Durham College DATA1204 Assignment2

Chinedu Onyeka

10/20/2021

Load the required libraries

```
# Load required libraries
library(tidyverse)
```

Steps for Hypothesis testing:
1. State the Null and Alternative hypothesis.
2. Check conditions for independence and normality.
3. Compute mean, standard, deviation of each group.
4. Compute the test statistic.
5. Compute the p-value.
6. Reject or Accept null hypothesis based on p-value and level of significance.

```
# Read the BloodPressure file
url <- "https://raw.githubusercontent.com/chinedu2301/DC_Analytics/main/BloodPressure.csv"
blood_pressure <- read_csv(url)
```

```
# check the head of the blood pressure dataset
head(blood_pressure)
```

```
## # A tibble: 6 x 3
##   Subject Before After
##     <dbl>  <dbl> <dbl>
## 1       1    135   127
## 2       2    142   145
## 3       3    137   131
## 4       4    122   125
## 5       5    147   132
## 6       6    151   147
```

Compute relevant statistics

```
mean_before <- mean(blood_pressure$Before)
mean_after <- mean(blood_pressure$After)
std_before <- round(sd(blood_pressure$Before), 2)
std_after <- round(sd(blood_pressure$After), 2)
n_before <- nrow(blood_pressure)
n_after <- nrow(blood_pressure)
SE_before <- round(std_before/sqrt(n_before), 2)
SE_after <- round(std_after/sqrt(n_after), 2)
```

```r
# Put all the values in a dataframe
headers <- c("Before", "After")
row_names <- c("Mean", "Std", "n", "SE")
bp_mean <- c(mean_before, mean_after)
bp_std <- c(std_before, std_after)
bp_n <- c(n_before, n_after)
bp_se <- c(SE_before, SE_after)
bp_table <- data.frame(bp_mean, bp_std, bp_n, bp_se)
bp_table <- t.data.frame(bp_table) # transpose the dataframe to make the table more intuitive
row.names(bp_table) <- row_names # give more intuitive row names
colnames(bp_table) <- headers # give more intuitive header names
bp_table
```

```
##      Before  After
## Mean 138.28 130.28
## Std   11.70   7.96
## n     25.00  25.00
## SE     2.34   1.59
```

Hypothesis Testing:

Null Hypothesis, H_0: There is no difference in the blood pressure before and after.$\mu_{after} - \mu_{before} = 0$
Alternative Hypothesis, H_1: There is difference in the blood pressure before and after.$\mu_{after} - \mu_{before} \neq 0$

Check conditions:
Independence: The samples come from a simple random sample. Hence, Independence is satisfied.
Normality: The sample size is large enough we can move forward with applying the t-distribution for this problem.

Compute Test Statistics:
$SE = \sqrt{\frac{(std_{before})^2}{n_{before}} + \frac{(std_{after})^2}{n_{after}}}$

Test statistic, $t = \frac{\bar{x}_{diff} - \mu_{diff}}{SE_{diff}}$

```r
# compute the t statistic and the p_value
alpha <- 0.05 # level of confidence
df <- n_before + n_after - 2
mean_diff <- mean_after - mean_before
mu_diff <- 0
SE_diff <- sqrt(SE_before**2 + SE_after**2)
t <- (mean_diff - mu_diff)/SE_diff
p_value <- round(2*pt(t, df, lower.tail = FALSE), 4)
paste0("Since the p value is ", p_value, " which is greater than ", alpha,
       ", we do not reject the null hypothesis")
```

```
## [1] "Since the p value is 1.9932 which is greater than 0.05, we do not reject the null hypothesis"
```

Conclusion:
From the hypothesis testing of the difference in means of the blood pressure of before and after, we can infer that there is insufficient statistical evidence to reject the null hypothesis that there is no difference in blood pressure before and after. Hence, we accept the null hypothesis at 0.05 level of significance that there is no difference in mean blood pressure of an individual before and after exercising regularly for the past 30 -days. i.e. An individual's blood pressure has not changed significantly after exercising regularly for the past 30 -days.