Homework 8

Luke Bennett

12/3/2021

Problem 1

Is there a statistically significant difference between the male and female responses?

Step 1:

$$X \sim \text{Bern}[\omega_x = 0.32]$$
$$Y \sim \text{Bern}[\omega_y = 0.37]$$

Step 2:

$$H_0 : \omega_x - \omega_y = 0$$
$$H_a : \omega_x - \omega_y \neq 0$$

Test Statistic:

$$z = \frac{\omega_x - \omega_y}{\sqrt{\frac{\omega_x(1-\omega_x)}{n_x} + \frac{\omega_y(1-\omega_y)}{n_y}}}$$

Evaluate

$$z = \frac{0.32 - 0.37}{\sqrt{\frac{0.32(1-0.32)}{540} + \frac{0.37(1-0.37)}{560}}} = -1.75$$

State $\alpha = 0.05$

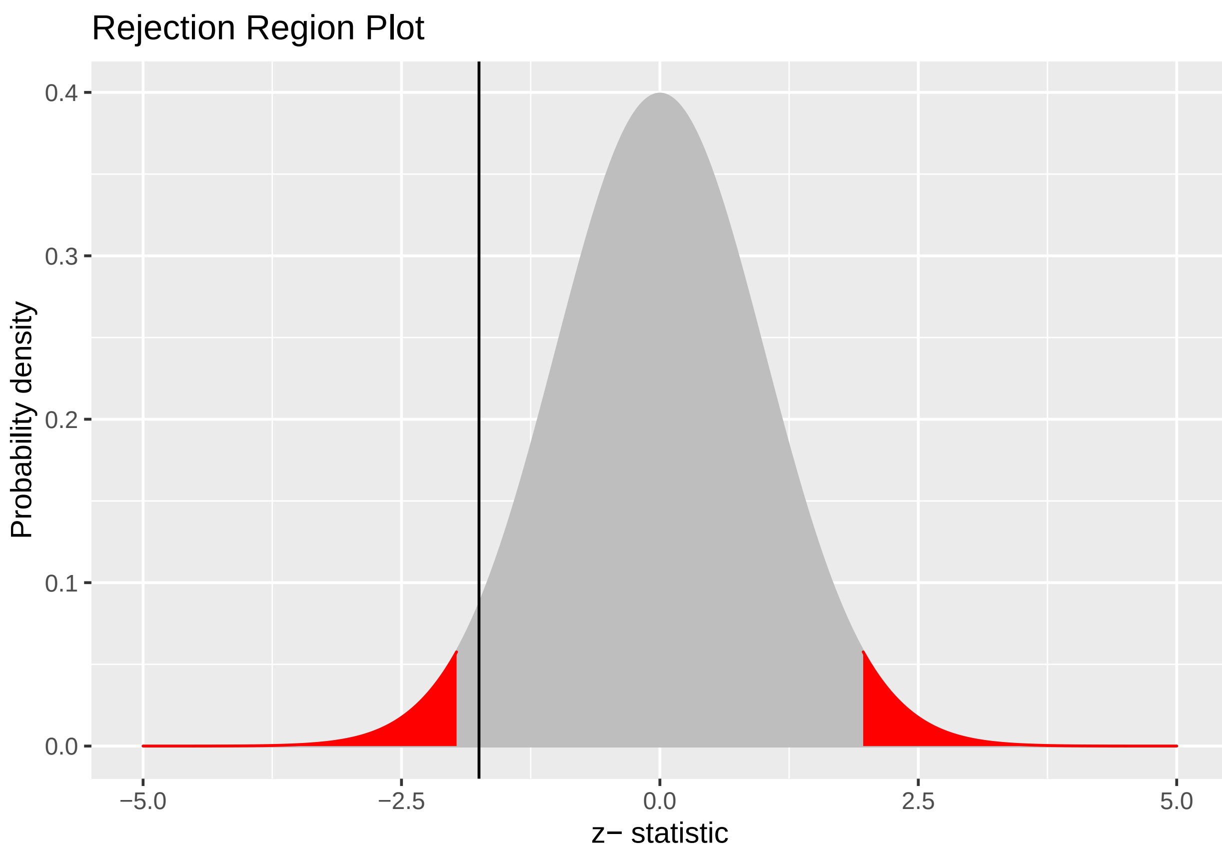
Compute p-value under H_0

$$p = 2(1 - \mathbf{P}(z \geq |-1.75|)) = 0.08$$

Decision The test is not statistically significant because the test statistic falls within the acceptance region, and so we can conclude that males and females are equally as bother when someone is one their phone during an interaction. See graph.

```
x <- seq(-5, 5, length = 1000)
pdf <- dnorm(x, mean = 0, sd = 1)
theoretical <- data.frame (x = x, pdf = pdf)
z <- -1.75
library(ggplot2)
ggplot() + geom_polygon(data = theoretical,
                        mapping = aes (x = x, y = pdf), color = "grey",
                        fill = "grey") +
          geom_area(data = subset(theoretical, x <= -1.96),
                    mapping = aes (x = x, y = pdf ), color = "red",
                    fill = "red") +
          geom_area(data = subset (theoretical, x >= 1.96),
                    mapping = aes(x = x, y = pdf), color = "red",
```

```r
               fill = "red") +
    labs(x = "z- statistic", y = "Probability density") +
    geom_vline(xintercept = z) + ggtitle("Rejection Region Plot")
```



Problem 2

Determine if the variable Height is normally distributed. Use an appropriate test at $\alpha = 0.05$.

$$H_0 : \text{Height variable is normally distributed}$$

$$H_a : \text{Height variable is not normally distributed}$$

```r
shapiro.test(trees$Height)
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  trees$Height
## W = 0.96545, p-value = 0.4034
```

p value is $.40 \geq 0.05 = \alpha$

Conclude that Height variable is Normally distributed.

We can use Pearson Correlation with another normal RV

```r
cor.test(trees$Height, trees$Girth)
```

```
##
##  Pearson's product-moment correlation
##
## data:  trees$Height and trees$Girth
## t = 3.2722, df = 29, p-value = 0.002758
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2021327 0.7378538
## sample estimates:
##       cor
## 0.5192801
```

$p = 0.0028 < \alpha = 0.05$ so the correlation is significant. Reject null hypothesis.

```r
t.test(trees$Height, mu = 70, alternative = "g")
```

```
##
##  One Sample t-test
##
## data:  trees$Height
## t = 5.2429, df = 30, p-value = 5.866e-06
## alternative hypothesis: true mean is greater than 70
## 95 percent confidence interval:
##  74.05764      Inf
## sample estimates:
## mean of x
##        76
```

p value is close to zero smaller than $\alpha \Rightarrow$ reject H_0