

Home Work 1-MATH 588

Md Ismail Hossain

1/27/2022

1(a)

Null Hypothesis: Mean weights are same in group 0 and 1.

Alternative Hypothesis: Mean weights are not same in group 0 and 1.

```
library(readr)
fullBumpus <- read_table2("E:/NMT MS/Spring 22/MATH 588/Home_Work/Spring-2022---MATH-588-01-Advanced-Da

ttst1 = t.test(Weight~Survive,var.equal=TRUE,data = fullBumpus)
ttst1

##
## Two Sample t-test
##
## data: Weight by Survive
## t = 2.6093, df = 134, p-value = 0.0101
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## 0.1569291 1.1399459
## sample estimates:
## mean in group 0 mean in group 1
## 25.86094 25.21250
```

Comments: According to the p-value [considering variance are same in both group] (0.0101) it can be said that there is statistically significant difference between two group (0 and 1) in terms of weights, at level of significance is 0.05.

```
ttst2 = t.test(Weight~Survive,var.equal=FALSE,data = fullBumpus)
ttst2

##
## Welch Two Sample t-test
##
## data: Weight by Survive
## t = 2.5703, df = 117.95, p-value = 0.01141
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## 0.1488463 1.1480287
## sample estimates:
## mean in group 0 mean in group 1
## 25.86094 25.21250
```

Comments: According to the p-value [considering variance are not equal in both group] (0.01141) it can be said that there is statistically significant difference between two group (0 and 1) in terms of weights, at level of significance is 0.05.

So, in both case (variance same or different), the null hypothesis is rejected.

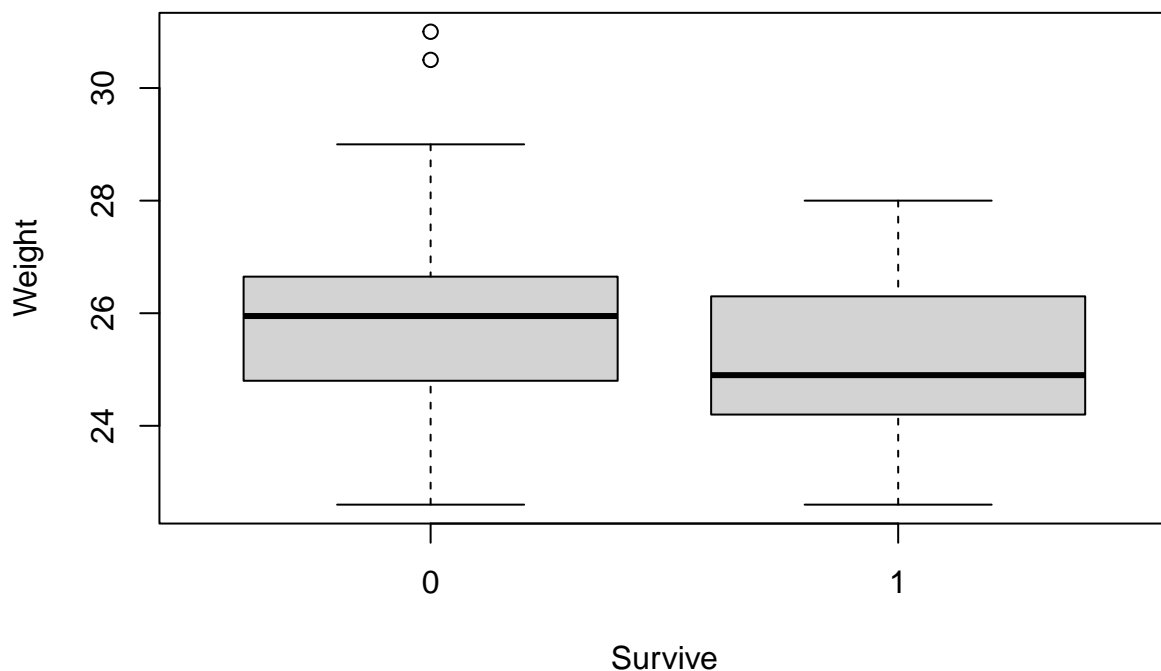
1(b)

After observing the normal qqplot of the residual we can interpret that the data does not follow normal distribution. Because the lower tail and upper part in the plot not goes close to the straight line. We should take some sort of transformation before carry out the t-test. The data distribution seems right skewed, so log-transformation could leads the data to normal shape and we know normality is the key assumption for t-test.

```
res = resid(lm(Weight~Survive, data = fullBumpus))
qqnorm(res)
qqline(res)
```

1(c)

```
boxplot(Weight~Survive,data=fullBumpus)
```



```
# Showing IQR
aggregate(Weight~Survive,fullBumpus,IQR)
```

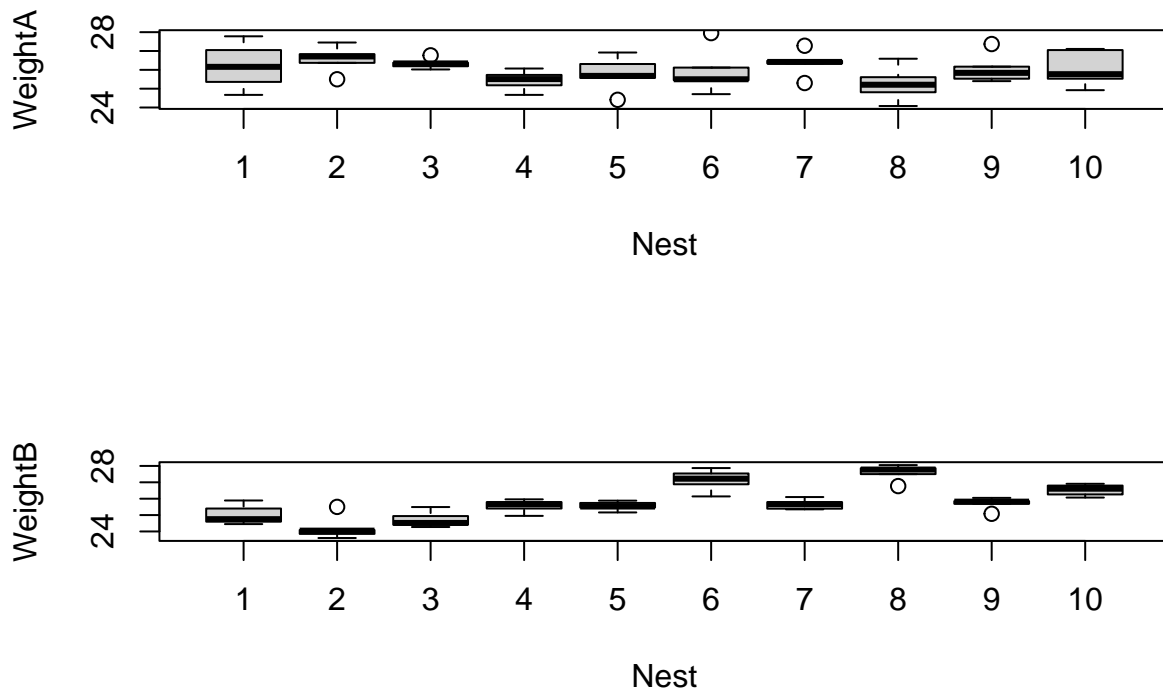
```
##   Survive Weight
## 1      0  1.775
## 2      1  2.100
```

As we know if the IQRs is between 0.5 and 2.0, there is no cause for concern about unequal variances. But in our case we observed that the IQR is greater than 2 for the surviving sparrow (1). So, the variability seems different between this two groups for the weight.

1(d)

```
HW1FakeCor <- read_table2("E:/NMT MS/Spring 22/MATH 588/Home_Work/Spring-2022---MATH-588-01-Advanced-Da

{par(mfrow=c(2,1))
boxplot(WeightA~Nest,data=HW1FakeCor)
boxplot(WeightB~Nest,data=HW1FakeCor)}
```



From the boxplot we have seen that Nest 4 and 9 have close mean values and similar types of weight distribution. So, birds in nest 4 and 9 are similar in weights and have correlated errors.

2(a)

```
fullBumpus <- read_table2("E:/NMT MS/Spring 22/MATH 588/Home_Work/Spring-2022---MATH-588-01-Advanced-Da

mdl <- lm(Alar~Female+Weight,data = fullBumpus)
summary(mdl)

##
## Call:
```

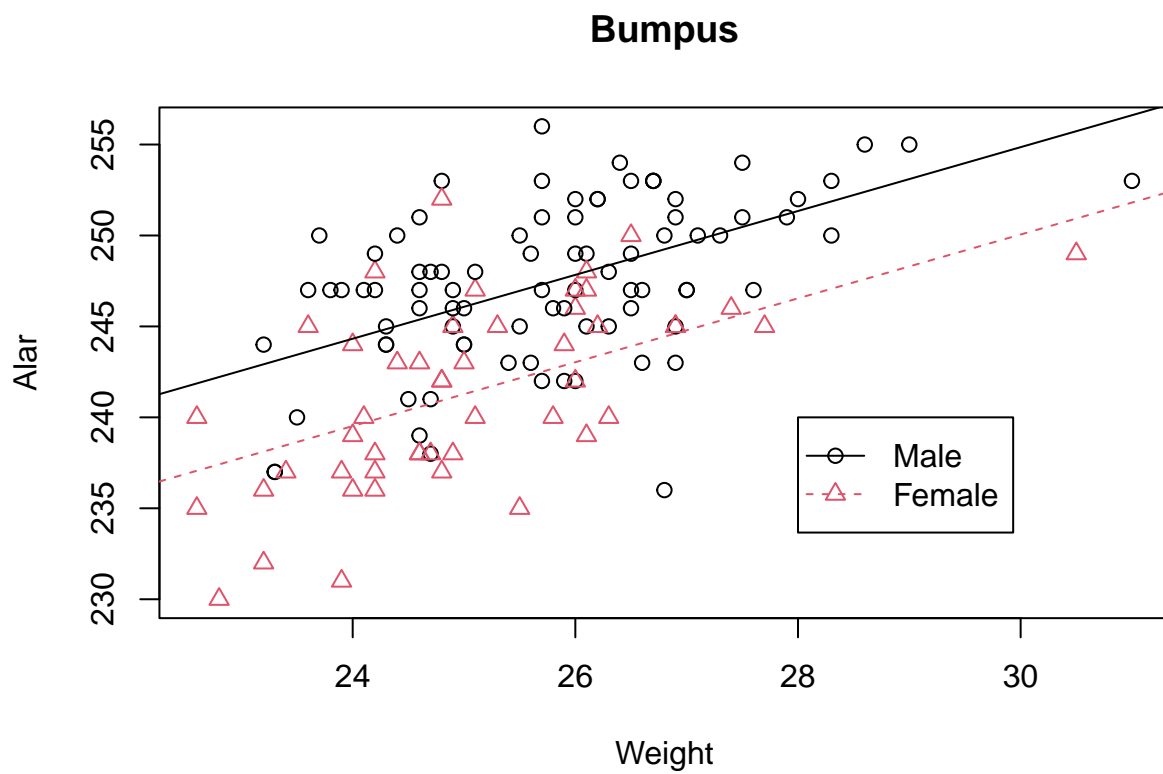
```
## lm(formula = Alar ~ Female + Weight, data = fullBumpus)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.2387  -2.6125   0.2613   2.8729  11.0747
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  202.1958     6.1318  32.975  < 2e-16 ***
## Female       -4.8027     0.7271  -6.605 8.71e-10 ***
## Weight        1.7553     0.2372   7.401 1.37e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.942 on 133 degrees of freedom
## Multiple R-squared:  0.4961, Adjusted R-squared:  0.4885
## F-statistic: 65.47 on 2 and 133 DF,  p-value: < 2.2e-16
```

2(b)

```
b0M = coef(mdl)[1]
b0F = coef(mdl)[1] + coef(mdl)[2]
b1 = coef(mdl)[3]
```

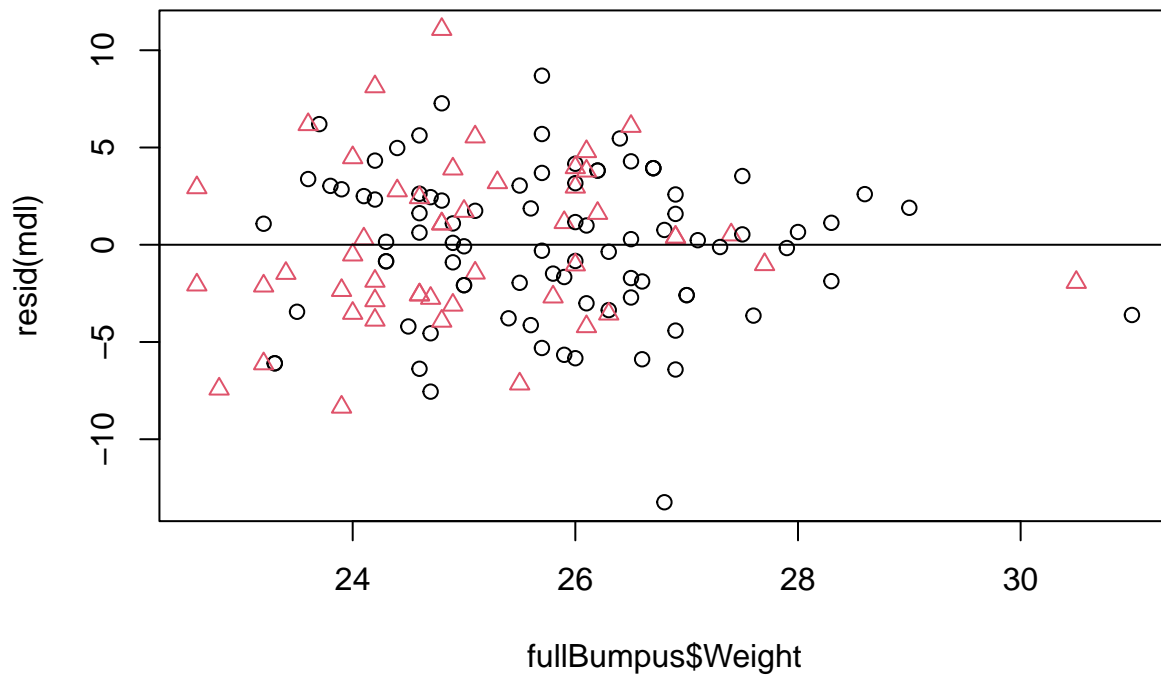
2(c)

```
fullBumpus$Female <- as.factor(fullBumpus$Female)
#with(fullBumpus, table(Female, as.numeric(Female)))
{plot(Alar~Weight, pch=as.numeric(Female),col=as.numeric(Female), main="Bumpus",data=fullBumpus)
abline(b0M, b1, col=1, lty=1)
abline(b0F, b1, col=2, lty=2)
legend(28, 240, c("Male", "Female"), col=1:2, lty=1:2, pch=1:2)}
```



Without the interaction between Female and weights, the fitted lines are parallel. Let's explore residual vs X plot

```
{plot(resid mdl)~fullBumpus$Weight, col=as.numeric(fullBumpus$Female),
pch=as.numeric(fullBumpus$Female))
abline(h=0)}
```



2(d)

```
mdlI = lm(Alar~Female*Weight, data=fullBumpus)
summary(mdlI)
```

```
##
## Call:
## lm(formula = Alar ~ Female * Weight, data = fullBumpus)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
##	-13.0332	-2.5531	0.0527	2.8415	11.1550

```
##
## Coefficients:
```

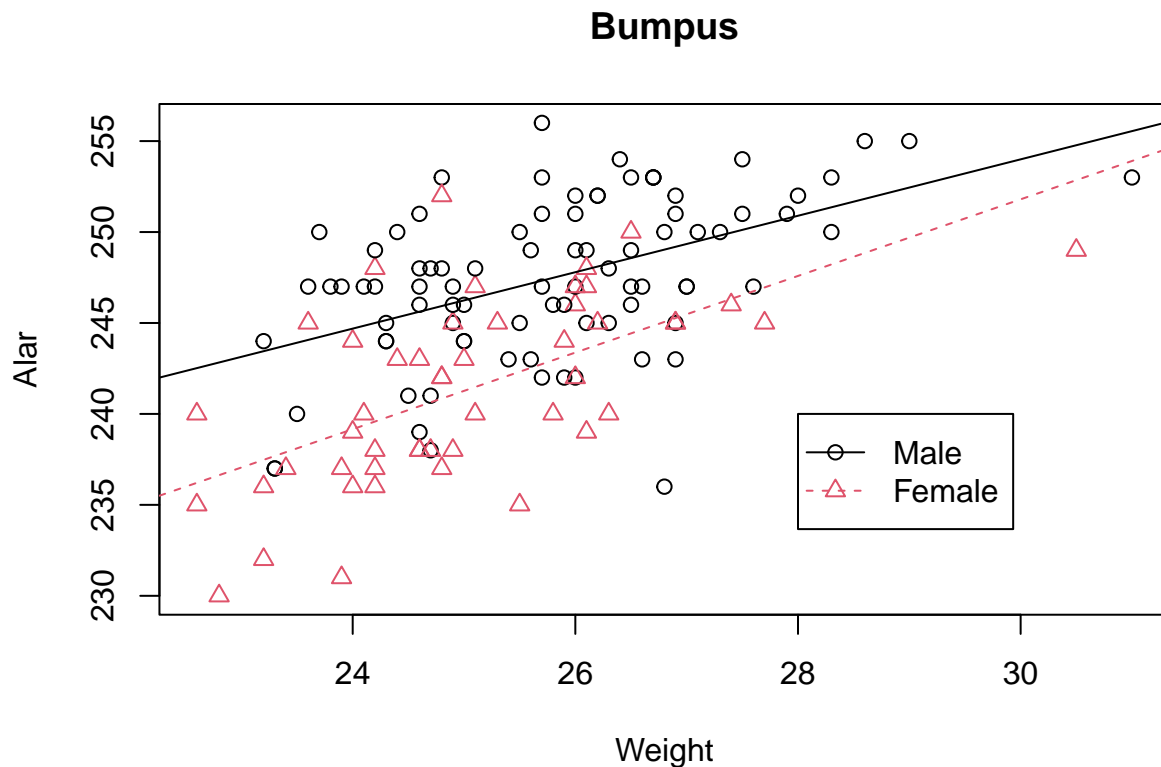
	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	207.4608	7.6956	26.958	< 2e-16 ***
## Female1	-18.8586	12.4580	-1.514	0.132
## Weight	1.5512	0.2979	5.207	7.18e-07 ***
## Female1:Weight	0.5554	0.4914	1.130	0.260

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.938 on 132 degrees of freedom
## Multiple R-squared:  0.5009, Adjusted R-squared:  0.4896
## F-statistic: 44.16 on 3 and 132 DF, p-value: < 2.2e-16
```

```

b0M = coef(md1I)[1]
b0F = coef(md1I)[1] + coef(md1I)[2]
b1M = coef(md1I)[3]
b1F = coef(md1I)[3] + coef(md1I)[4]
{with(fullBumpus, plot(Alar~Weight, pch=as.numeric(Female),
col=as.numeric(Female), main="Bumpus"))
abline(b0M, b1M, col=1, lty=1)
abline(b0F, b1F, col=2, lty=2)
legend(28, 240, c("Male", "Female"), col=1:2, lty=1:2, pch=1:2)}

```



2(e)

```
anova mdl,mdlI)
```

```

## Analysis of Variance Table
##
## Model 1: Alar ~ Female + Weight
## Model 2: Alar ~ Female * Weight
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1     133 2067.1
## 2     132 2047.3  1     19.81 1.2773 0.2605

```

The p-value for the F-statistic is greater than 0.05. So, at 5% level of significance we conclude that the non-parallel slopes are not significant or the interaction term is insignificant. So, we don't have good evidence of non-parallel slopes.

2(f)

```
confint mdlI)
```

```
##                2.5 %      97.5 %  
## (Intercept)  192.2381377 222.683434  
## Female1     -43.5018377   5.784645  
## Weight       0.9619145   2.140502  
## Female1:Weight -0.4166585   1.527374
```

The 95% CI for the difference of slopes (female-male)=[-0.4166585 ,1.527374] because interaction term representing difference between male and female.

3(a)

```
#install.packages("Sleuth3")  
library(Sleuth3)  
head(ex0327)
```

```
##   PollenRemoved DurationOfVisit BeeType  
## 1           0.07              2  Queen  
## 2           0.10              5  Queen  
## 3           0.11              7  Queen  
## 4           0.12             11  Queen  
## 5           0.15             12  Queen  
## 6           0.19             11  Queen
```