stats101B_HW1

Ryan Chu

2024-04-05

#Q1

1.

```r
meth_data
```

```
## # A tibble: 20 x 6
##    Name              City      Age Sex    Score1 Score2
##    <chr>             <chr>   <dbl> <chr>   <dbl>  <dbl>
##  1 Hailey Kimura     Akkeshi    19 Female      7      9
##  2 Aline Pasquier    Akkeshi    36 Female      2      6
##  3 Ayako Sakaguchi   Akkeshi    35 Female      9      8
##  4 Riley Hall        Akkeshi    57 Male        6      7
##  5 Momoko Connolly   Akkeshi    56 Female      7      9
##  6 Leif Bager        Akkeshi    24 Male       10      6
##  7 Mary Morris       Akkeshi    51 Female     10     10
##  8 Roger Aitken      Akkeshi    69 Male        7     10
##  9 Kakuji Yamasaki   Akkeshi    42 Male        9      9
## 10 Claire Lund       Akkeshi    42 Female      8      9
## 11 Josh Jackson      Akkeshi    43 Male        8      7
## 12 Minami Arai       Akkeshi    43 Female      1      1
## 13 Haruki Arai       Akkeshi    15 Male        3      2
## 14 Ian Edwards       Akkeshi    69 Male        8      8
## 15 Gunnar Solberg    Akkeshi    52 Male        8     10
## 16 Riley Morris      Akkeshi    50 Male        9      9
## 17 Anamica Brown     Akkeshi    49 Female      8      5
## 18 Allen Macdonald   Akkeshi    82 Male        9      9
## 19 Katharina Larsen  Akkeshi    21 Female      8      8
## 20 Rina Mardia       Akkeshi    41 Female      8     10
```

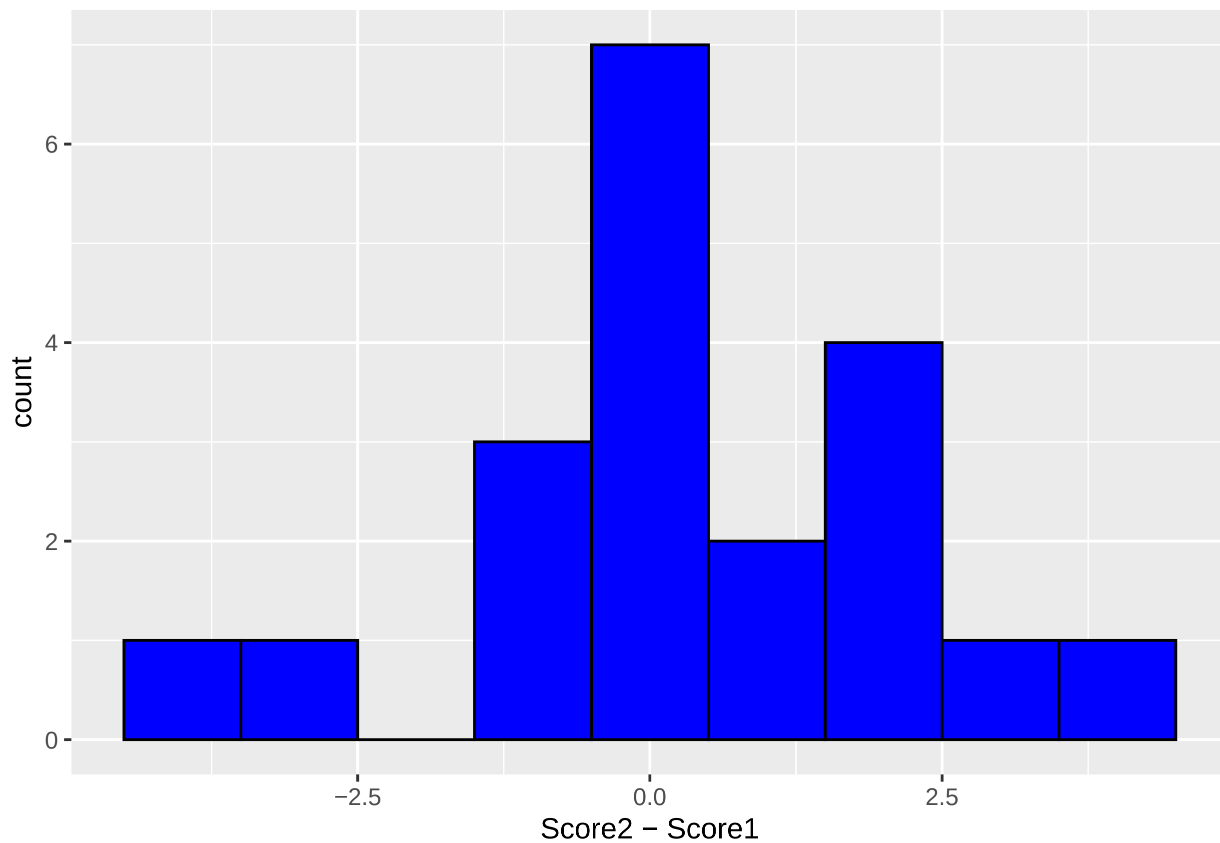
2.

```r
s1 <- meth_data$Score1
s2 <- meth_data$Score2
t.test(s1, s2, paired = TRUE, alternative = "two.sided")
```

```
##
```

```
##  Paired t-test
##
## data:  s1 and s2
## t = -0.82405, df = 19, p-value = 0.4201
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -1.238968  0.538968
## sample estimates:
## mean difference
##           -0.35
```

```
ggplot(meth_data, aes(x=Score2-Score1)) + geom_histogram(binwidth=1, color = "black", fill = "blue")
```

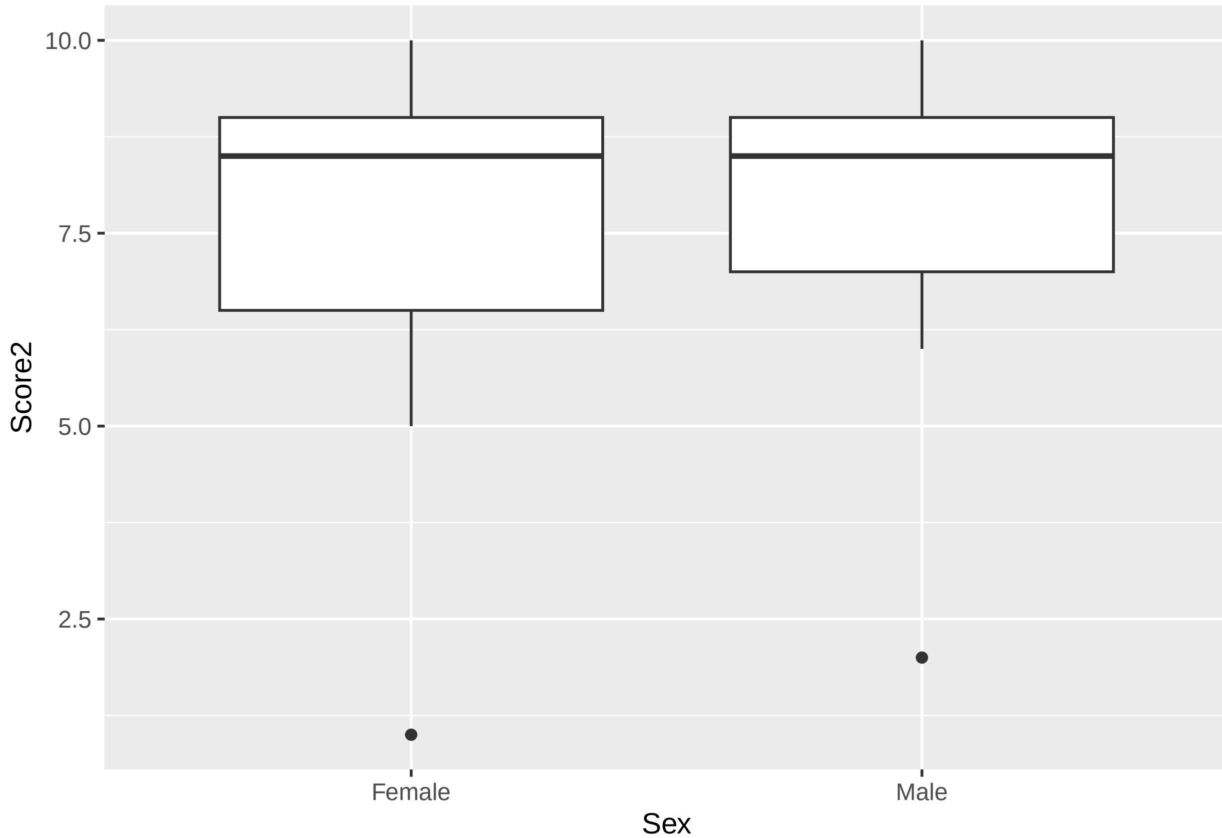


Ho: True mean difference is equal to 0 Ha: True mean difference is not equal to 0

Since the p-value(0.4201) is greater than 0.05, we fail to reject the null hypothesis.

3.

```
ggplot(meth_data, aes(x=Sex, y=Score2)) + geom_boxplot()
```



4.

```r
aggregate(data = meth_data, Score2~Sex, var)
```

```
##      Sex   Score2
## 1 Female 7.833333
## 2   Male 5.788889
```

```r
var.test(data=meth_data, Score2~Sex)
```

```
##
##  F test to compare two variances
##
## data:  Score2 by Sex
## F = 1.3532, num df = 9, denom df = 9, p-value = 0.6596
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.3361075 5.4478424
## sample estimates:
## ratio of variances
##           1.353167
```

From the var test we can conclude that there isn't a difference between the Female and Male variances.

5.

```r
t.test(Score2~Sex, data = meth_data, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  Score2 by Sex
## t = -0.17136, df = 18, p-value = 0.8659
## alternative hypothesis: true difference in means between group Female and group Male is not equal to
## 95 percent confidence interval:
##  -2.652076  2.252076
## sample estimates:
## mean in group Female   mean in group Male
##                  7.5                  7.7
```

Since our p-value is greater than 0.05, we fail to reject the null that the true difference in means between Females and Males is greater than 0.

Q2

a.

H0 = mean breaking strength = 150 HA = mean breaking strength <= 150

b.

```r
y <- c(145, 153, 150, 147)

(mean(y) - 150)/(2/2)
```

```
## [1] -1.25
```

```r
pnorm(-1.25, lower.tail = TRUE)
```

```
## [1] 0.1056498
```

Since the p-value is greater than 0.05, we fail to reject the null hypothesis.

c.

0.1056498

d.

```r
mean(y) + c(-1, 1) * (1.96 * 2)/2
```

```
## [1] 146.79 150.71
```

```r
mean(y)
```

```
## [1] 148.75
```

CI = 146.79 to 150.71

(mean of y is within this confidence interval, so makes sense that we failed to reject the null)

Q3

a.

```r
values <-  c(65, 81, 57, 66, 82, 82, 67, 59, 75, 70, 64, 71, 83, 59, 65, 56, 69, 74, 82, 79)
type <- c("type1", "type1", "type1", "type1", "type1", "type1", "type1", "type1", "type1", "type1", "ty

q3_df <- data.frame(values, type)

aggregate(values~type, data = q3_df, var)
```

```
##    type   values
## 1 type1 85.82222
## 2 type2 87.73333
```

```r
var.test(values~type, data = q3_df, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  values by type
## F = 0.97822, num df = 9, denom df = 9, p-value = 0.9744
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.2429752 3.9382952
## sample estimates:
## ratio of variances
##          0.9782168
```

Since the p value is greater than 0.05, we fail to reject the null that the two variances are equal. (So we can say the varainces are equal)

b.

```r
t.test(values~type, data = q3_df, var.equal = TRUE)
```

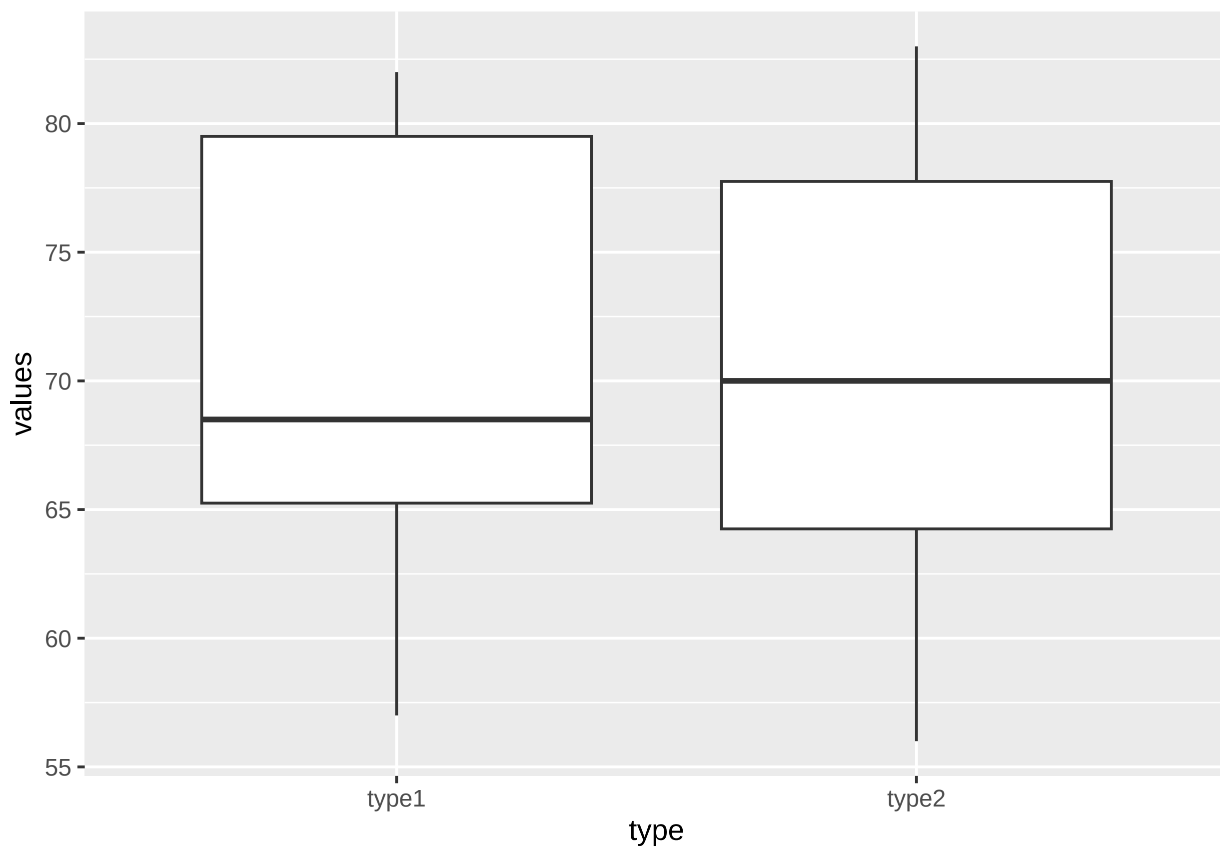
```
##
##  Two Sample t-test
##
## data:  values by type
## t = 0.048008, df = 18, p-value = 0.9622
## alternative hypothesis: true difference in means between group type1 and group type2 is not equal to
## 95 percent confidence interval:
##  -8.552441  8.952441
## sample estimates:
## mean in group type1 mean in group type2
##                70.4                70.2
```

Since our p value is greater than 0.05, we fail to reject the null that different between the means is equal to 0.

c.

```r
library(ggplot2)

ggplot(q3_df, aes(type, values)) + geom_boxplot()
```



Looking at the boxplots, Type 1 and Type 2 seem to violate the normality assumption

Q4

```r
anova_table <- matrix(ncol = 5, nrow = 3)

colnames(anova_table) <- c('DF', 'SS', 'MS', 'F', 'P')
rownames(anova_table) <- c('Factor', 'Error', 'Total')

anova_table[1, 1] <- 3
anova_table[1, 2] <- 36.15
anova_table[3, 1] <- 19
anova_table[3, 2] <- 196.04

anova_table[2, 1] <- anova_table[3, 1] - anova_table[1, 1]
anova_table[2, 2] <- anova_table[3, 2] - anova_table[1, 2]

anova_table[1, 3] <- anova_table[1, 2]/anova_table[1, 1]
anova_table[2, 3] <- anova_table[2, 2]/anova_table[2, 1]

anova_table[1, 4] <- anova_table[1, 3]/anova_table[2, 3]

anova_table[1, 5] <- pf(anova_table[1, 4], 3, 16, lower.tail=FALSE)

anova_table
```

```
##        DF     SS        MS        F         P
## Factor  3  36.15 12.050000 1.205829 0.3395233
## Error  16 159.89  9.993125       NA        NA
## Total  19 196.04        NA       NA        NA
```