Data Analysis using R

T-test/Wilcoxon test demonstration

When to use T-test?
-When comparing means of 2 groups of continuous data that meet the following assumptions:

Assumptions for T-test:
1. Data independent.
2. Normal distribution.
3. Equal variance.

When to use a Wilcoxon test?
-When the assumptions for T-test are not met; (ie. if data not normal)

Exercise 1: Using Excercise1Assign6.csv file.

In R:

1. Import the Data:

```
> #Assignment6
> #1.Import the Data:
> getwd()
[1] "C:/Users/Edison/Documents"
> file <- read.csv("Exercise1Assign6.csv")
> file
   T_treatment CO2_treatment Gender PupalWeight Frass
1      ambient           280      0       0.244 1.900
2      ambient           280      1       0.319 2.770
3      ambient           280      0       0.221    NA
4      ambient           280      0       0.280 1.996
5      ambient           280      0       0.257 1.069
6      ambient           280      1       0.333 2.257
7      ambient           280      0       0.275 2.198
8      ambient           280      1       0.312 1.873
```

2. T-Test Assumptions:

1. Data is independent.
2. Data is normally distributed.
3. Variance is unknown, but equal.

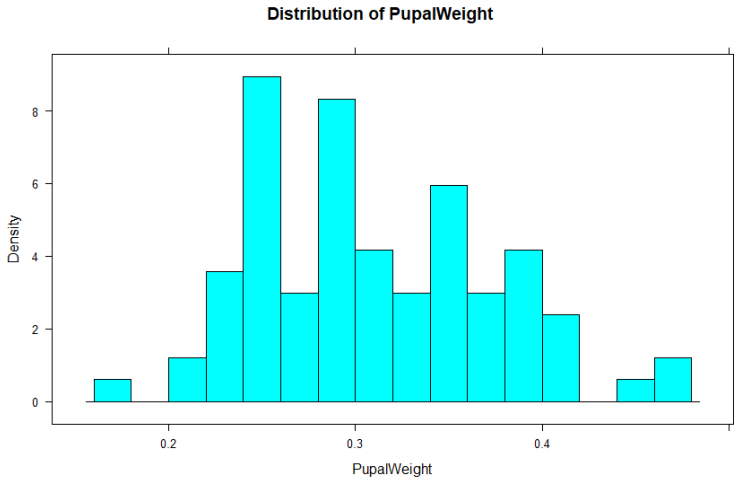
3. Does PupalWeight qualify for a t-test? YES

```
> summary(file)
 T_treatment         CO2_treatment       Gender        PupalWeight         Frass
 Length:84          Min.   :280.0   Min.   :0.0000   Min.   :0.1720   Min.   :0.986
 Class :character   1st Qu.:280.0   1st Qu.:0.0000   1st Qu.:0.2562   1st Qu.:1.515
 Mode  :character   Median :400.0   Median :0.0000   Median :0.2975   Median :1.818
                    Mean   :344.3   Mean   :0.4487   Mean   :0.3110   Mean   :1.846
                    3rd Qu.:400.0   3rd Qu.:1.0000   3rd Qu.:0.3560   3rd Qu.:2.095
                    Max.   :400.0   Max.   :1.0000   Max.   :0.4730   Max.   :3.117
                                    NA's   :6                         NA's   :1
```

A. Check for independence:
-Since the treatments data were measured on different groups of Pupal, the Pupalweight data is considered to be independent of each other.

B. Check for Normality:

```
summary(file)
histogram( ~ PupalWeight, dat=file, breaks=12, type="density", xlab="PupalWeight", main="Distribution of PupalWeight")

# Is PupalWeight Normal?
shapiro.test(file$Pupalweight)
```

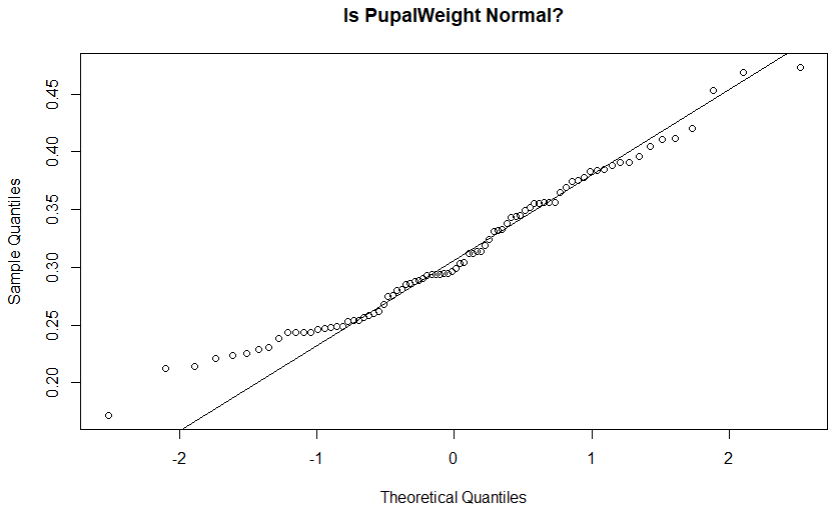



```
> shapiro.test(file$Pupalweight)

        Shapiro-Wilk normality test

data:  file$Pupalweight
W = 0.97488, p-value = 0.09943
```

H0: Distribution of Pupalweight is Normal.

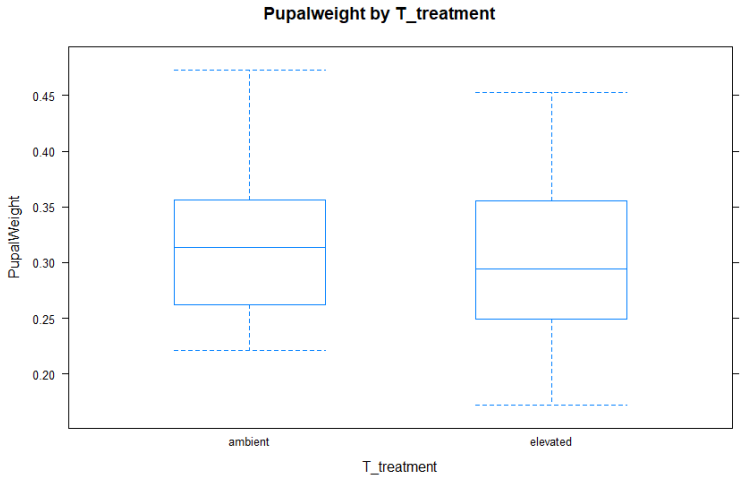
Conclusion: P-value of 0.09943 suggests that we failed to reject the hypothesis that “PupalWeight” is normal under a significance level of 0.05.

```
#QQLine:
qqnorm(file$Pupalweight, main="Is PupalWeight Normal?")
qqline(file$Pupalweight)
```



C. Check Variance using F-Test:

```
#Compare two treatment groups:
bwplot(Pupalweight ~ T_treatment, data=file,
       main="Pupalweight by T_treatment",
       xlab="T_treatment",  pch = '|')
```



From plot, observations are not so different in btw ambient and elevated.

```
> ftest <- var.test(PupalWeight ~ T_treatment, data = file)
> ftest

        F test to compare two variances

data:  PupalWeight by T_treatment
F = 1.0831, num df = 36, denom df = 46, p-value = 0.7907
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5863872 2.0494742
sample estimates:
ratio of variances
          1.083135
```

p-value = 0.7907 > 0.05 (Failed to reject hypothesis)
=> Therefore no significant difference in variances.

All 3 assumptions are met, we can now use T-test.

4. T-test on PupalWeight and T_treatment:

H0: There is no true difference in means btw group ambient and group elevated.
Level of significance: 0.05
All 3 T-test assumptions are met.

```
> # 4.T-test:
> res <- t.test(PupalWeight ~ T_treatment, data = file, var.equal = TRUE)
> res

        Two Sample t-test

data:  PupalWeight by T_treatment
t = 1.4385, df = 82, p-value = 0.1541
alternative hypothesis: true difference in means between group ambient and group elevated is not equal to 0
95 percent confidence interval:
 -0.007715698  0.048012420
sample estimates:
 mean in group ambient mean in group elevated
             0.3222973              0.3021489
```

p-value = 0.1541 > 0.05 (Failed to reject H0)
=> There is no true difference in means btw group ambient and group elevated.

5. Repeat above using Wilcoxon test:
H0: Group ambient and group have the same distribution and the same median.
Level of significance: 0.05
Assume that the distribution is not normal, we use a Wilcoxon Test.

```
> # 5.Wilcoxian Test: non-parametric test
> res_w <- wilcox.test(PupalWeight ~ T_treatment, data = file,
+                      exact = FALSE)
> res_w

        Wilcoxon rank sum test with continuity correction

data:  PupalWeight by T_treatment
W = 1017.5, p-value = 0.1838
alternative hypothesis: true location shift is not equal to 0
```

p-value = 0.1838 > 0.05 (Failed to reject hypothesis)
=> There is no true difference in medians btw group ambient and group elevated.

6. PupalWeight visualization:

In this case, a box-and-whisker is well suited as it shows that there is no significant difference in the means and distributions in btw the 2 T_treatment groups.

```
#Compare two treatment groups:
bwplot(Pupalweight ~ T_treatment, data=file,
       main="Pupalweight by T_treatment",
       xlab="T_treatment",  pch = '|')
```