EDDA - Assignment 1

Exercise 4

a)

Disregarding the type of drink, test whether the run times before drink and after are correlated.

```
data <- read.table(file="data/run.txt",header=TRUE)
cor(data$before, data$after)
```

```
## [1] 0.638803
```

Run times before and after the drink seem to be positively correlated.

b)

```
# calculate differences
data <- data %>%
  mutate(diff = before - after)

# filter for lemo

lemo <- data %>%
  filter(drink == "lemo")

t.test(lemo$before, lemo$after)
```

```
##
##  Welch Two Sample t-test
##
## data:  lemo$before and lemo$after
## t = -0.5036, df = 18.989, p-value = 0.6203
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.7476599  0.4576599
## sample estimates:
## mean of x mean of y
##  7.554167  7.699167
```

```
# filter for energy

energy <- data %>%
```

```r
  filter(drink == "energy")

t.test(energy$before, energy$after)
```

```
##
##  Welch Two Sample t-test
##
## data:  energy$before and energy$after
## t = 0.97741, df = 18.854, p-value = 0.3407
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1761393  0.4844726
## sample estimates:
## mean of x mean of y
##  7.732500  7.578333
```

For both energy and soft-drink groups there does not seem to be a significant difference in running times.

c)

For each pupil compute the time difference between the two running tasks. Test whether these time differences are effected by the type of drink.

```r
# perform t-test

t.test(lemo$diff, energy$diff)
```

```
##
##  Welch Two Sample t-test
##
## data:  lemo$diff and energy$diff
## t = -1.4764, df = 16.509, p-value = 0.1586
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.7276409  0.1293076
## sample estimates:
##  mean of x  mean of y
## -0.1450000  0.1541667
```

The p-value is > 0.05 therefore the means of the two populations are not significantly different.

d)

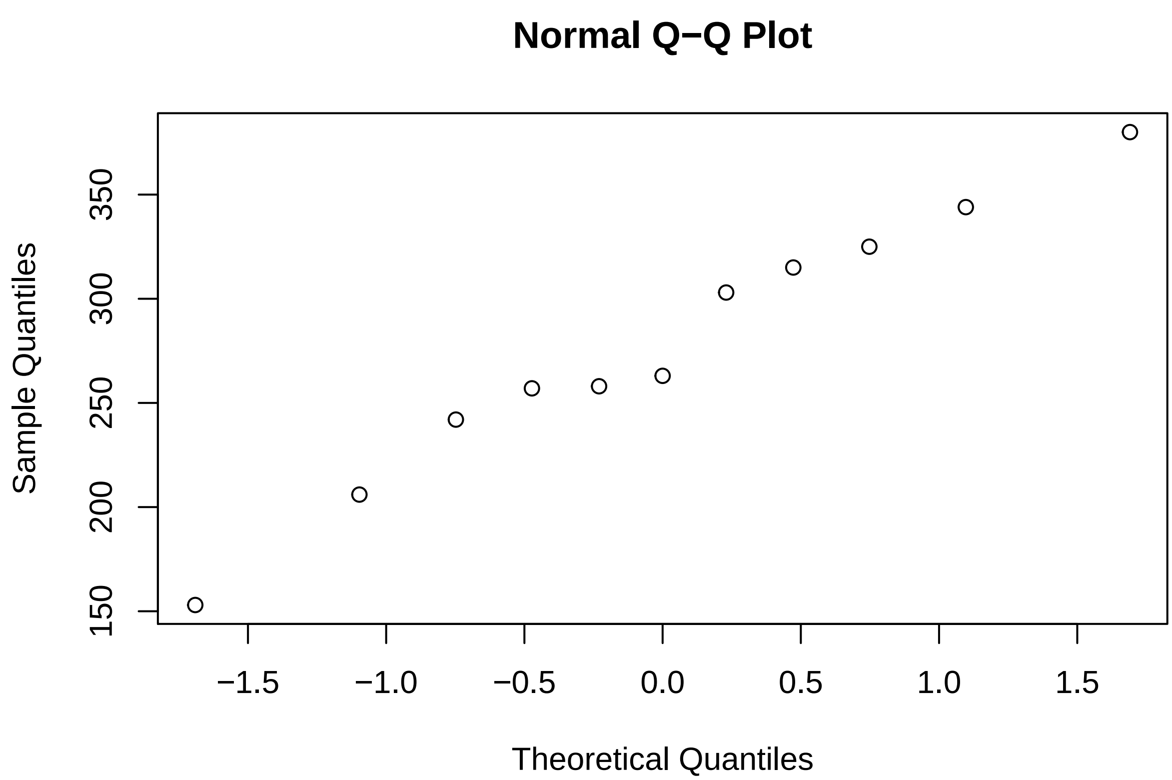
Can you think of a plausible objection to the design of the experiment in b) if the main aim was to test whether drinking the energy drink speeds up the running? Is there a similar objection to the design of the experiment in c)? Comment on all your fndings in this exercise.

Exercise 5

a)

Test whether the distributions of the chicken weights for meatmeal and sunflower groups are different by performing three tests: the two samples t-test (argue whether the data are paired or not), the Mann-Whitney test and the Kolmogorov-Smirnov test. Comment on your fndings.

```r
# filter for meatmeal

meatmeal <- chickwts %>%
  filter(feed == "meatmeal") %>%
  select(weight)

# filter for sunflower

sunflower <- chickwts %>%
  filter(feed == "sunflower") %>%
  select(weight)

# check for data normality

qqnorm(meatmeal$weight)
qqnorm(meatmeal$weight)
```



```r
# perform t-test, the data is not paired

t.test(meatmeal, sunflower)
```

```
##
##  Welch Two Sample t-test
##
## data:  meatmeal and sunflower
## t = -2.1564, df = 18.535, p-value = 0.04441
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -102.572435   -1.442716
## sample estimates:
## mean of x mean of y
##  276.9091  328.9167
```

```r
# Mann-Whitney test

wilcox.test(meatmeal$weight, sunflower$weight)
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  meatmeal$weight and sunflower$weight
## W = 36, p-value = 0.06882
## alternative hypothesis: true location shift is not equal to 0
```

```r
# Kolmogorov-Smirnov test

ks.test(meatmeal$weight, sunflower$weight)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  meatmeal$weight and sunflower$weight
## D = 0.47727, p-value = 0.1085
## alternative hypothesis: two-sided
```

Data in chickwts is not paired as the "treatment" of different feed was applied to different newly-hatched chicks not the same chick. From t-test we can see that the p-values <0.05, therefore the means between the two groups are significantly different. From Mann-Whitney test we can see that p-value is >0.05 therefore we can not conclude that the medians of the two datasets are different. From Kolgomorov-Smirnov test we can not conclude that the means are different.

b)

Conduct a one-way ANOVA to determine whether the type of feed supplement has an effect on the weight of the chicks. Give the estimated chick weights for each of the six feed supplements. What is the best feed supplement?

```r
chickaov <- lm(weight~feed, data = chickwts)
# performing one-way ANOVA
anova(chickaov)
```

```
## Analysis of Variance Table
##
## Response: weight
##           Df Sum Sq Mean Sq F value    Pr(>F)
## feed       5 231129   46226  15.365 5.936e-10 ***
## Residuals 65 195556    3009
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#extracting more information
summary_table <- summary(chickaov)
summary_table$coefficients
```

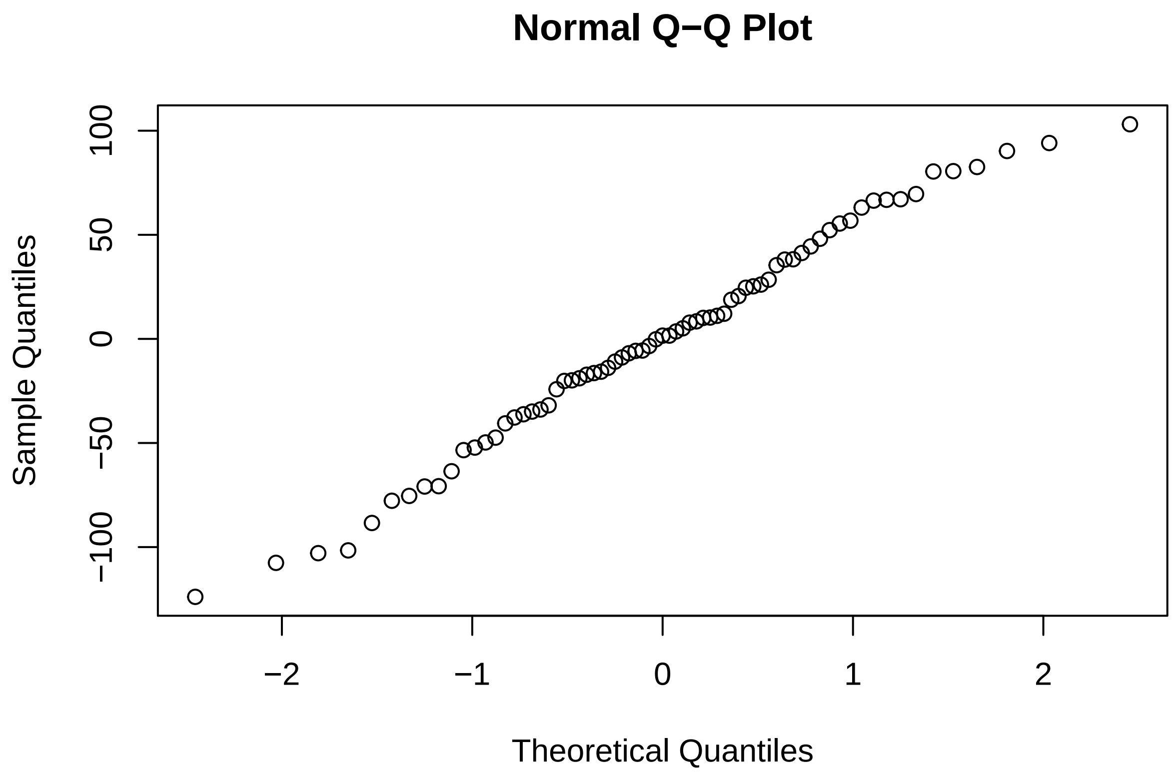
```
##                  Estimate Std. Error    t value     Pr(>|t|)
## (Intercept)    323.583333   15.83391 20.4360920 5.325090e-30
## feedhorsebean -163.383333   23.48549 -6.9567776 2.067997e-09
## feedlinseed   -104.833333   22.39254 -4.6816194 1.493344e-05
## feedmeatmeal   -46.674242   22.89580 -2.0385502 4.556672e-02
## feedsoybean    -77.154762   21.57799 -3.5756235 6.654079e-04
## feedsunflower    5.333333   22.39254  0.2381746 8.124949e-01
```

From the results of one-way ANOVA we can see that the p-values is <0.05 therefore we can conclude that the means between all of the feed varieties are significantly different. From summary statistics it seems that "sunflower" feed is the feed resulting in highest weight. therefore it is the best.

c)

Check the ANOVA model assumptions by using relevant diagnostic tools.

```r
# check for normality
qqnorm(chickaov$residuals)
```




```
# check if the variances are equal
chickwts %>%
  group_by(feed) %>%
  summarise(variance = var(weight))
```

```
## # A tibble: 6 x 2
##   feed      variance
## * <fct>        <dbl>
## 1 casein       4152.
## 2 horsebean    1492.
## 3 linseed      2729.
## 4 meatmeal     4212.
## 5 soybean      2930.
## 6 sunflower    2385.
```

From qqplot assumption of normality holds. However the assumption of equal variances does not hold.

d)

Does the Kruskal-Wallis test arrive at the same conclusion about the effect of feed supplement as the test in b)? Explain possible differences between conclusions of the Kruskal-Wallis and ANOVA tests.

```
kruskal.test(weight~feed, data = chickwts)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  weight by feed
## Kruskal-Wallis chi-squared = 37.343, df = 5, p-value = 5.113e-07
```

With Kruskal-Wallis test we arrive to the same conclusion as with ANOVA.