Assignment 1

Ayato Tanemura (atan524)

Question 1

(a) Essential assumption check

(i) Normality Distribution

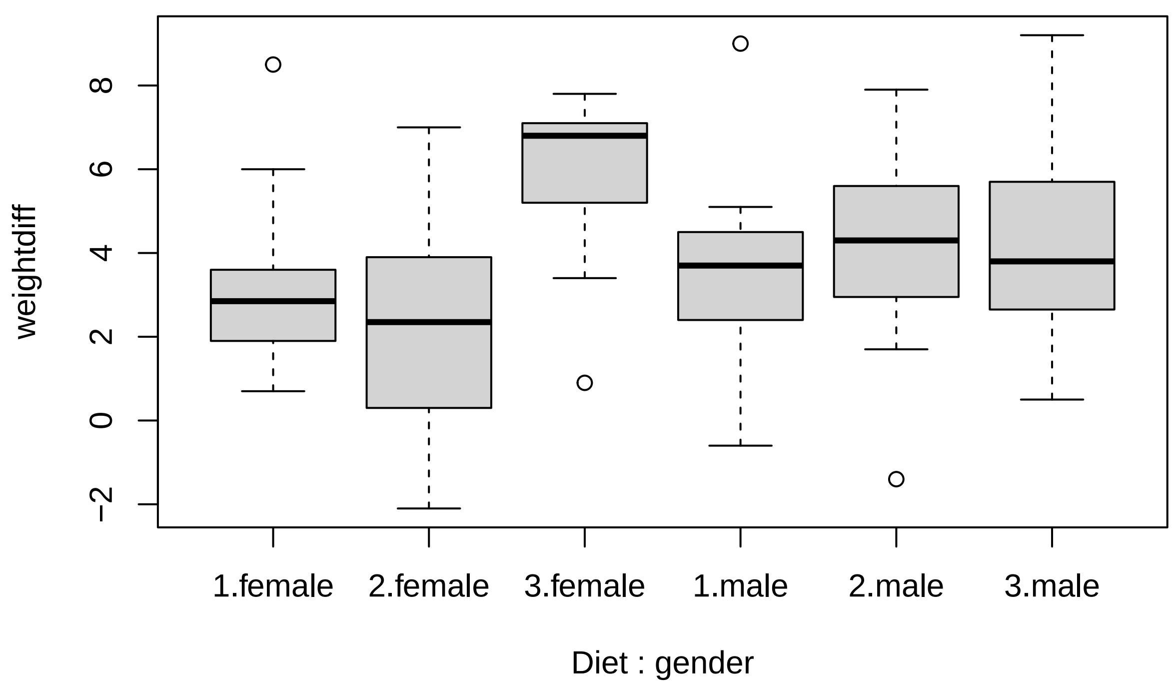
SW test:

```
>
>   Shapiro-Wilk normality test
>
> data:  weightdiff
> W = 0.98991, p-value = 0.802
```

We want to see non-significant result.

P-value for weight difference test of SW is more than 0.05. H0 is not rejected and we conclude that the assumption f normality is satisfied.

Normality plot:



- weightdiff shows the difference between before and after six weeks.
- The formula is `weightdiff = (weight before diet) - (weight 6 weeks after)`

Outliers check: 4 outliers are identified via this plot.

There are 1 (Diet1, Male), 1 (Diet1, Female), 1 (Diet2, Male), and 1 (Diet3, Female) observations should be omitted.

Therefore, 4 observations will be deleted as outliers.

(ii) Homogeneity of Variance

Levenne's test

```
> Levene's Test for Homogeneity of Variance (center = median)
>       Df F value Pr(>F)
> group  5  1.5479 0.1867
>       68
```

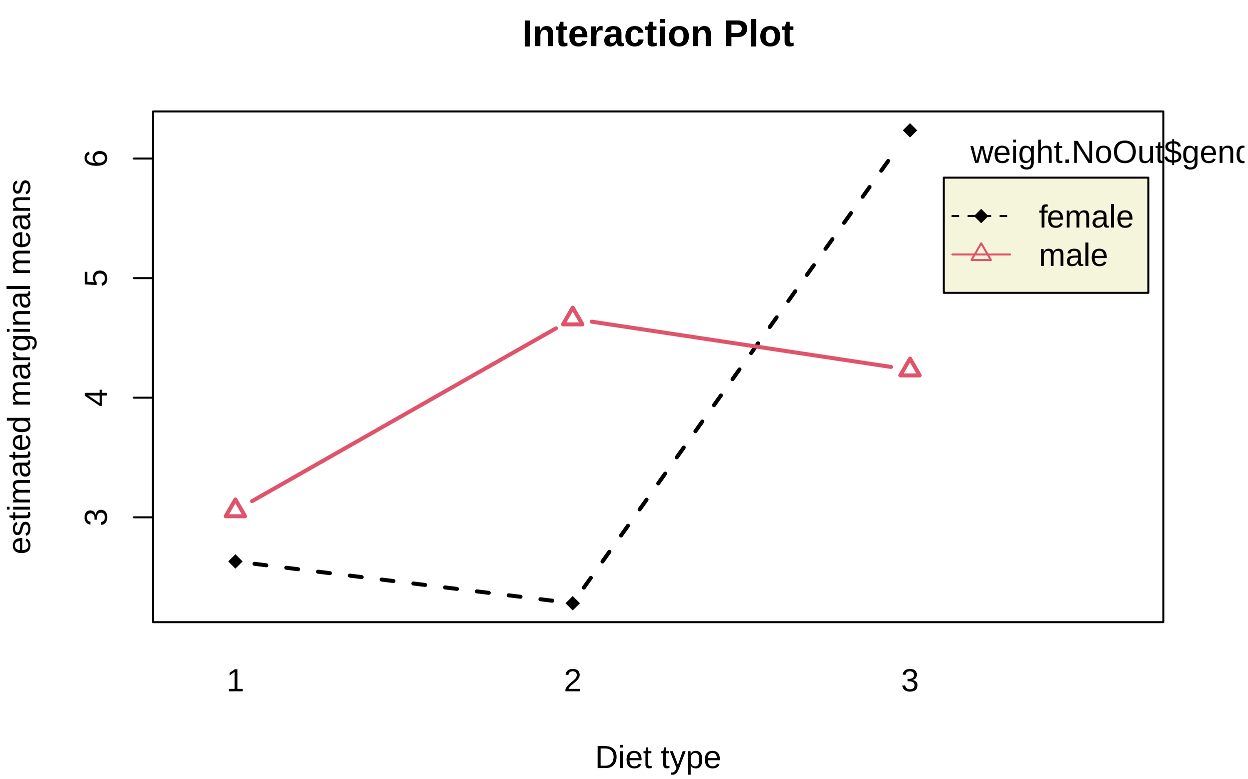
The p-value is more than 0.05 (level of significance), which means the result is non-significant.

Therefore, Homogeneity is met.

(iii) Independence

????

(b) Interaction Plot



(c) Difference between Male Diet 1 and 2

ANOVA summary:??

```
>             Df Sum Sq Mean Sq F value   Pr(>F)
> gender       1   2.27    2.27   0.582 0.448008
> Diet         2  89.25   44.63  11.474 5.09e-05 ***
> gender:Diet  2  60.81   30.41   7.818 0.000879 ***
> Residuals   68 264.48    3.89
> ---
> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tukey test:

```
>   Tukey multiple comparisons of means
>     95% family-wise confidence level
>
> Fit: aov(formula = weightdiff.NoOut ~ gender * Diet, data = weight.NoOut)
>
> $gender
>                  diff        lwr      upr     p adj
> male-female 0.3546137 -0.5726118 1.281839 0.4480081
>
> $Diet
>          diff        lwr      upr     p adj
> 2-1 0.4002877 -0.9685892 1.769165 0.7639065
> 3-1 2.4883944  1.1195175 3.857271 0.0001344
> 3-2 2.0881066  0.7775068 3.398707 0.0008487
>
> $`gender:Diet`
>                         diff         lwr       upr     p adj
> male:1-female:1    0.4247863 -2.08306614 2.9326388 0.9961356
> female:2-female:1 -0.3495192 -2.50900778 1.8099693 0.9968828
> male:2-female:1    2.0292308 -0.40339788 4.4618594 0.1552935
> female:3-female:1  3.6049451  1.37738325 5.8325069 0.0001568
> male:3-female:1    1.6025641 -0.71265064 3.9177788 0.3366516
> female:2-male:1   -0.7743056 -3.18405746 1.6354464 0.9339493
> male:2-male:1      1.6044444 -1.05284658 4.2617355 0.4912294
> female:3-male:1    3.1801587  0.70921862 5.6510988 0.0044074
> male:3-male:1      1.1777778 -1.37246393 3.7280195 0.7536003
> male:2-female:2    2.3787500  0.04738508 4.7101149 0.0428650
> female:3-female:2  3.9544643  1.83795493 6.0709736 0.0000098
> male:3-female:2    1.9520833 -0.25649077 4.1606574 0.1132848
> female:3-male:2    1.5757143 -0.81884270 3.9702713 0.3934275
> male:3-male:2     -0.4266667 -2.90297256 2.0496392 0.9958122
> male:3-female:3   -2.0023810 -4.27756015 0.2727982 0.1161146
```

(d) Highest Efficiency

As you can see form the interaction plot, Female Diet 3 has the largest impact on weight difference. It is regarding both gender and diet type.

As can bee seen the difference between before and after six weeks, the larger difference of weight for Male Diet 1 and 2 while it has larger impact on weight difference for Female Diet 3.

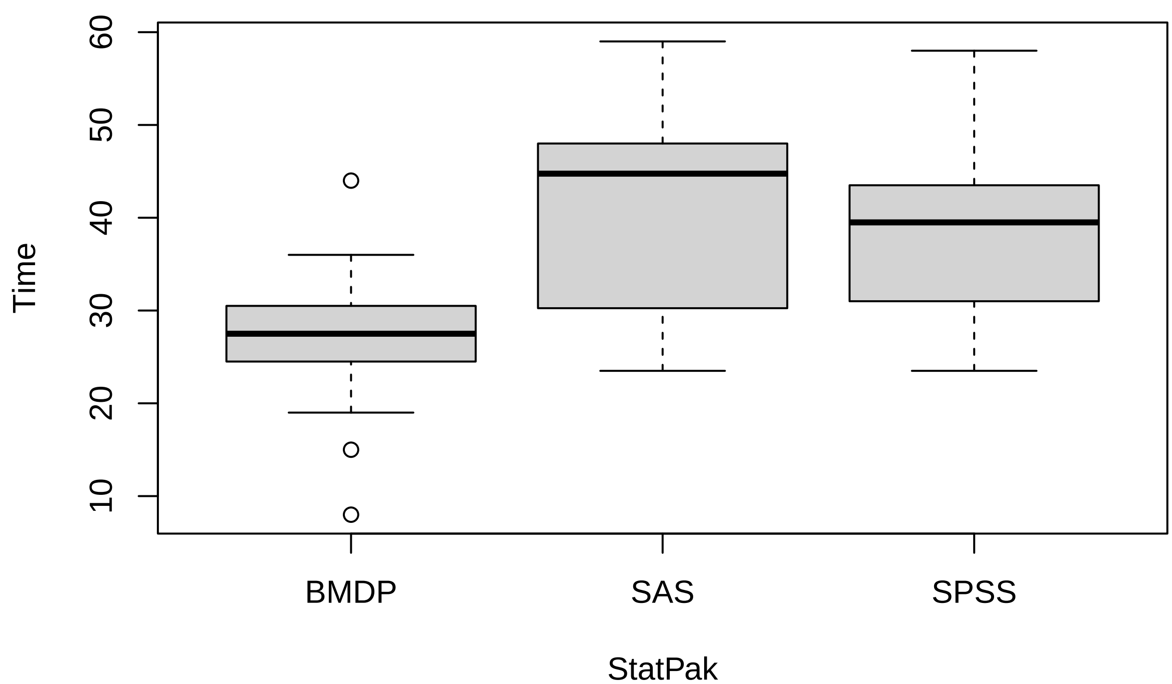
Question 2

```
df.stat <- read.csv("Data/STATPAK.csv") %>%
  mutate(StatPak = factor(StatPak))
```

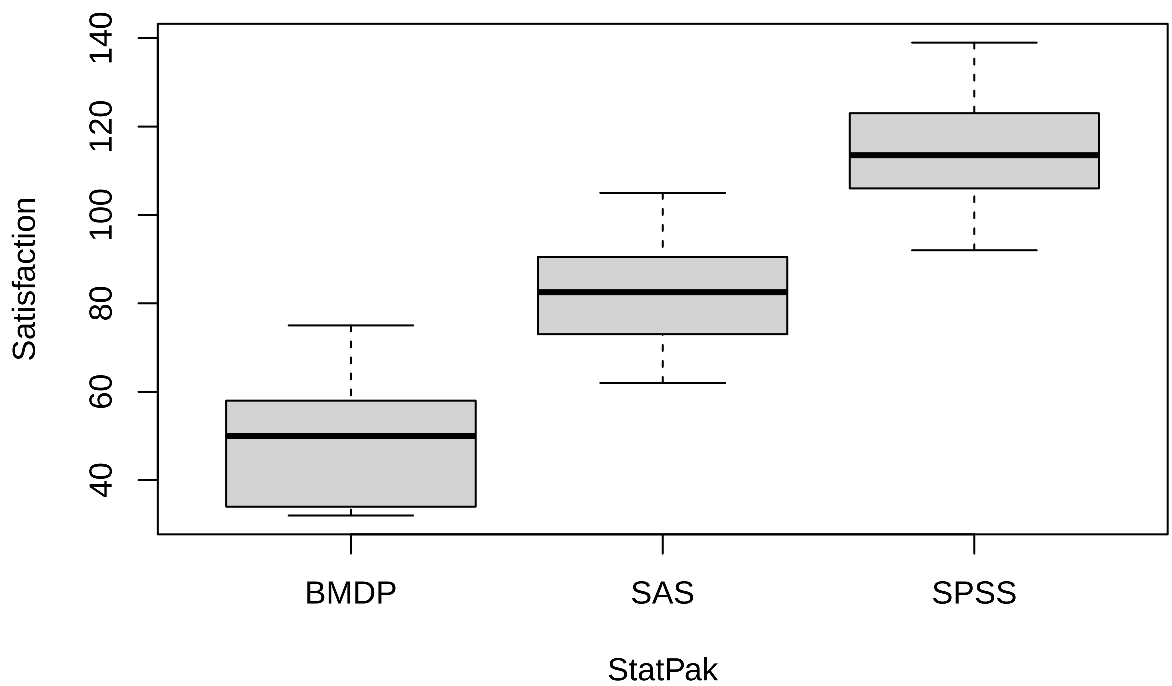
(a)Outliers & Homogeneity

Outliers

```
# Outliers Detestion
olTime <- boxplot(Time ~ StatPak, data = df.stat)$out
```




```
olSatisfaction <- boxplot(Satisfaction ~ StatPak, data = df.stat)$out
```




```r
olTime
```

```
> [1] 15 44  8
```

```r
olSatisfaction
```

```
> numeric(0)
```

```r
out <- df.stat[df.stat$StatPak == "BMDP" & (df.stat$Time == 15 | df.stat$Time == 44 | df.stat$Time == 8
out
```

```
>    No StatPak Platform Experience Comp Time Satisfaction
> 35 35    BMDP  Windows         12   29   15           58
> 36 36    BMDP      Mac          0   88   44           44
> 65 65    BMDP  Windows         10   38    8           58
```

```r
# Remove Outliers
df.NoOutlier <- df.stat[-which(df.stat$No %in% out$No),] %>%
  mutate(Satisfaction = as.numeric(Satisfaction))
```

Homogeneity of Variance and covariance

Levene's test for Time

```
> Levene's Test for Homogeneity of Variance (center = median)
>       Df F value   Pr(>F)
> group  2  5.8648 0.004279 **
>       76
> ---
> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

P-value of Levene's for Time is 0.004 < 0.05 -> assumption is not satisfied.

Levene's test for Satisfaction

```
> Levene's Test for Homogeneity of Variance (center = median)
>       Df F value Pr(>F)
> group  2  0.4613 0.6322
>       76
```

P-value of Levene;s for Satisfaction is 0.6322 < 0.05 -> assumption is satisfied.

Box's M test

```
>
>   Box's M-test for Homogeneity of Covariance Matrices
>
> data:  df.NoOutlier[, 6:7]
> Chi-Sq (approx.) = 17.523, df = 6, p-value = 0.007541
```

P-value of Box test is 0.007 < 0.05 -> Homogeneity of cov matrices assumption not met. We should use Pillai's Trace for interpretation of MANOVA results.

(b) Effects of stat packages

(c) Independent Group