

My_Rmarkdown

My_assignment

2024-08-17

Load the required libraries for the analysis

loading the dataset to be used into R document

```
library(tidyverse,quietly = TRUE)
```

```
## Warning: package 'tidyverse' was built under R version 4.2.3
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
## Warning: package 'tibble' was built under R version 4.2.3
```

```
## Warning: package 'tidyr' was built under R version 4.2.3
```

```
## Warning: package 'readr' was built under R version 4.2.3
```

```
## Warning: package 'purrr' was built under R version 4.2.3
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
## Warning: package 'forcats' was built under R version 4.2.3
```

```
## Warning: package 'lubridate' was built under R version 4.2.3
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
## v dplyr      1.1.2      v readr      2.1.4
```

```
## v forcats    1.0.0      v stringr   1.5.0
```

```
## v ggplot2    3.4.3      v tibble    3.2.1
```

```
## v lubridate  1.9.2      v tidyr     1.3.0
```

```
## v purrr      1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
labs<- read.csv("Labs.csv", header=TRUE, colClasses=c("integer","factor","numeric"))
labs
```

```
##      id treat postwt
```

```
## 1     1     1   80.2
```

```
## 2     2     1   80.1
```

```
## 3     3     1   86.4
```

```
## 4     4     1   86.3
```

```
## 5     5     1   76.1
```

```
## 6     6     1   78.1
```

```
## 7     7     1   75.1
```

```
## 8     8     1   86.7
```

```
## 9     9     1   73.5
```

```
## 10    10     1   84.6
```

##	11	11	1	77.4
##	12	12	1	79.5
##	13	13	1	89.6
##	14	14	1	81.4
##	15	15	1	81.8
##	16	16	1	77.3
##	17	17	1	84.2
##	18	18	1	75.4
##	19	19	1	79.5
##	20	20	1	73.0
##	21	21	1	88.3
##	22	22	1	84.7
##	23	23	1	81.4
##	24	24	1	81.2
##	25	25	1	88.2
##	26	26	1	78.8
##	27	27	2	82.2
##	28	28	2	85.6
##	29	29	2	81.4
##	30	30	2	81.9
##	31	31	2	76.4
##	32	32	2	103.6
##	33	33	2	98.4
##	34	34	2	93.4
##	35	35	2	73.4
##	36	36	2	82.1
##	37	37	2	96.7
##	38	38	2	95.3
##	39	39	2	82.4
##	40	40	2	72.5
##	41	41	2	90.9
##	42	42	2	71.3
##	43	43	2	85.4
##	44	44	2	81.6
##	45	45	2	89.1
##	46	46	2	83.9
##	47	47	2	82.7
##	48	48	2	75.7
##	49	49	2	82.6
##	50	50	2	100.4
##	51	51	2	85.2
##	52	52	2	83.6
##	53	53	2	84.6
##	54	54	2	96.2
##	55	55	2	86.7
##	56	56	3	95.2
##	57	57	3	94.3
##	58	58	3	91.5
##	59	59	3	91.9
##	60	60	3	100.3
##	61	61	3	76.7
##	62	62	3	76.8
##	63	63	3	101.6
##	64	64	3	94.9

```
## 65 65      3   75.2
## 66 66      3   77.8
## 67 67      3   95.5
## 68 68      3   90.7
## 69 69      3   92.5
## 70 70      3   93.8
## 71 71      3   91.7
## 72 72      3   98.0
```

inspecting the data for the missing values

```
head(labs)
```

```
##   id treat postwt
## 1  1     1   80.2
## 2  2     1   80.1
## 3  3     1   86.4
## 4  4     1   86.3
## 5  5     1   76.1
## 6  6     1   78.1
```

```
tail(labs)
```

```
##   id treat postwt
## 67 67     3   95.5
## 68 68     3   90.7
## 69 69     3   92.5
## 70 70     3   93.8
## 71 71     3   91.7
## 72 72     3   98.0
```

```
summary(is.na(labs))
```

```
##      id      treat      postwt
## Mode :logical Mode :logical Mode :logical
## FALSE:72      FALSE:72      FALSE:72
```

from the above code missing values are not present in the dataset meaning the dataset is clean and tidy and ready for the analysis.

The code below gives the number of patients who received different treatments

```
table(labs$treat)
```

```
##
##  1  2  3
## 26 29 17
```

from the output there were 72 patients in which 26 received treatment 1, 29 received treatment 2, 17 received treatment 3

Calculating different descriptive statistics for the sample that is the means, sd and median for the different treatments

```
labs %>%
  group_by(treat) %>%
  summarize(mean=mean(postwt), median=median(postwt),
            sd=sd(postwt))
```

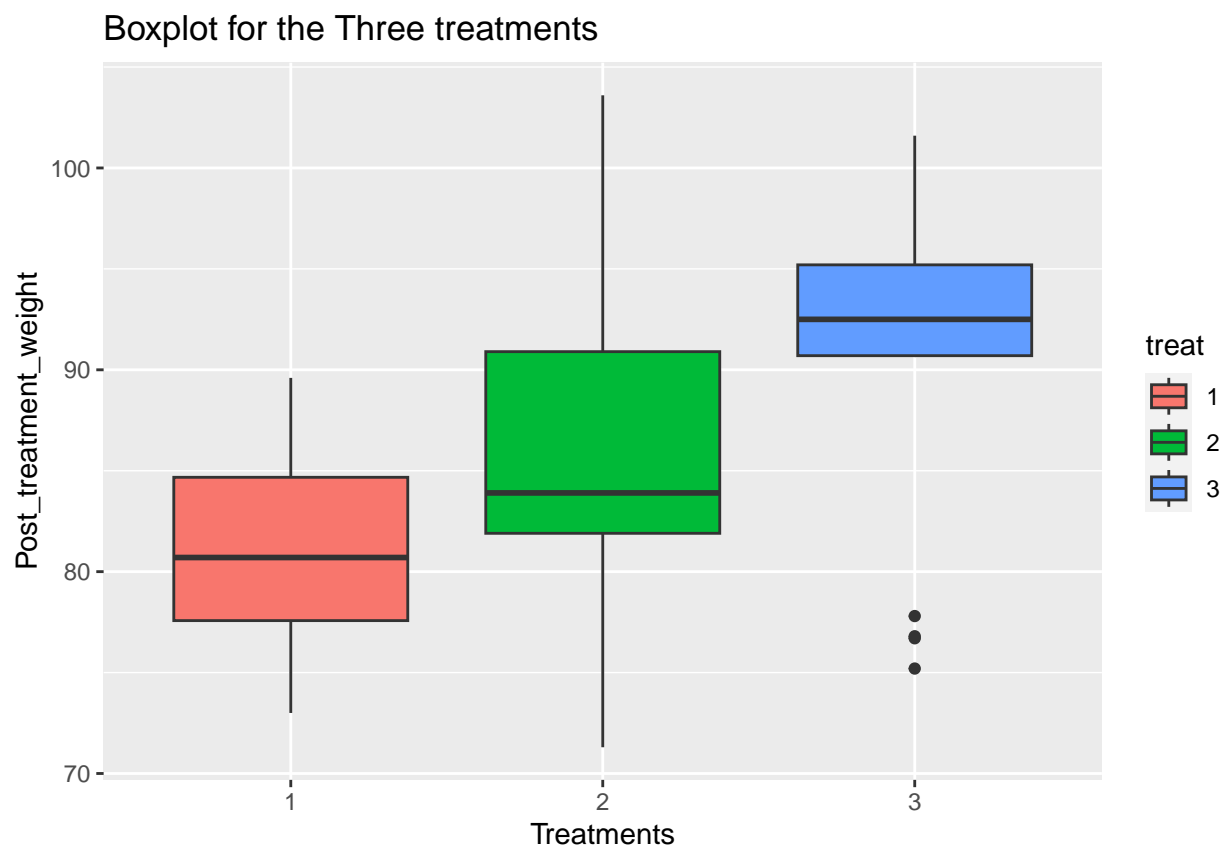
```
## # A tibble: 3 x 4
```

```
##   treat  mean median   sd
##   <fct> <dbl> <dbl> <dbl>
## 1 1      81.1  80.7  4.74
## 2 2      85.7  83.9  8.35
## 3 3      90.5  92.5  8.48
```

From the output the patients that received treatment 1 had a mean of 81.10769, median of 80.7 and standard deviation of 4.7... and for the rest of treatment as show above

Constructing the boxplots to compare the post treatment weights for the three treatments

```
ggplot(labs, aes(x=treat, y=postwt, fill=treat))+
  geom_boxplot()+
  ggtitle("Boxplot for the Three treatments")+
  xlab("Treatments")+
  ylab("Post_treatment_weight")
```



From the graph the control group had a median value slightly above 80 while cbt group had a median weight of almost 85 and the family therapy group has a median post weight of above 90 In the cbt group the IQR is larger than that of control group and Family therapy group.

```
alpha <- 0.05
```

Perfoming the T_Test for the control group which is treatment1

H0: $\mu = 79$ H1: $\mu > 79$

```
#Filter data for Control group
control_group <- labs %>% filter(treat == 1)
```

```
# Perform the upper-tail T-test
t_test_upper <- t.test(control_group$postwt, mu = 79, alternative = "greater")
t_test_upper
```

```
##
## One Sample t-test
##
## data: control_group$postwt
## t = 2.2653, df = 25, p-value = 0.01621
## alternative hypothesis: true mean is greater than 79
## 95 percent confidence interval:
## 79.5184 Inf
## sample estimates:
## mean of x
## 81.10769
```

```
p_value1 <- t_test_upper$p.value
if(p_value1 <= alpha){
  print("Reject H0.")
}else{
  print("Fail to Reject H0")
}
```

```
## [1] "Reject H0."
```

In testing the above hypothesis the alpha value was 0.05, from the output the t_vale is 2.2653 and the sample has a degree of freedom of 25, since our calculated p_value 0.01621 which is less than our significant level which is 0.05 we reject null hypothesis which suggest that there is enough evidence that the mean post_treat_weight is greater than 79 lbs

H0: $\mu = 83$ H1: $\mu \neq 83$

```
# Perform the two-sided T-test
t_test_two_tail <- t.test(control_group$postwt, mu = 83, alternative = "two.sided")
t_test_two_tail
```

```
##
## One Sample t-test
##
## data: control_group$postwt
## t = -2.0338, df = 25, p-value = 0.05271
## alternative hypothesis: true mean is not equal to 83
## 95 percent confidence interval:
## 79.19145 83.02394
## sample estimates:
## mean of x
## 81.10769
```

```
p_value2 <- t_test_two_tail$p.value
if(p_value2 <= alpha){
  print("Reject H0.")
}else{
  print("Fail to Reject H0")
}
```

```
## [1] "Fail to Reject H0"
```

Since our calculated p_value 0.05271 is greater than our alpha which is 0.05 we fail to reject null hypothesis

which suggest that the mean post_treatment_weight is not equal to 83 meaning it can either be below 83 or above 83 for the control treatment group.

for the Two_sample T_test we set our null and alternative hypothesis as follows H0: mean post_treatment_weight for control group is equal to mean post_treatment_weight for cbt group. H1: The mean post_treatment_weight for control group and cbt group are not equal

```
# Filter data for CBT group
cbt_group <- labs %>% filter(treat == 2)

# Perform the lower-tail T-test
t_test_independent <- t.test(control_group$postwt, cbt_group$postwt,
                             alternative = "less", var.equal = FALSE)
t_test_independent
```

```
##
## Welch Two Sample t-test
##
## data: control_group$postwt and cbt_group$postwt
## t = -2.5372, df = 45.221, p-value = 0.007345
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -1.551765
## sample estimates:
## mean of x mean of y
## 81.10769 85.69655

p_value3 <- t_test_independent$p.value
if(p_value3 <= alpha){
  print("Reject H0.")
}else{
  print("Fail to Reject H0")
}
```

```
## [1] "Reject H0."
```

since our calculated p_value 0.007345 which is less than our significant level which is 0.05 we reject null hypothesis this suggesting that there is enough evidence that the mean_post_treatment_weight for control treatment group and Cbt group are not equal.

Question 2

loading the lab2b dataset

```
lab2b<- read.csv("lab2b.csv", header=TRUE, colClasses=c("integer","factor","numeric"))
lab2b
```

```
##   id stage  wt
## 1   1     1 80.7
## 2   2     1 89.4
## 3   3     1 91.8
## 4   4     1 74.0
## 5   5     1 78.1
## 6   6     1 88.3
## 7   7     1 87.3
## 8   8     1 75.1
## 9   9     1 80.6
## 10 10     1 78.4
```

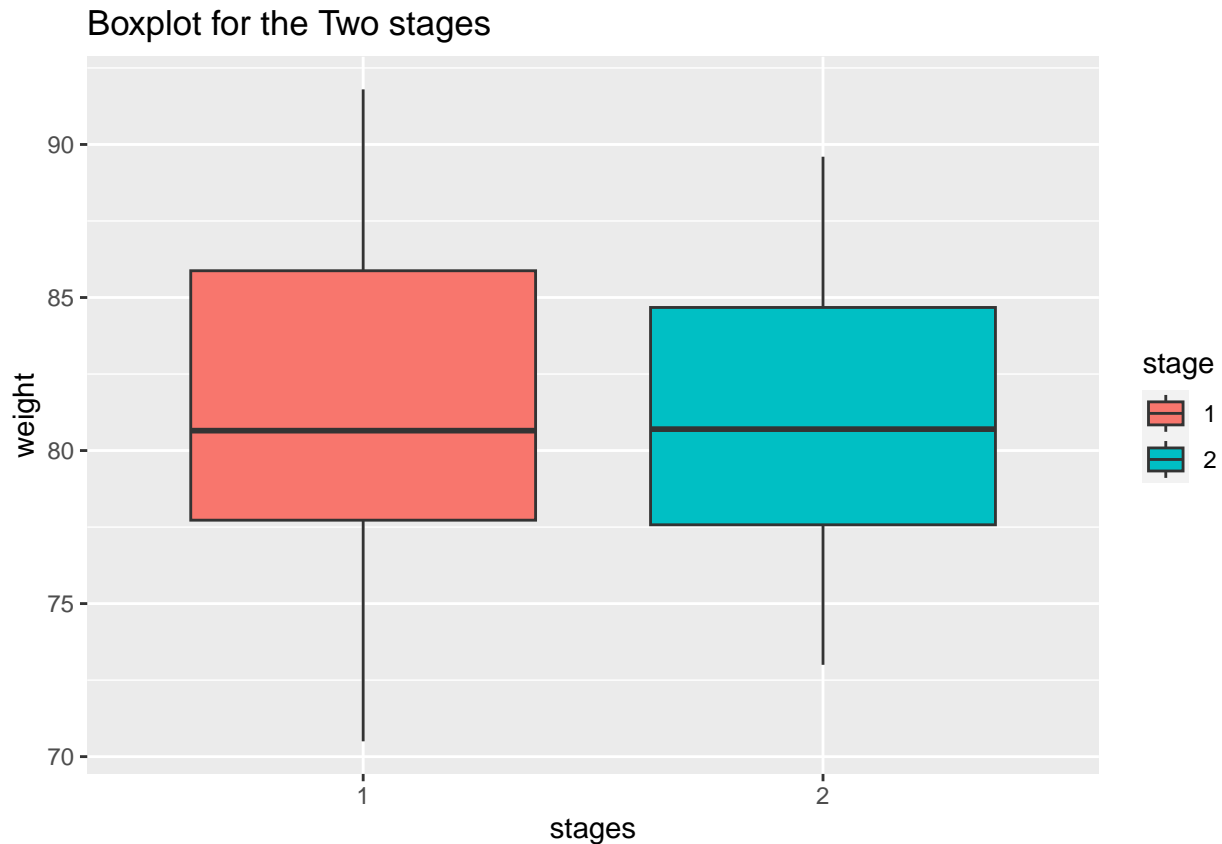
```
## 11 11      1 77.6
## 12 12      1 88.7
## 13 13      1 81.3
## 14 14      1 78.1
## 15 15      1 70.5
## 16 16      1 77.3
## 17 17      1 85.2
## 18 18      1 86.0
## 19 19      1 84.1
## 20 20      1 79.7
## 21 21      1 85.5
## 22 22      1 84.4
## 23 23      1 79.6
## 24 24      1 77.5
## 25 25      1 72.3
## 26 26      1 89.0
## 27  1      2 80.2
## 28  2      2 80.1
## 29  3      2 86.4
## 30  4      2 86.3
## 31  5      2 76.1
## 32  6      2 78.1
## 33  7      2 75.1
## 34  8      2 86.7
## 35  9      2 73.5
## 36 10      2 84.6
## 37 11      2 77.4
## 38 12      2 79.5
## 39 13      2 89.6
## 40 14      2 81.4
## 41 15      2 81.8
## 42 16      2 77.3
## 43 17      2 84.2
## 44 18      2 75.4
## 45 19      2 79.5
## 46 20      2 73.0
## 47 21      2 88.3
## 48 22      2 84.7
## 49 23      2 81.4
## 50 24      2 81.2
## 51 25      2 88.2
## 52 26      2 78.8
```

```
lab2b %>%
  group_by(stage) %>%
  summarize(mean=mean(wt),median=median(wt),
            sd=sd(wt))
```

```
## # A tibble: 2 x 4
##   stage mean median   sd
##   <fct> <dbl> <dbl> <dbl>
## 1 1      81.6   80.6  5.71
## 2 2      81.1   80.7  4.74
```

Boxplots showing wt for each stages.

```
ggplot(lab2b, aes(x=stage, y=wt, fill=stage))+
  geom_boxplot()+
  ggtitle("Boxplot for the Two stages")+
  xlab("stages")+
  ylab("weight")
```



From the box plots above the two stages has a mean of more than 80lb

```
# Filter data for each stage
stage_1 <- lab2b %>% filter(stage == 1)
stage_2 <- lab2b %>% filter(stage == 2)

# Perform the lower-tail T-test
t_test_stages <- t.test(stage_1$wt, stage_2$wt,
                        alternative = "less", var.equal = FALSE)
t_test_stages

##
## Welch Two Sample t-test
##
## data: stage_1$wt and stage_2$wt
## t = 0.30918, df = 48.385, p-value = 0.6207
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 2.890773
## sample estimates:
## mean of x mean of y
```



```
## 81.55769 81.10769
p_value4 <- t_test_stages$p.value
if(p_value4 <= alpha){
  print("Reject H0.")
}else{
  print("Fail to Reject H0")
}
```

```
## [1] "Fail to Reject H0"
```

Since our calculated `p_value` 0.6207 is greater than `alpha` which is 0.05 we fail to reject null hypothesis.