Homework 8

Kevin Jin

4/24/2022

Problem 1

Step 1: Specify a model.

Proportion of medically insured adults: $X \sim Bern(\omega)$

Step 2: Identify null (H_0) and alternative (H_α) hypotheses.

$H_0 : \omega = \omega_0 = 0.281$

$H_\alpha : \omega \neq 0.281$

Step 3: Specify a test statistic.

$z = \frac{\bar{X} - \omega_0}{\sqrt{\frac{\omega_0(1-\omega_0)}{n}}} \sim N(0, 1)$

Step 4: Compute the observed value of the test statistic.

$\bar{X} = \frac{30}{75} = 0.4$

$z = \frac{0.4 - 0.281}{\sqrt{\frac{0.281(1-0.281)}{75}}} = 0.292767347$

Step 5: Compute the p-value under H_0, where $p = P$.

```
x_bar <- 30 / 75
w_0 <- 0.281
n <- 75
z <- (x_bar - w_0) / sqrt(w_0 * (1 - w_0) / n)

2 * (1 - pnorm(abs(z), mean = 0, sd = 1))
```

```
## [1] 0.0218614
```

Step 6: Specify a significance level α.

We select the p-value range of $(0.01, 0.05]$, making it statistically significant and placing it within significance level $\alpha = 0.05$.

Step 7: Compare the p-value and the significance level α.

The p-value $= 0.0218614 < \alpha = 0.05$, indicating that we can reject H_0; therefore, there is a statistically significant difference between the proportion of college graduates age 21-24 not medically insured and the nationwide proportion of adults not medically insured.

```
prop.test(30, 75, p = 0.281, alternative = "two.sided")
```

```
##
##  1-sample proportions test with continuity correction
##
```

```
## data:  30 out of 75, null probability 0.281
## X-squared = 4.6843, df = 1, p-value = 0.03044
## alternative hypothesis: true p is not equal to 0.281
## 95 percent confidence interval:
##  0.2905787 0.5197370
## sample estimates:
##   p
## 0.4
```

Problem 2

Step 1: Specify a model.

Proportion of iPhone returns: $X \sim Bern(\omega_x)$

Proportion of Galaxy returns: $Y \sim Bern(\omega_y)$

Step 2: Identify null (H_0) and alternative (H_α) hypotheses.

$H_0 : \omega_x - \omega_y = 0$

$H_\alpha : \omega_x - \omega_y < 0$

Step 3: Specify a test statistic.

$z = \frac{\bar{X}-\bar{Y}}{\sqrt{\hat{\omega}(1-\hat{\omega})(\frac{1}{n_x}+\frac{1}{n_y})}} \sim N(0,1)$

$\hat{\omega} = \frac{n_x\bar{x}+n_y\bar{y}}{n_x+n_y}$

Step 4: Compute the observed value of the test statistic.

$\bar{X} = \frac{14}{150} = 0.09\bar{3}$

$\bar{Y} = \frac{15}{125} = 0.12$

$\hat{\omega} = \frac{150*0.09\bar{3}+125*0.12}{150+125} = 0.105\bar{45}$

$z = \frac{0.09\bar{3}-0.12}{\sqrt{0.105\bar{45}(1-0.105\bar{45})(\frac{1}{150}+\frac{1}{125})}} = -0.716916$

Step 5: Compute the p-value under H_0, where $p = P$.

```r
n_x <- 150
n_y <- 125
x_bar <- 14/n_x
y_bar <- 15/n_y
w_hat <- (n_x * x_bar + n_y * y_bar) / (n_x + n_y)
z <- (x_bar - y_bar) / sqrt(w_hat * (1 - w_hat) * (1 / n_x + 1 / n_y))

pnorm(z, mean = 0, sd = 1)
```

```
## [1] 0.2367125
```

Step 6: Specify a significance level α.

We select the p-value range of $(0.01, 0.05]$, making it statistically significant and placing it within significance level $\alpha = 0.05$.

Step 7: Compare the p-value and the significance level α.

The p-value $= 0.2367125 > \alpha = 0.05$, indicating that we cannot reject H_0; therefore, we cannot conclude that there is statistical evidence that iPhones have a smaller chance of being returned than Galaxys.

```r
prop.test(c(14, 15), c(150, 125), alternative = "less")
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(14, 15) out of c(150, 125)
## X-squared = 0.27016, df = 1, p-value = 0.3016
## alternative hypothesis: less
## 95 percent confidence interval:
##  -1.00000000  0.04240782
## sample estimates:
##     prop 1     prop 2
## 0.09333333 0.12000000
```

Problem 3

Step 1: Specify a model.

Fathers' ages: $X \sim N(\mu_x, \sigma_x^2)$

Mothers' ages: $Y \sim N(\mu_y, \sigma_y^2)$

Step 2: Identify null (H_0) and alternative (H_α) hypotheses.

$H_0 : \mu_x - \mu_y = 0$

$H_\alpha : \mu_x - \mu_y < 0$

Step 3: Specify a test statistic.

$t = \frac{|\bar{x}|}{\sqrt{\frac{s^2}{n}}}$

Step 4: Compute the observed value of the test statistic.

```r
# Compute the mean difference between the two sample populations
library(UsingR)
```

```
## Loading required package: MASS

## Loading required package: HistData

## Loading required package: Hmisc

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

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

##
## Attaching package: 'UsingR'
```

```
## The following object is masked from 'package:survival':
##
##     cancer
```

```
data(babies)
diff <- babies$age - babies$dage
n <- length(babies$age - babies$dage) # n = 1236
x_bar <- mean(diff) # mean = -3.365696
sd <- sd(diff) # sd = 6.803471
```

$t = \frac{|-3.365696|}{\sqrt{\frac{6.803471^2}{1236}}} = 17.3922$

Step 5: Compute the p-value under H_0, where $p = P$.

```
t <- 17.3922
n <- length(babies$age - babies$dage)
2 * pt(abs(t), df = n - 1, lower.tail = FALSE) # p-value for paired t-test
```

```
## [1] 9.029009e-61
```

Step 6: Specify a significance level α.

We select the p-value range of $(0.01, 0.05]$, making it statistically significant and placing it within significance level $\alpha = 0.05$.

Step 7: Compare the p-value and the significance level α. The p-value $= 9.029009 * 10^{-61} < \alpha = 0.05$, indicating that we can reject H_0; therefore, we conclude that there is a statistically significant difference between the mothers' ages and the fathers' ages.

```
t.test(babies$age, babies$dage, paired = TRUE, alternative = "two.sided")
```

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