STAT 3355 - HW 7

Yebom Kim

2024-04-19

Problem 1

```
# Step 1: Define hypotheses
# H0: The fill of coca cola cans  = 16.00 ounces
# H1: The fill of coca cola cans are less than 16.00 ounces

# Step 2: Data
fill_data <- c(15.997, 16.005, 15.981, 15.954, 15.986, 16.021, 15.985, 16.001, 16.018, 16.056)
# Significance level is 0.05

# Step 3: Calculate sample mean and standard deviation
sample_mean <- mean(fill_data)
sample_sd <- sd(fill_data)

# Step 4: Calculate the test statistic (t-value)
n <- length(fill_data)
t_value <- (sample_mean - 16) / (sample_sd / sqrt(n))

# Step 5: Determine the p-value
p_value <- pt(t_value, df = n - 1, lower.tail = TRUE)

# Step 6: Interpret the results
alpha <- 0.05

if (p_value < alpha) {
  cat("There is statistical evidence that the fill of the cola cans
      is less than 16.00 ounces (p-value =", p_value, "<", alpha, ")\n")
} else {
  cat("There is no statistical evidence to conclude that the fill of the
      cola cans is less than 16.00 ounces (p-value =", p_value, ">= ", alpha, ")\n")
}
```

```
## There is no statistical evidence to conclude that the fill of the
##       cola cans is less than 16.00 ounces (p-value = 0.5178072 >=  0.05 )
```

Problem 2

```r
# Step 1: Define hypotheses
# H0: Proportion of stressed college students = 0.70
# H1: Proportion of stressed college students is not 0.70

# Step 2: Data
n <- 200
x <- 130

# Step 3: Calculate sample proportion
p_hat <- x / n

# Step 4: Calculate the standard error of the proportion
se <- sqrt(p_hat * (1 - p_hat) / n)

# Step 5: Calculate the test statistic
p_null <- 0.70  # Assumed population proportion under the null hypothesis
z_score <- (p_hat - p_null) / se

# Step 6: Determine the p-value
p_value <- 2 * pnorm(abs(z_score), lower.tail = FALSE)  # Two-sided test

# Step 7: Interpret the results
alpha <- 0.05

if (p_value < alpha) {
  cat("There is statistical evidence that the stress percentage of
      college students in the country is different from 70% (p-value =", p_value, "<", alpha, ")\n")
} else {
  cat("There is no statistical evidence to conclude that the stress
      percentage of college students in the country is different from
      70% (p-value =", p_value, ">= ", alpha, ")\n")
}
```

```
## There is no statistical evidence to conclude that the stress
##       percentage of college students in the country is different from
##       70% (p-value = 0.1382077 >=  0.05 )
```

Problem 3

```r
# Step 1: Define hypotheses
# H0: Mean data usage for age > 50 = Mean data usage for age 18-50
# H1: Mean data usage for age > 50 < Mean data usage for age 18-50

# Step 2: Data
n1 <- 0.7 * 500  # Number of customers aged 18-50
n2 <- 0.3 * 500  # Number of customers aged > 50

mean_18_50 <- 2  # Average data usage for age 18-50
mean_above_50 <- 1.85  # Average data usage for age > 50

sd_18_50 <- 0.812  # Sample standard deviation for age 18-50
```

```r
sd_above_50 <- 0.837  # Sample standard deviation for age > 50

# Step 3: Calculate the standard error of the difference in means
se_diff <- sqrt((sd_18_50^2 / n1) + (sd_above_50^2 / n2))

# Step 4: Calculate the test statistic
mean_diff <- mean_18_50 - mean_above_50
t_value <- mean_diff / se_diff

# Step 5: Determine the p-value
p_value <- pt(t_value, df = n1 + n2 - 2, lower.tail = TRUE)

# Step 6: Interpret the results
alpha <- 0.05  # Significance level

if (p_value < alpha) {
  cat("There is statistical evidence that people older than 50
      years use less cell phone data than people aged 18-50 (p-value =", p_value, "<", alpha, ")\n")
} else {
  cat("There is no statistical evidence to conclude that people older
      than 50 years use less cell phone data than people aged 18-50 (p-value =", p_value, ">= ", alpha,
}
```

```
## There is no statistical evidence to conclude that people older
##       than 50 years use less cell phone data than people aged 18-50 (p-value = 0.9677487 >=  0.05 )
```

Problem 4

```r
# Step 1: Define hypotheses
# H0: Proportion of returns for Apple = Proportion of returns for Samsung
# H1: Proportion of returns for Apple < Proportion of returns for Samsung

# Step 2: Calculate sample proportions
n_apple <- 150  # Number of Apple iPhones sold
n_samsung <- 125  # Number of Samsung Galaxy phones sold

returns_apple <- 14  # Number of returns for Apple iPhones
returns_samsung <- 15  # Number of returns for Samsung Galaxy phones

p_apple <- returns_apple / n_apple  # Sample proportion of returns for Apple
p_samsung <- returns_samsung / n_samsung  # Sample proportion of returns for Samsung

# Step 3: Calculate the overall pooled proportion
p_pool <- (returns_apple + returns_samsung) / (n_apple + n_samsung)

# Step 4: Calculate the standard error of the difference in proportions
se_diff <- sqrt(p_pool * (1 - p_pool) * (1/n_apple + 1/n_samsung))

# Step 5: Calculate the test statistic
z_score <- (p_apple - p_samsung) / se_diff
```

```r
# Step 6: Determine the p-value
p_value <- pnorm(z_score, lower.tail = TRUE)

# Step 7: Interpret the results
alpha <- 0.05  # Significance level

if (p_value < alpha) {
  cat("There is statistical evidence that Apple has a
      smaller chance of being returned than Samsung (p-value =", p_value, "<", alpha, ")\n")
} else {
  cat("There is no statistical evidence to conclude that Apple has
      a smaller chance of being returned than Samsung (p-value =", p_value, ">= ", alpha, ")\n")
}
```

```
## There is no statistical evidence to conclude that Apple has
##       a smaller chance of being returned than Samsung (p-value = 0.2367125 >=  0.05 )
```

Problem 5

```r
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)

# Step 1: Define hypotheses
# H0: Mean mother's age = Mean father's age
# H1: Mean mother's age is not Mean father's age

# Step 2: Data
mother_age <- babies$age
father_age <- babies$dage

# Step 3: Calculate sample means
mean_mother_age <- mean(mother_age)
mean_father_age <- mean(father_age)

# Step 4: Calculate sample standard deviations
```

```r
sd_mother_age <- sd(mother_age)
sd_father_age <- sd(father_age)

# Step 5: Calculate the standard error of the difference in means
n_mother <- length(mother_age)
n_father <- length(father_age)
se_diff <- sqrt((sd_mother_age^2 / n_mother) + (sd_father_age^2 / n_father))

# Step 6: Calculate the test statistic
mean_diff <- mean_mother_age - mean_father_age
t_value <- mean_diff / se_diff

# Step 7: Determine the p-value
p_value <- 2 * pt(abs(t_value), df = n_mother + n_father - 2, lower.tail = FALSE)

# Interpret the results
alpha <- 0.05  # Significance level

if (p_value < alpha) {
  cat("There is statistical evidence that the ages of mothers
      and fathers in the sampled population are different (p-value =", p_value, "<", alpha, ")\n")
} else {
  cat("There is no statistical evidence to conclude that the
      ages of mothers and fathers in the sampled population are different (p-value =", p_value, ">= ", 
}
```

```
## There is statistical evidence that the ages of mothers
##       and fathers in the sampled population are different (p-value = 8.071277e-28 < 0.05 )
```