# Standard deviation of the sample mean
sd_sample_mean <- sd_pop / sqrt(n)
desired_mean <- 16.01
z_score <- (desired_mean - mean_pop) / sd_sample_mean

# Probability that the sample mean is greater than 16.01
probability <- 1 - pnorm(z_score)
probability
```

```
## [1] 0.03259821
```

Problem 2

```r
# Given parameters
mu <- 8.2     # population mean
sigma <- 1    # population standard deviation
n <- 60       # sample size
mean_sample <- mu
sd_sample <- sigma / sqrt(n)

# Results
mean_sample  # Mean of the sample mean
```

```
## [1] 8.2
```

```r
sd_sample    # Standard deviation of the sample mean
```

```
## [1] 0.1290994
```

```r
# 90th percentile for the sample mean
percentile_90 <- qnorm(0.90, mean = mu, sd = sd_sample)
percentile_90
```

```
## [1] 8.365448
```

```r
# Output result as a complete sentence
cat("The 90th percentile for the sample mean time for app engagement is approximately", round(percentil
```

```
## The 90th percentile for the sample mean time for app engagement is approximately 8.37 minutes. This r
```

```r
prob_1_sd <- pnorm(mu + sd_sample) - pnorm(mu - sd_sample)
prob_2_sd <- pnorm(mu + 2 * sd_sample) - pnorm(mu - 2 * sd_sample)
prob_3_sd <- pnorm(mu + 3 * sd_sample) - pnorm(mu - 3 * sd_sample)
#results
prob_1_sd  # Probability within ±1 SD
```

```
## [1] 3.330669e-16
```

```r
prob_2_sd  # Probability within ±2 SD
```

```
## [1] 9.992007e-16
```

```r
prob_3_sd  # Probability within ±3 SD
```

```
## [1] 2.775558e-15
```

```r
# Given parameters
mu <- 8.2    # population mean
sigma <- 1   # population standard deviation
n <- 60      # sample size

# (a) Mean and standard deviation of the sampling distribution
mean_sample <- mu
sd_sample <- sigma / sqrt(n)

# Output results for part (a)
cat("Mean of the sampling distribution: ", mean_sample, "\n")
```

```
## Mean of the sampling distribution:  8.2
```

```r
cat("Standard deviation (standard error) of the sampling distribution: ", sd_sample, "\n")
```

```
## Standard deviation (standard error) of the sampling distribution:  0.1290994
```

```r
# (b) 90th percentile for the sample mean
percentile_90 <- qnorm(0.90, mean = mu, sd = sd_sample)
cat("90th percentile for the sample mean: ", percentile_90, "\n")
```

```
## 90th percentile for the sample mean:  8.365448
```

```r
# (c) Probabilities for ±1, ±2, and ±3 standard deviations from the mean
prob_1_sd <- pnorm(mu + sd_sample) - pnorm(mu - sd_sample)
prob_2_sd <- pnorm(mu + 2 * sd_sample) - pnorm(mu - 2 * sd_sample)
prob_3_sd <- pnorm(mu + 3 * sd_sample) - pnorm(mu - 3 * sd_sample)

# Output results for part (c) --- USING EMPIRICAL RULE
cat("Probability within ±1 standard deviation: ", prob_1_sd, "\n")
```

```
## Probability within ±1 standard deviation:  3.330669e-16
```

```r
cat("Probability within ±2 standard deviations: ", prob_2_sd, "\n")
```

```
## Probability within ±2 standard deviations:  9.992007e-16
```

```r
cat("Probability within ±3 standard deviations: ", prob_3_sd, "\n")
```

```
## Probability within ±3 standard deviations:  2.775558e-15
```

```r
# Given parameters
N <- 5           # trials
omega <- 0.5     # probability of success
n <- 75          # sample size

# Mean and standard deviation for a single trial
mu_single <- N * omega
sigma_single <- sqrt(N * omega * (1 - omega))

# Standard error of the sample mean
se_sample <- sigma_single / sqrt(n)

# Mean and standard deviation for the total stress score
mu_total <- mu_single * n
sd_total <- sigma_single * sqrt(n)

# (a) Probability that the average stress score is less than 2.25
z_a <- (2.25 - mu_single) / se_sample
prob_a <- pnorm(z_a)
cat("The probability that the average stress score is less than 2.25 is:", prob_a, "\n")
```

```
## The probability that the average stress score is less than 2.25 is: 0.02640376
```

```r
# (b) 90th percentile for the average stress score
percentile_90 <- qnorm(0.90, mean = mu_single, sd = se_sample)
cat("The 90th percentile for the average stress score is:", percentile_90, "\n")
```

```
## The 90th percentile for the average stress score is: 2.665448
```

```r
# (c) Probability that the total stress score is less than 200
z_c <- (200 - mu_total) / sd_total
prob_c <- pnorm(z_c)
cat("The probability that the total stress score is less than 200 is:", prob_c, "\n")
```

```
## The probability that the total stress score is less than 200 is: 0.9016472
```

```r
# (d) 90th percentile for the total stress score
percentile_total_90 <- qnorm(0.90, mean = mu_total, sd = sd_total)
cat("The 90th percentile for the total stress score is:", percentile_total_90, "\n")
```

```
## The 90th percentile for the total stress score is: 199.9086
```

```r
# Given parameters
lambda <- 1 / 2  # rate parameter
n <- 80  # sample size
mu_single <- 2  # mean
se_sample <- 2 / sqrt(n)  # standard error

# (a) Probability that one customer’s excess data use is larger than 2.5 Gb
prob_a <- exp(-lambda * 2.5)
cat("The probability that one customer’s excess data use is larger than 2.5 Gb is:", prob_a, "\n")
```

```
## The probability that one customer’s excess data use is larger than 2.5 Gb is: 0.2865048
```

```r
# (b) Probability that the average excess data used by the 80 customers is larger than 2.5 Gb
z_sample <- (2.5 - mu_single) / se_sample
prob_b <- 1 - pnorm(z_sample)
cat("The probability that the average excess data used by the 80 customers is larger than 2.5 Gb is:", p
```

```
## The probability that the average excess data used by the 80 customers is larger than 2.5 Gb is: 0.01
```

```r
# (c) Explanation of why probabilities are different
cat("The difference arises because part (a) refers to a single customer's data usage, which is more vari
    "while part (b) involves the average of 80 customers, which has less variability due to the larger s
```

```
## The difference arises because part (a) refers to a single customer’s data usage, which is more varia
```

```r
# Given data
n <- 30          # sample size
x <- 22          # number of students who like video games
p_hat <- x / n   # sample proportion

# Confidence level and Z-score for 95%
Z <- 1.96
SE <- sqrt((p_hat * (1 - p_hat)) / n)  # standard error

# Confidence Interval
```

```r
ME <- Z * SE  # margin of error
CI_lower <- p_hat - ME
CI_upper <- p_hat + ME

cat("The 95% confidence interval for the proportion of students who like video games is (",
    round(CI_lower, 3), ",", round(CI_upper, 3), ")\n")
```

```
## The 95% confidence interval for the proportion of students who like video games is ( 0.575 , 0.892 )
```

```r
# Given data
n <- 15                 # number of people
mu <- 180               # mean weight
sigma <- 25             # standard deviation
safe_limit <- 3500      # weight limit
# Total mean and standard deviation for 15 people
mu_T <- n * mu
sigma_T <- sqrt(n) * sigma

# probability P(T > 3500)
prob_exceeds <- 1 - pnorm(safe_limit, mean = mu_T, sd = sigma_T)

cat("The probability that the total weight of 15 people exceeds 3500 lbs is:", prob_exceeds, "\n")
```

```
## The probability that the total weight of 15 people exceeds 3500 lbs is: 1.110223e-16
```

```r
# Given data
mean_daily <- 25
variance_daily <- 25
days <- 30
target_sales <- 600

# Calculate the mean and standard deviation for 30-day sales
mean_30days <- days * mean_daily
sd_30days <- sqrt(days * variance_daily)

# Calculate the probability that sales exceed 600 bottles
prob_exceeds <- 1 - pnorm(target_sales, mean = mean_30days, sd = sd_30days)

cat("The probability that the restaurant will sell more than 600 bottles in the next 30 days is:", prob
```

```
## The probability that the restaurant will sell more than 600 bottles in the next 30 days is: 1
```

```r
# Given data
n <- 30                     # sample size
left_handed <- 2            # number of left-handed students in sample
p_hat <- left_handed / n    # sample proportion
z <- 1.96                   # z-score for 95% confidence

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

```r
# Confidence interval
lower_bound <- p_hat - z * SE
upper_bound <- p_hat + z * SE

cat("The 95% confidence interval for the proportion of left-handed students is:",
    "(", max(0, lower_bound), ",", upper_bound, ")\n")
```

```
## The 95% confidence interval for the proportion of left-handed students is: ( 0 , 0.155929 )
```

```r
# Load the necessary package and dataset
library(UsingR)
```

```
## Loading required package: MASS

## Loading required package: HistData

## Loading required package: Hmisc

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```r
data("babies")

# Extract the mother and father ages
mother_age <- babies$age
father_age <- babies$dage

# Remove NA values
mother_age <- na.omit(mother_age)
father_age <- na.omit(father_age)

# Calculate sample means and standard deviations
mean_mother <- mean(mother_age)
sd_mother <- sd(mother_age)
n_mother <- length(mother_age)

mean_father <- mean(father_age)
sd_father <- sd(father_age)
n_father <- length(father_age)

# Calculate the difference in means and standard error
mean_diff <- mean_mother - mean_father
SE_diff <- sqrt((sd_mother^2 / n_mother) + (sd_father^2 / n_father))

# 95% confidence interval
z <- 1.96
lower_bound <- mean_diff - z * SE_diff
```

```r
upper_bound <- mean_diff + z * SE_diff

cat("The 95% confidence interval for the difference in mean age (mother - father) is:",
    "(", lower_bound, ",", upper_bound, ")\n")
```

```
## The 95% confidence interval for the difference in mean age (mother - father) is: ( -3.961765 , -2.769
```

```r
# Check if the interval contains 0
if (lower_bound <= 0 && upper_bound >= 0) {
    cat("The confidence interval contains 0, suggesting no significant difference in mean ages.\n")
} else {
    cat("The confidence interval does not contain 0, suggesting a significant difference in mean ages.\n
}
```

```
## The confidence interval does not contain 0, suggesting a significant difference in mean ages.
```