

# 705604096\_stats101b\_hw1

Jade Gregory

2023-08-16

## Question 1

```
island_data <- read.csv("101b_Island_HW1.csv", header = TRUE)
island_data
```

##	name	age	gender	pulse.before.in.bpm	pulse.after.in.bpm
## 1	Kelda Larsen	42	female	69	124
## 2	Amber Watanabe	15	female	91	146
## 3	Auna Collins	44	male	87	145
## 4	Andrew Sato	93	male	71	124
## 5	Carina Ghosh	83	female	67	124
## 6	Magritte Sato	40	female	77	133
## 7	Pia Bahadur	21	female	83	137
## 8	Kazuki Summers	90	male	70	127
## 9	Christian Blomgren	23	male	66	120
## 10	Jonas Clausen	67	male	73	129
## 11	Signe Solberg	39	female	65	120
## 12	Natsume Masuda	20	female	73	127
## 13	Harold Solberg	42	male	67	123
## 14	William Blomgren	66	male	65	120
## 15	Edwin Mitchell	52	male	70	126
## 16	William Jones	37	male	67	121
## 17	Illona Lund	27	female	62	119
## 18	Lucien Connolly	19	male	79	132
## 19	Laranya Ramanan	74	female	74	127
## 20	Sophia Varma	17	female	58	112
##	time.in.seconds				
## 1	63.2				
## 2	90.0				
## 3	71.6				
## 4	65.7				
## 5	79.0				
## 6	114.0				
## 7	72.4				
## 8	79.7				
## 9	55.4				
## 10	63.0				
## 11	56.6				
## 12	62.9				
## 13	57.1				

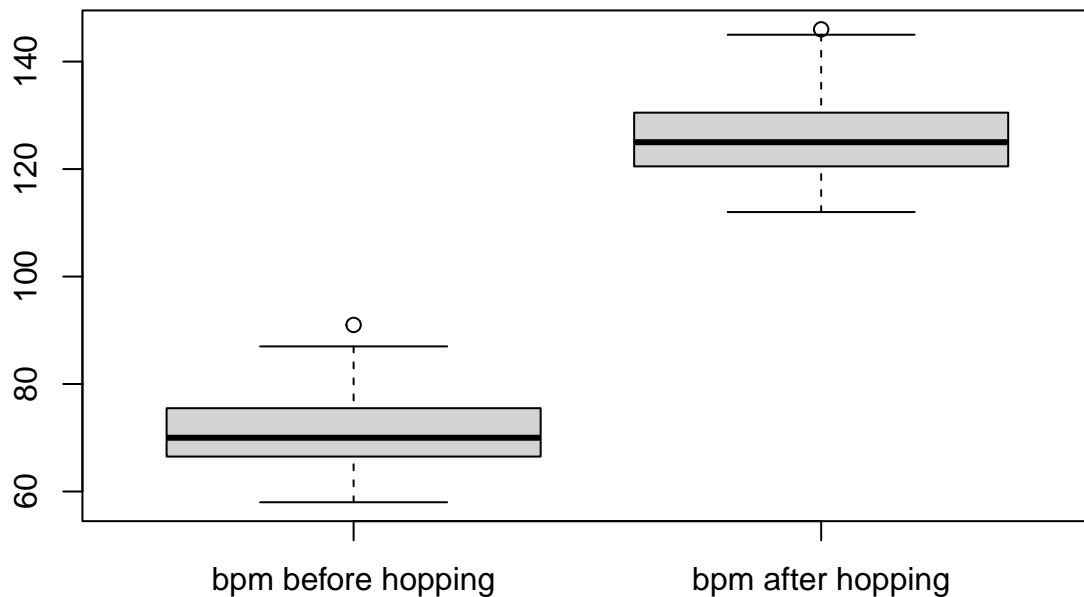
```
## 14          64.9
## 15          88.9
## 16          66.0
## 17          54.6
## 18          55.9
## 19         102.0
## 20          79.8
```

## Part 1

```
t.test(island_data$pulse.after.in.bpm, island_data$pulse.before.in.bpm, paired = TRUE)
```

```
##
## Paired t-test
##
## data: island_data$pulse.after.in.bpm and island_data$pulse.before.in.bpm
## t = 166.13, df = 19, p-value < 2.2e-16
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  54.40582 55.79418
## sample estimates:
## mean difference
##           55.1
```

```
boxplot(island_data$pulse.before.in.bpm, island_data$pulse.after.in.bpm, xaxt = "n")
axis(1, at = 1:2, labels = c("bpm before hopping", "bpm after hopping"))
```



Our null hypothesis is that the true mean difference is equal to zero while our alternative hypothesis states that the true mean difference does not equal zero. From our paired t test, we see that our p-value is less than  $2.2e-16$  which is less than our significance level of 0.05 and we therefore reject our null hypothesis. From this we can conclude that the true mean difference does not equal zero. From our boxplot, we can clearly see how the two means are not equal, and therefore the true mean difference is not zero. This means that the average bpm of an individual before hopping is not equal to the average bpm of an individual after hopping. We can see that the hopping did in fact raise the Islander's pulse. Our sources of variation come from the differences amongst the surveyed individuals, including age and gender. To hold variables constant one can limit the area they choose Islanders from, or limit the age range that they choose from.

## Part 2

```
t.test(island_data$pulse.before.in.bpm ~ island_data$gender)
```

```
##
## Welch Two Sample t-test
##
## data: island_data$pulse.before.in.bpm by island_data$gender
## t = 0.10495, df = 15.906, p-value = 0.9177
## alternative hypothesis: true difference in means between group female and group male is not equal to
## 95 percent confidence interval:
## -7.683671 8.483671
## sample estimates:
## mean in group female mean in group male
```

```
##                71.9                71.5

t.test(island_data$pulse.after.in.bpm ~ island_data$gender)

##
## Welch Two Sample t-test
##
## data: island_data$pulse.after.in.bpm by island_data$gender
## t = 0.051355, df = 16.944, p-value = 0.9596
## alternative hypothesis: true difference in means between group female and group male is not equal to 0
## 95 percent confidence interval:
## -8.018628  8.418628
## sample estimates:
## mean in group female    mean in group male
##                126.9                126.7
```

For our t test of the pulses before hopping, we fail to reject our null hypothesis that states the true difference in means between group female and group male is equal to 0 since our p-value of 0.9177 is greater than our significance level of 0.05. This means that males and females have the same average pulse before hopping. In our second t test of the pulses after hopping, we fail to reject our null hypothesis that states the true difference in means between group female and group male is equal to 0 since our p value of 0.9596 is greater than our significance level of 0.05. We can conclude that males and females have the same average pulse after hopping.

### Part 3

We could have taken their pulse during the hopping exercise to see how quickly it was increasing, and then we could have taken it five minutes after they stopped to see if the pulse returned to their original pulse before the hopping. We could have also recorded the experiment on the same individuals over a longer period of time, say over the course of a few months, to see if we would get consistent results. I would also limit the age range and area for which I chose Islanders in order to get more statistically significant results of a group of Islanders. I would analyze this data by determining the average pulse before hopping and then after hopping, to see if the pulse differed. The experimental design would be similar to the one that occurred in part (1).

### Question 2

```
m_diff <- 11.5
df <- 24
t_stat <- -1.88
p_val <- 0.0723
```

a)

```
SE <- abs(m_diff / t_stat)
SE
```

```
## [1] 6.117021
```

Our standard error is 6.117021.

- b) From a t table, we can determine that it is a two sided test since we can see that a p-value of 0.0723 correlates to a t-statistic of -1.88.
- c) Since our p-value of 0.0723 is greater than our confidence level of 0.05 we fail to reject the null hypothesis and can conclude that there is no difference in the means.
- d)

```
confint_num <- 2.064
ci_lower <- -1 * m_diff - confint_num * SE
ci_upper <- -1 * m_diff + confint_num * SE
confint <- c(ci_lower, ci_upper)
confint
```

```
## [1] -24.125532  1.125532
```

Our 95% confidence interval is (-24.125532, 1.125532).

### Question 3

- a) The null hypothesis to be tested in this experiment is that there is no difference in the two means. The alternative hypothesis states that there is a difference in the two means.

$$H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2$$

- b)

```
machine1 <- c(16.03, 16.04, 16.05, 16.05, 16.02, 16.01, 15.96, 15.98, 16.02, 15.99)
machine2 <- c(16.02, 15.97, 15.96, 16.01, 15.99, 16.03, 16.04, 16.02, 16.01, 16.00)
mean1 <- mean(machine1)
mean2 <- mean(machine2)
sd1 <- 0.015
sd2 <- 0.018
n <- 10

z <- ( mean1 - mean2 ) / ( sqrt( (sd1^2 / n) + (sd2^2 / n) ) )
z
```

```
## [1] 1.349627
```

Since our z value is 1.349627 which is less than 1.96 and greater than -1.96, we fail to reject the null hypothesis and can conclude that there is no difference in the two means.

- c) From a two tailed z table, our p-value is 0.17702 for a rounded z value of 1.35.
- d)

```
confint_d_lower <- ( mean1 - mean2 ) - 1.96 * sqrt( (sd1^2 / n) + (sd2^2 / n) )
confint_d_upper <- ( mean1 - mean2 ) + 1.96 * sqrt( (sd1^2 / n) + (sd2^2 / n) )
confint_d <- c(confint_d_lower, confint_d_upper)
confint_d
```

```
## [1] -0.004522529  0.024522529
```

Our 95% confidence interval is (-0.004522529, 0.024522529).

## Question 4

a)

```
mydata <- data.frame("x10" = c(1530, 1530, 1440),
                     "x15" = c(1610, 1650, 1500),
                     "x20" = c(1560, 1730, 1530),
                     "x25" = c(1500, 1490, 1510)
                     )
df_aov <- aov(values ~ ind, data = stack(mydata))
summary(df_aov)
```

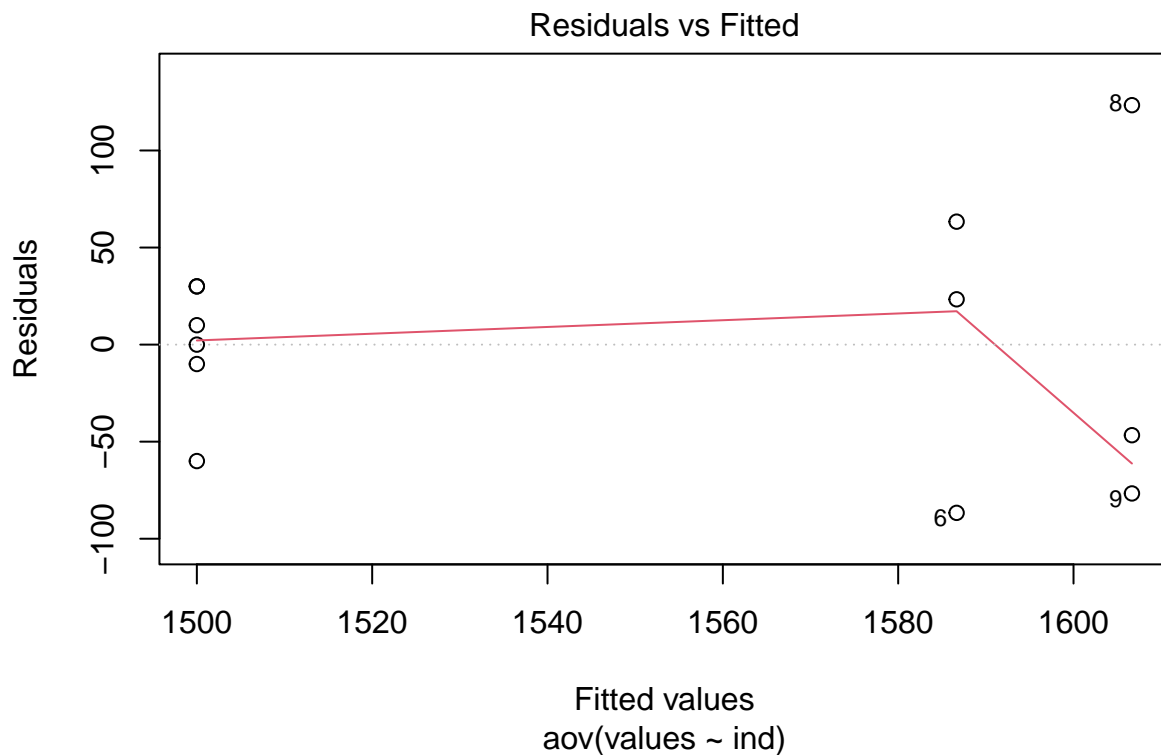
```
##           Df Sum Sq Mean Sq F value Pr(>F)
## ind         3  28633    9544   1.865  0.214
## Residuals   8  40933     5117
```

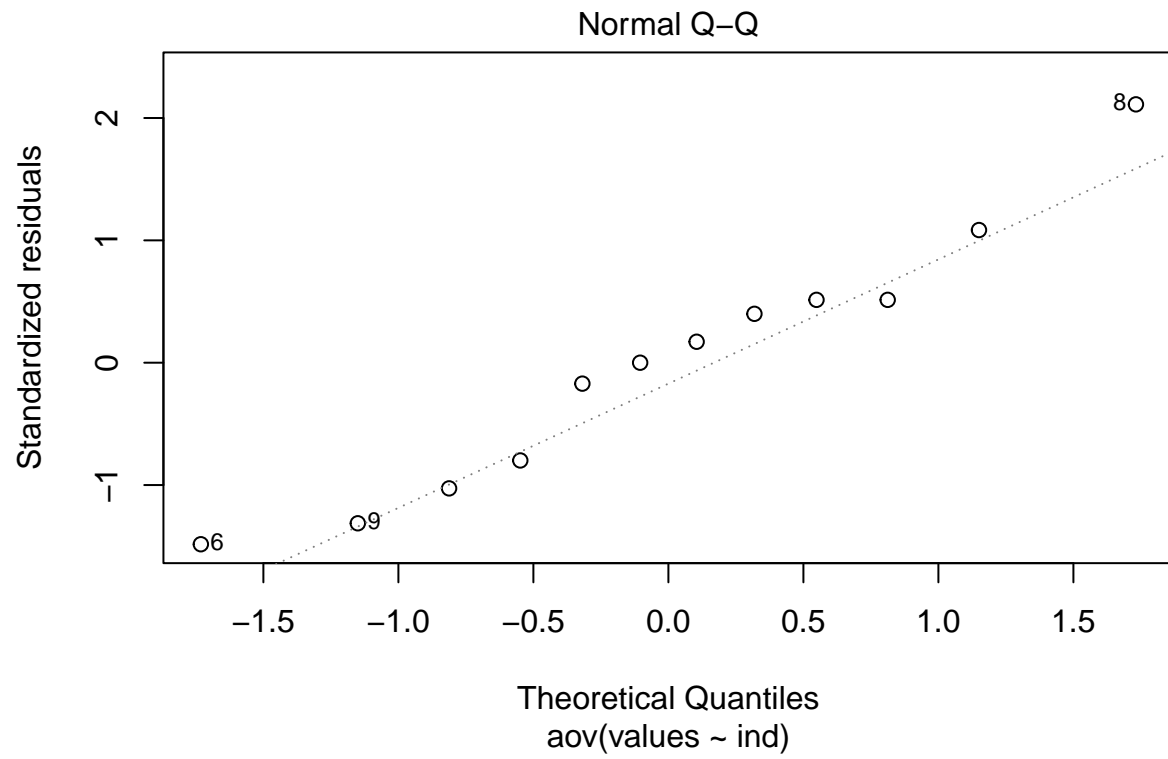
Since our p-value is 0.214 which is greater than our significance level of 0.05, we fail to reject the null hypothesis and can conclude that there is no difference in compressive strength due to rodding level.

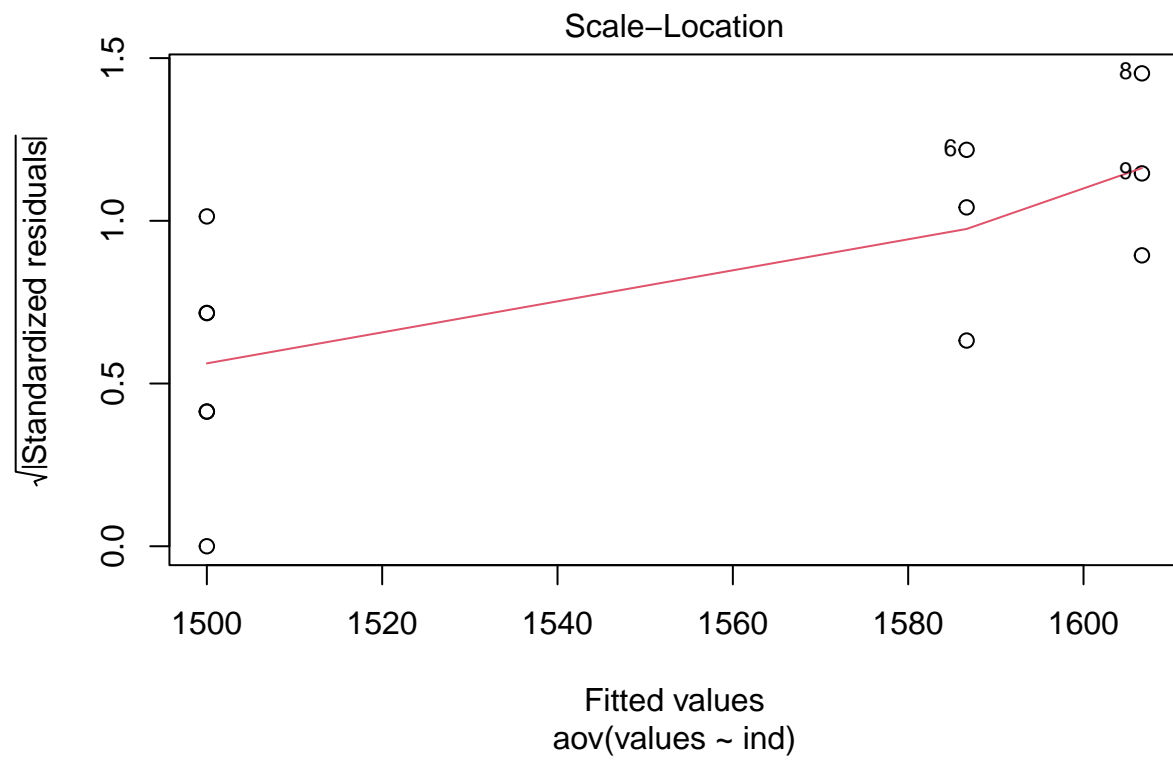
b) From part (a), an F statistic of 1.865 has a p-value of 0.214.

c)

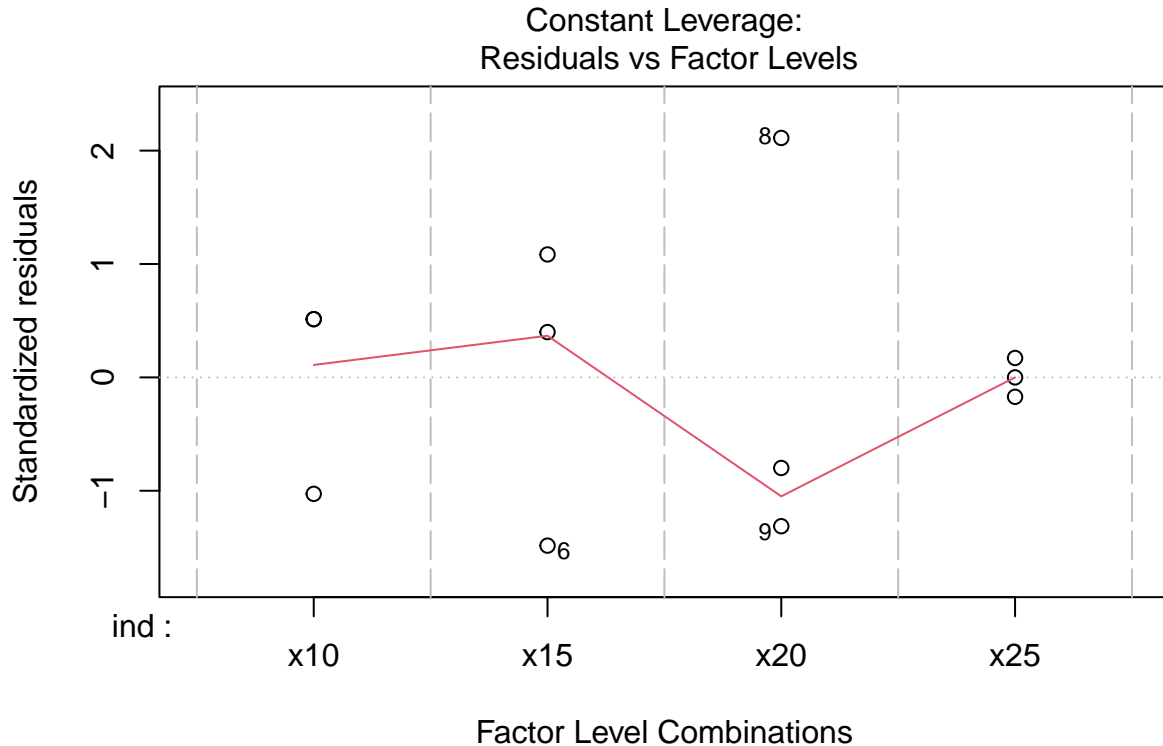
```
plot(df_aov)
```











In our residuals vs fitted plot, we do not see a noticeable trend in our plotted points as they are all roughly equal at each level of the x-values. This suggests that the constant variance assumption is held for our model. In our QQ norm plot, we can see that the plotted points follow the dashed line closely and never stray far from the line, suggesting that our normality assumption is held in our model. In our scale-location plot, we can see that the points are plotted randomly and horizontally across the graph. There is a slight increasing trend in the values, which could suggest that the model does not actually hold the constant variance assumption. But, the lack of data points on the graph makes it harder to come to a precise conclusion about this. In our residuals vs factor levels plot, we can see that the points are plotted randomly and horizontally across the graph which suggests that the model appropriately accounts for the contribution of each factor level in our model. Overall, I would conclude that this model is satisfactory for our data.

d)

```
curve(dt(x, df = 3), from = -4, to = 4, type = "l", ylab = " ", yaxt = "n", xaxt = "n", xlab= "Mean Comp
axis(1, at = -3:3, labels = c(1418, 1459, 1500, 1541, 1582, 1623, 1664))
points(x = c(-1, 1.1, 1.6), y = c(0,0, 0))
points(x = c(-1, 1.1, 1.6), y = c(0.025, 0.025 ,0.025), col = "white")
text(c(-1, 1.1, 1.6), c(0.025, 0.025 ,0.025), c("(10, 25)", "(15)", "(25)"))
```

### Scaled t Distribution

