Statistical Methods - Assignment 3

Michel Mooiweer (1866761) Thomas Webbers (2560695) Eirik Kultorp (2544992)

December 2016

Exercise 3.1

$$n = \frac{[z_{\alpha/2}]^2 \hat{p}\hat{q}}{E^2}$$

Sample Proportion $\hat{p} = 0.15$

$1 - \hat{p} = \hat{q} = 0.85$

Significance level $\alpha = 0.05$

$\alpha/2 = 0.025$

Critical value: $z_{\alpha/2} = 1.96$

$E = 0.03$

$$n = \frac{1.96^2 \cdot 0.15 \cdot 0.85}{0.03^2}$$
$$n = 544.226 \rightarrow 545$$

Sample size needed is 545

Exercise 3.3

a)

```
## [1] "Point estimate :  0.2024 (Alice works 0.2024 hours more than Bob)"
## [1] "Confidence interval (0.95) : -0.057038284558802 : 0.461838284558803"
```

b)

We do a t-test because we're testing for difference between population means. The conditions are also met:

- The sampling method for each sample is simple random sampling
- The samples are independent.
- Each population is at least 20 times larger than its respective sample. (We don't know this, but we assume it)
- The sampling distribution is approximately normal and the sample size is greater than 40, without outliers.

```
##
##  Welch Two Sample t-test
##
## data:  Alice and Bob
```

```
## t = 1.5495, df = 91.778, p-value = 0.1247
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.05703828  0.46183828
## sample estimates:
## mean of x mean of y
##    3.9416    3.7392
```

- null hypothesis: true difference in means is equal to 0
- alternative hypothesis: true difference in means is not equal to 0
- significance level: 5%
- test statistic: difference in means
- observed value: 0.2024
- p-value: 0.1247

We fail to reject the null hypothesis because p>0.05. In other words, we can't disprove the manager's claim that Alice and Bob work the same amount of hours. We could also have reached the same conclusion just by looking at the confidence interval, which includes 0 (indicating a test statistic of 0 would not be unexpected with this significance level).

c)

We do a t-test for the same reasons as above. This time we test the claim that Alice works less than Bob

```
##
##  Welch Two Sample t-test
##
## data:  Alice and Bob
## t = 1.5495, df = 91.778, p-value = 0.06235
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.01464775         Inf
## sample estimates:
## mean of x mean of y
##    3.9416    3.7392
```

- null hypothesis: true difference in means is less than or equal to 0
- alternative hypothesis: true difference in means is greater than 0
- significance level: 5%
- test statistic: difference in means
- observed value: 0.2024
- p-value: 0.06235

We fail to reject the null hypothesis because p>0.05. In other words, we can't confirm Alice's claim that she works more than Bob. We could also have reached the same conclusion just by looking at the confidence interval, which has `<0` on the left side (indicating a test statistic of `<0` would not be unexpected with this significance level).

d)

```
0.1247/0.06235==2
```

Huh.

3.4

a)

```
## [1] 0.16
```

b)

We perform a test for population proportions.

```
##    T  F
## A 40 10
## B 32 18
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  d
## X-squared = 2.4306, df = 1, p-value = 0.0595
## alternative hypothesis: greater
## 95 percent confidence interval:
##  -0.005344017  1.000000000
## sample estimates:
## prop 1 prop 2
##   0.80   0.64
```

- null hypothesis: true difference in proportions is less than or equal to 0
- alternative hypothesis: true difference in proportions is greater than 0
- significance level: 5%
- test statistic: difference in means
- observed value: 0.16
- p-value: 0.0595

We fail to reject the null hypothesis because p>0.05. In other words, we can't confirm Alice's claim claim that the proportion of evenings on which she worked more than 3.5 hours is larger than the proportion of evenings on which Bob worked more than 3.5 hours. We could also have reached the same conclusion just by looking at the confidence interval, which has `<0` on the left side (indicating a test statistic of `<0` would not be unexpected with this significance level).