Homework 6

2025-11-05

Problem 1

```r
# Defines parameters
shape <- 256000
rate <- 16000
n <- 34
x_bar <- 16.01

# Calculates mu and sigma parameters
mu <- shape / rate
sigma <- sqrt(shape) / rate

# Calculates standard error (se)
se <- sigma / sqrt(n)

# Calculates z-score
z <- (x_bar - mu) / se

# Calculates probability that the sample mean (x_bar) > 16.01
probability <- pnorm(z, lower.tail = FALSE)
probability
```

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

The probability that the average fill of the sampled soda cans is greater than 16.01 ounces is `0.0325982`.

Problem 2

```r
# Below is the code for parts a - c (and the explanation for part d) of problem 2
# Defines parameters
sigma <- 1
sample <- 60
```

a)

```r
# Calculates mean and standard deviation
mu <- 8.2
mu
```

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

```
sd <- sigma / sqrt(sample)
sd
```

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

The mean of the sampling distribution of the sample mean is `8.2` and the standard deviation is `0.1290994`.

b)

```
# Calculates 90th percentile for sample mean
prob_90_q2 <- qnorm(0.90, mean = mu, sd = sd)
prob_90_q2
```

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

Interpretation: The value obtained from part b implies that 90% of possible random users from the sample will have a tablet usage time of `8.3654476` minutes or less.

c)

```
# Calculates + or - 1 standard deviation
p1 <- pnorm(1) - pnorm(-1)
# Calculates + or - 2 standard deviation
p2 <- pnorm(2) - pnorm(-2)
# Calculates + or - 3 standard deviation
p3 <- pnorm(3) - pnorm(-3)
c(p1, p2, p3)
```

```
## [1] 0.6826895 0.9544997 0.9973002
```

$\pm$**1** Standard Deviation: `0.6826895`

$\pm$**2** Standard Deviation: `0.9544997`

$\pm$**3** Standard Deviation: `0.9973002`

d) We can solve part c by using the Empirical Rule (the 68-95-99.7% rule): A basic generalization is that $\pm$**1** Standard Deviation is $\approx$ 68%, $\pm$**2** Standard Deviation is $\approx$ 95%, and $\pm$**3** Standard Deviation is $\approx$ 99.7%.

Problem 3

```
# Below is the code for parts a - d of problem 3
# Defines parameters
n <- 75
mu <- 5 * 0.5
sigma <- sqrt(5 * 0.5 * 0.5)
sd <- sigma / sqrt(n)
x_bar <- 2.25
```

a)

```r
# Calculates P(x_bar < 2.25)
z <- (x_bar - mu) / sd
prob_avg_score <- pnorm(z)
prob_avg_score
```

```
## [1] 0.02640376
```

b)

```r
# Calculates 90th percentile
prob_90_q3b <- qnorm(0.90, mean = mu, sd = sd)
prob_90_q3b
```

```
## [1] 2.665448
```

c)

```r
z_total_score <- (200 - (n * mu)) / (sqrt(n) * sigma)
prob_total <- pnorm(z_total_score)
prob_total
```

```
## [1] 0.9016472
```

d)

```r
prob_90_q3d <- qnorm(0.90, mean = n * mu, sd = sqrt(n) * sigma)
prob_90_q3d
```

```
## [1] 199.9086
```

Problem 4

a)

```r
# Defines parameters
lambda <- 0.5
x <- 2.5

# Calculates probability of data usage being > 2.5
pexp(x, rate = lambda, lower.tail = FALSE)
```

```
## [1] 0.2865048
```

b)

```r
mu <- 2
sigma <- 2 / sqrt(80)
x_bar <- 2.5

pnorm(x_bar, mean = mu, sd = sigma, lower.tail = FALSE)
```

```
## [1] 0.01267366
```

c) The probabilities are different because part a follows an exponential distribution and measures the probability of an individual, whereas part b follows a normal distribution (central limit theorem) and measures the probability of a sample mean.

Problem 5

```r
# Defines parameters
p_hat <- 22/30
n <- 30
N <- 70

# Calculates standard error
se <- sqrt(p_hat * (1 - p_hat) / n) * sqrt((N - n) / (N - 1))

# Calculates confidence interval
ci_q5 <- p_hat + c(-1, 1) * qnorm(0.975) * se
ci_q5
```

```
## [1] 0.6128497 0.8538169
```

The 95% confidence interval is (`0.6128497, 0.8538169`) and 65% falls under that range.

Problem 6

```r
# Define parameters
mu <- 180
sigma <- 25
n <- 15
max_weight_per_person <- 3500

# Calculate total mean and standard deviation
mu_total <- n * mu
sd_total <- sqrt(n) * sigma

# Calculates z score and probablity of total weight > 3500
z <- (max_weight_per_person - mu_total) / sd_total
prob_exceeds_weight <- pnorm(z, lower.tail = FALSE)
prob_exceeds_weight
```

```
## [1] 7.140742e-17
```

Problem 7

```r
# Defines parameters
n <- 30
mu_nightly <- 25
var_nightly <- 25

# Calculates total mean and variance values
mu_total <- n * mu_nightly
var_total <- sqrt(n * var_nightly)

# Calculates z score
z <- (600 - mu_total) / var_total

# Calculates probability
prob_bottles_sold <- pnorm(z, lower.tail = FALSE)
prob_bottles_sold
```

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

Problem 8

```r
# Defines parameters
p_hat <- 2/30
n <- 30
N <- 54

# Calculates standard error
se <- sqrt(p_hat * (1 - p_hat) / n) * sqrt((N - n) / (N - 1))

# Calculates confidence interval
ci_q8 <- p_hat + c(-1, 1) * qnorm(0.975) * se
ci_q8
```

```
## [1] 0.006600778 0.126732556
```

The 95% confidence interval is (`0.0066008, 0.1267326`) and 13.1% does not fall in that range as it's greater, which suggests that the class has a lower rate of left-handed people than the overall US population.

Problem 9

```r
# Loads 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)

# Ensures the ages are within a valid range
valid_ages <- complete.cases(babies$age, babies$dage)

# Performing the t test on the ages
t.test(babies$dage[valid_ages], babies$age[valid_ages], paired = TRUE, conf.level = 0.95)
```

```
##
##  Paired t-test
##
## data:  babies$dage[valid_ages] and babies$age[valid_ages]
## t = 17.392, df = 1235, p-value < 2.2e-16
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  2.986035 3.745356
## sample estimates:
## mean difference
##        3.365696
```

Interpretation: The confidence interval does not contain 0. The data implies that on average, fathers are a few years older than mothers, but there is no significant difference in ages.