

Homework6

2024-04-07

Question 1

a)

```
mu <- 8.2
sigma <- 1
sample_size <- 60
```

```
standard_error <- sigma / sqrt(sample_size)
sampledist_mean <- mu
sampledist_sd <- round(standard_error, 3)
sampledist_mean
```

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

```
sampledist_sd
```

```
## [1] 0.129
```

b)

```
percentile_90 <- round(qnorm(0.90, mean = mu, sd = standard_error), 3)
percentile_90
```

```
## [1] 8.365
```

The 90th percentile for the sample mean time for app engagement for a tablet user is 8.365 minutes. This means that 90% of the sample means fall below 8.365 minutes.

c)

```
diff(pnorm(mu + c(-1, 1) * standard_error, mu, standard_error))
```

```
## [1] 0.6826895
```

```
diff(pnorm(mu + c(-2, 2) * standard_error, mu, standard_error))
```

```
## [1] 0.9544997
```

```
diff(pnorm(mu + c(-3, 3) * standard_error, mu, standard_error))
```

```
## [1] 0.9973002
```

d)

We can use the 68-95-99.7% rule (empirical rule), which tells us that for a normal distribution, around 68% of the data will fall within one standard deviation of the mean, around 95% will fall within two standard deviations of the mean, and 99.7% will fall within three standard deviations of the mean.

Question 2

a)

Mean (μ) = 2

$\lambda = \frac{1}{2} = 0.5$ Excess data threshold = 2.5 GB.

```
# Probability that an individual customer's excess data use is larger than 2.5 GB
1 - pexp(2.5, 0.5)
```

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

b)

To find the sample mean of 80 customers, we can use the Central Limit Theorem to approximate it using a normal distribution.

So, Sample Mean = $N(2, 2/\sqrt{80})$

```
# Probability that the average excess data use of 80 customers is larger than 2.5 GB
1 - pnorm(2.5, 2, (2/sqrt(80)))
```

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

c)

The probabilities in parts (a) and (b) are different because they are two different cases, and we used two different methods to calculate their probabilities.

In part (a), we are calculating the probability that a single customer's excess data use exceeds 2.5 GB, using the exponential distribution function.

In part (b), we are calculating the probability that the average excess data use of 80 customers (sample mean) exceeds 2.5 GB by using the Central Limit Theorem. The large sample size (80 customers) allows us to use the normal approximation to the sample mean, which results in a very small probability (almost 0) due to the high Z-score.

Question 3

Population proportion, $p = 0.65$

Sample proportion, $\hat{p} = 22/30 = 0.733$

Sample size, $n = 30$

Standard Error, $SE = \sqrt{\hat{p}(1 - \hat{p})/n} = \sqrt{0.7333(1 - 0.7333)/30} = 0.08074$

For 95% confidence interval, $z = 1.96$

Confidence interval = $\hat{p} \pm z * SE$

Confidence interval = $0.7333 \pm 1.96(0.08074) = (0.575, 0.891)$

So, the population proportion of 0.65 is within the confidence interval.

Question 4

Population proportion, $p = 0.131$

Sample proportion, $\hat{p} = 2/30 = 0.0667$

Sample size, $n = 30$

Standard Error, $SE = \sqrt{\hat{p}(1 - \hat{p})/n} = \sqrt{0.0667(1 - 0.0667)/30} = 0.04555$

For 95% confidence interval, $z = 1.96$

Confidence interval = $\hat{p} \pm z * SE$

Confidence interval = $0.0667 \pm 1.96(0.04555) = (-0.223, 0.1559)$

So, the population proportion of 0.131 is within the confidence interval.

Question 5

```
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

data(babies)
m_age <- babies$age
d_age <- babies$dage

length_m <- length(babies$age)
length_d <- length(babies$dage)

mean_m <- mean(m_age)
sd_m <- sd(m_age)

mean_d <- mean(d_age)
sd_d <- sd(d_age)

diff_mean = mean_m - mean_d

z <- 1.96

sd_m2 <- (sd_m)^2
sd_d2 <- (sd_d)^2

SE = sqrt(((sd_m2)/length_m) + ((sd_d2)/length_d))
c(diff_mean - z*SE, diff_mean + z*SE)

## [1] -3.961765 -2.769626
```

Looking at the confidence interval, which is (-3.961765, -2.769626), we can see that it does not contain 0. This suggests that there is a significant difference between the mean ages of mothers and fathers.