mkhan_hw7

2023-11-30

```r
shape_parameter <- 256000
rate_parameter <- 16000

sample_size <- 34

# Mean and standard deviation of the gamma distribution
mean_gamma <- shape_parameter / rate_parameter
sd_gamma <- sqrt(shape_parameter) / rate_parameter

# Calculate the probability using the complement of the CDF
probability_greater_than_16.01 <- 1 - pgamma(16.01, shape = shape_parameter, rate = rate_parameter, lowe

se <- sd_gamma / sqrt(sample_size)

# Calculate the z-score
z_score <- (16.01 - mean_gamma) / se

probability_z_score <- pnorm(z_score, lower.tail = FALSE)

cat("Probability that the average fill is greater than 16.01 ounces:", probability_greater_than_16.01,
```

```
## Probability that the average fill is greater than 16.01 ounces: 0.3756898
```

```r
cat("Probability using z-score approach:", probability_z_score, "\n")
```

```
## Probability using z-score approach: 0.03259821
```

```r
# Part a
mu <- 8.2
sigma <- 1
sample_size <- 60

mean_sample_mean <- mu
sd_sample_mean <- sigma / sqrt(sample_size)

cat("Mean of the sampling distribution of the sample mean:", mean_sample_mean, "minutes\n")
```

```
## Mean of the sampling distribution of the sample mean: 8.2 minutes
```

```r
cat("Standard deviation of the sampling distribution of the sample mean:", sd_sample_mean, "minutes\n")
```

```
## Standard deviation of the sampling distribution of the sample mean: 0.1290994 minutes
```

```r
# Part b
percentile <- 0.9

# Calculate the z-score for the 90th percentile
z_score <- qnorm(percentile)
```

```r
percentile_value <- mu + z_score * (sigma / sqrt(sample_size))

cat("90th percentile for the sample mean time:", percentile_value, "minutes\n")
```

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

```r
#Part c
prob_within_1_sd <- pnorm(mu + sigma / sqrt(sample_size), mean = mu, sd = sigma / sqrt(sample_size)) - p
prob_within_2_sd <- pnorm(mu + 2 * sigma / sqrt(sample_size), mean = mu, sd = sigma / sqrt(sample_size)
prob_within_3_sd <- pnorm(mu + 3 * sigma / sqrt(sample_size), mean = mu, sd = sigma / sqrt(sample_size)

cat("Probability that the sample mean is within ±1 standard deviation:", prob_within_1_sd, "\n")
```

```
## Probability that the sample mean is within ±1 standard deviation: 0.6826895
```

```r
cat("Probability that the sample mean is within ±2 standard deviations:", prob_within_2_sd, "\n")
```

```
## Probability that the sample mean is within ±2 standard deviations: 0.9544997
```

```r
cat("Probability that the sample mean is within ±3 standard deviations:", prob_within_3_sd, "\n")
```

```
## Probability that the sample mean is within ±3 standard deviations: 0.9973002
```

```r
# Part d
cat("Without using R you can state probablities of each standard deviation using the empirical rule:\n"
Probability within ±2 standard deviations: 95%
Probability within ±3 standard deviations: 99.7%", "\n")
```

```
## Without using R you can state probablities of each standard deviation using the empirical rule:
##  Probability within ±1 standard deviation: 68%
## Probability within ±2 standard deviations: 95%
## Probability within ±3 standard deviations: 99.7%
```

```r
#Part a
N <- 5
omega <- 0.5
sample_size <- 75
average_stress_score <- 2.25

# Calculate mean and standard deviation of the sampling distribution of the sample mean
mu_sample_mean <- N * omega
sd_sample_mean <- sqrt((N * omega * (1 - omega)) / sample_size)

z_score <- (average_stress_score - mu_sample_mean) / sd_sample_mean

probability_less_than_2.25 <- pnorm(z_score)

cat("Probability that the average stress score is less than 2.25:", probability_less_than_2.25, "\n")
```

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

```r
#Part b
percentile <- 0.9

# Calculate mean and standard deviation of the sampling distribution of the sample mean
mu_sample_mean <- N * omega
sd_sample_mean <- sqrt((N * omega * (1 - omega)) / sample_size)
```

```r
# Calculate the z-score for the 90th percentile
z_score <- qnorm(percentile)

percentile_value <- mu_sample_mean + z_score * sd_sample_mean

cat("90th percentile for the average stress score:", percentile_value, "\n")
```

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

```r
# Part c

# Calculate cumulative probability
cumulative_prob <- pbinom(199, size = sample_size * N, prob = omega)

cat("Probability that the total of the 75 stress scores is less than 200:", round(cumulative_prob, 4),
```

```
## Probability that the total of the 75 stress scores is less than 200: 0.8924
```

```r
# Part d

# Calculate mean and standard deviation of the binomial distribution
mu_binomial <- N * omega
sigma_binomial <- sqrt(N * omega * (1 - omega))

# Use normal approximation to find the z-score for the 90th percentile
z_score_90th_percentile <- qnorm(percentile, mean = mu_binomial, sd = sigma_binomial)

total_stress_score_90th_percentile <- round(z_score_90th_percentile * sigma_binomial + mu_binomial)

cat("90th percentile for the total stress score:", total_stress_score_90th_percentile, "\n")
```

```
## 90th percentile for the total stress score: 7
```

```r
# Part a
mean_sample <- 22
sd_sample <- 22 / sqrt(80)
threshold_value <- 20

z_score <- (threshold_value - mean_sample) / sd_sample

probability_greater_than_20 <- pnorm(z_score, lower.tail = FALSE)

cat("Probability that X-bar is greater than 20:", probability_greater_than_20, "\n")
```

```
## Probability that X-bar is greater than 20: 0.7919241
```

```r
# Part b

mean_population <- 22
sd_population <- 22
threshold_value <- 20

z_score <- (threshold_value - mean_population) / sd_population

probability_greater_than_20 <- pnorm(z_score, lower.tail = FALSE)

cat("Probability that X is greater than 20:", probability_greater_than_20, "\n")
```

```
## Probability that X is greater than 20: 0.5362176
```

```r
# Part c
cat("The probabilities differ because (a) calculates the probability for an individual customer's exces
```

```
## The probabilities differ because (a) calculates the probability for an individual customer's excess
```

```r
sample_size <- 30
sample_proportion <- 22 / 30
confidence_level <- 0.95
z_score <- qnorm((1 + confidence_level) / 2)  # Two-tailed interval

margin_of_error <- z_score * sqrt((sample_proportion * (1 - sample_proportion)) / sample_size)

# Calculate the confidence interval
confidence_interval_lower <- sample_proportion - margin_of_error
confidence_interval_upper <- sample_proportion + margin_of_error

cat("Calculated 95% Confidence Interval:", confidence_interval_lower, "to", confidence_interval_upper,
```

```
## Calculated 95% Confidence Interval: 0.575091 to 0.8915756
```

```r
sample_size <- 30
sample_proportion <- 2 / 30  # Proportion of left-handed students
confidence_level <- 0.95
z_score <- qnorm((1 + confidence_level) / 2)  # Two-tailed interval

margin_of_error <- z_score * sqrt((sample_proportion * (1 - sample_proportion)) / sample_size)

# Calculate the confidence interval
confidence_interval_lower <- sample_proportion - margin_of_error
confidence_interval_upper <- sample_proportion + margin_of_error

# Print the results
cat("Calculated 95% Confidence Interval:", confidence_interval_lower, "to", confidence_interval_upper,
```

```
## Calculated 95% Confidence Interval: -0.02259402 to 0.1559274
```

```r
library(UsingR)
```

```
## Warning: package 'UsingR' was built under R version 4.3.2

## Loading required package: MASS

## Loading required package: HistData

## Warning: package 'HistData' was built under R version 4.3.2

## Loading required package: Hmisc

## Warning: package 'Hmisc' was built under R version 4.3.2

##
## Attaching package: 'Hmisc'

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

```r
data(babies)

names(babies)
```

```
##  [1] "id"        "pluralty"  "outcome"   "date"      "gestation" "sex"
##  [7] "wt"        "parity"    "race"      "age"       "ed"        "ht"
## [13] "wt1"       "drace"     "dage"      "ded"       "dht"       "dwt"
## [19] "marital"   "inc"       "smoke"     "time"      "number"
```

```r
age = babies$age
dage = babies$dage

t.test(age, dage, var.equal = TRUE, paired = FALSE, conf.level = 0.95)
```

```
##
##  Two Sample t-test
##
## data:  age and dage
## t = -11.067, df = 2470, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.962047 -2.769345
## sample estimates:
## mean of x mean of y
##  27.37136  30.73706
```

```r
cat("For 95% confidence level, the population mean lies within (-3.96, -2.77)")
```

```
## For 95% confidence level, the population mean lies within (-3.96, -2.77)
```

```r
rm(list = ls())
x1 <- c(15.997,16.005,15.981,15.954,15.986,16.021,15.985,16.001,16.018,16.056)
x2<-NULL
t.test(x1, x2, alternative = "two.sided", paired = FALSE, mu = 16, conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  x1
## t = 0.045909, df = 9, p-value = 0.9644
## alternative hypothesis: true mean is not equal to 16
## 95 percent confidence interval:
##  15.98069 16.02011
## sample estimates:
## mean of x
##   16.0004
```

```r
shapiro.test(x1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x1
## W = 0.96609, p-value = 0.8524
```

```r
print("The observed effect size d is very small, 0.015. This indicates that the magnitude of the differe
```

```
## [1] "The observed effect size d is very small, 0.015. This indicates that the magnitude of the diffe
```

```r
x_bar <- 8.412  # sample mean
s <- 1.512      # sample standard deviation
n <- 60         # sample size
mu_0 <- 8.2     # hypothesized population mean
alpha <- 0.05   # significance level


t_stat <- (x_bar - mu_0) / (s / sqrt(n))

# Since it's a one-sided test and we are testing if the mean is greater, we look at the right tail
critical_value <- qt(1 - alpha, df = n - 1)

if (t_stat > critical_value) {
  cat("Reject the null hypothesis. There is enough evidence to suggest that the mean is more than 8.2 m
} else {
  cat("Fail to reject the null hypothesis. There is not enough evidence to suggest that the mean is mor
}
```

```
## Fail to reject the null hypothesis. There is not enough evidence to suggest that the mean is more th
```

```r
cat("\nTest Statistic:", round(t_stat, 4), "\nCritical Value:", round(critical_value, 4))
```

```
##
## Test Statistic: 1.0861
## Critical Value: 1.6711
```

```r
# Given data
n <- 200
x <- 130
p_null <- 0.70
alpha <- 0.05

p_hat <- x / n

se <- sqrt(p_null * (1 - p_null) / n)

# Calculate z-test statistic
z_stat <- (p_hat - p_null) / se

# Calculate two-tailed p-value
p_value <- 2 * pnorm(-abs(z_stat))

cat("Z-Test Statistic:", round(z_stat, 4), "\n")
```

```
## Z-Test Statistic: -1.543
```

```r
cat("P-Value:", p_value, "\n")
```

```
## P-Value: 0.1228226
```

```r
if (p_value < alpha) {
  cat("Reject the null hypothesis. There is enough evidence to suggest that the stress percentage is di
} else {
  cat("Fail to reject the null hypothesis. There is not enough evidence to suggest that the stress perc
}
```

```
## Fail to reject the null hypothesis. There is not enough evidence to suggest that the stress percentag
```

```r
n1 <- 30  # Sample size of male students
n2 <- 40  # Sample size of female students
x1 <- 22  # Number of male students playing video games
x2 <- 24  # Number of female students playing video games

p1 <- x1 / n1
p2 <- x2 / n2

# Calculate pooled sample proportion
p <- (x1 + x2) / (n1 + n2)

# Calculate test statistic
z <- (p1 - p2) / sqrt(p * (1 - p) * (1/n1 + 1/n2))

# Find critical value for one-sided test at alpha = 0.05
critical_value <- 1.96  # Critical value from the z-table for a 2-tailed test

cat("Test Statistic:", z, "\n")
```

```
## Test Statistic: 1.163038
```

```r
cat("Critical Value:", critical_value, "\n")
```

```
## Critical Value: 1.96
```

```r
if (z > critical_value) {
  cat("Reject the null hypothesis. There is statistical evidence that more males play video games among
} else {
  cat("Fail to reject the null hypothesis. There is no statistical evidence that more males play video g
}
```

```
## Fail to reject the null hypothesis. There is no statistical evidence that more males play video games
```

```r
age <- babies$age
dage <- babies$dage
x1bar <- mean(age)
x2bar <- mean(dage)

S1square <- var(age)
S2square <- var(dage)
t <- (x1bar - x2bar) / sqrt((S1square / 1236) + (S2square / 1236))

df = 2470
p.value <- dt(t, df)

t.test(age, dage, paired = FALSE, conf.level = 0.95, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  age and dage
## t = -11.067, df = 2470, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.962047 -2.769345
```

```
## sample estimates:
## mean of x mean of y
##  27.37136  30.73706
```