repeated anova - parametric

Aung Thura Htoo

2024-03-28

```r
setwd("D:/me/R-Language/Practice")
my_data <- read.csv("repeated_anova.csv")
```

I set my working directory and loaded the assigned csv file in R.

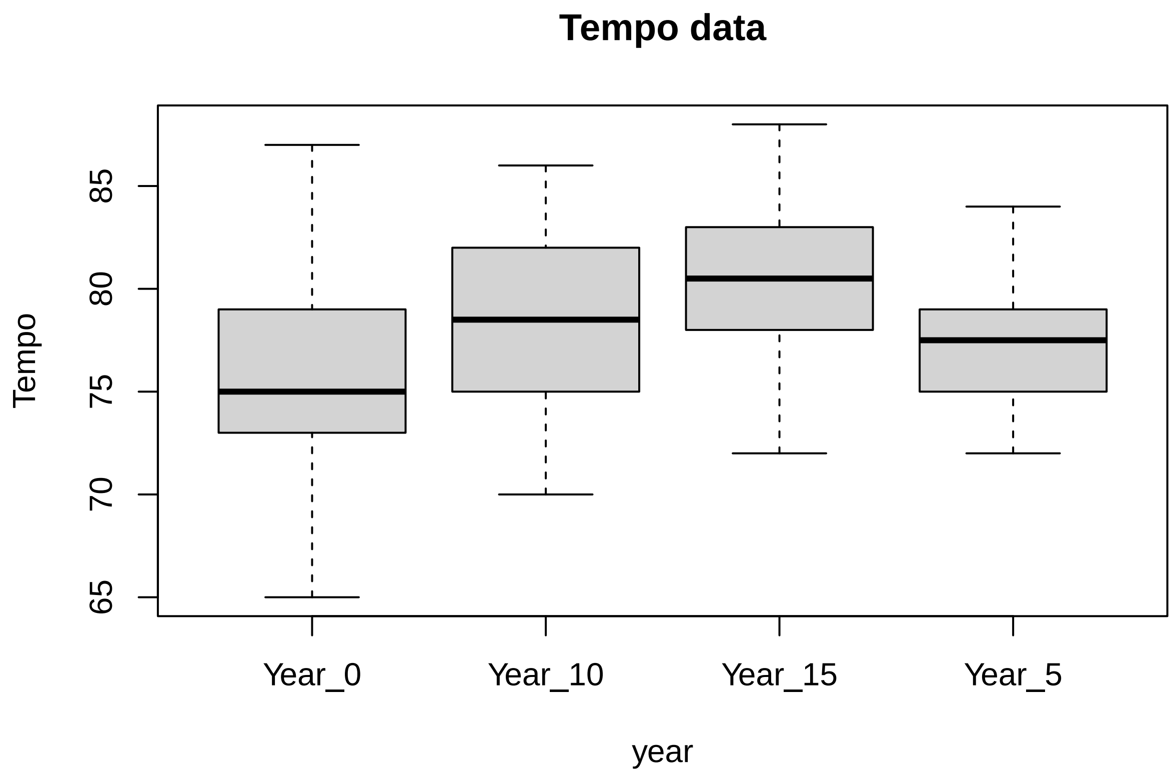
Then, I summarized the data in the assigned file to grasp the summary of the variables.

```r
summary(my_data)
```

```
##      year               tempo      
##  Length:120         Min.   :65.00  
##  Class :character   1st Qu.:75.00  
##  Mode  :character   Median :78.00  
##                     Mean   :77.96  
##                     3rd Qu.:81.00  
##                     Max.   :88.00
```

I plotted a box-plot of amylase level for each group to have a better understanding and visualization of the spread of my data.

```r
boxplot(tempo ~ year, data = my_data,
        main = "Tempo data",
        ylab = "Tempo",
        col = "lightgray")
```



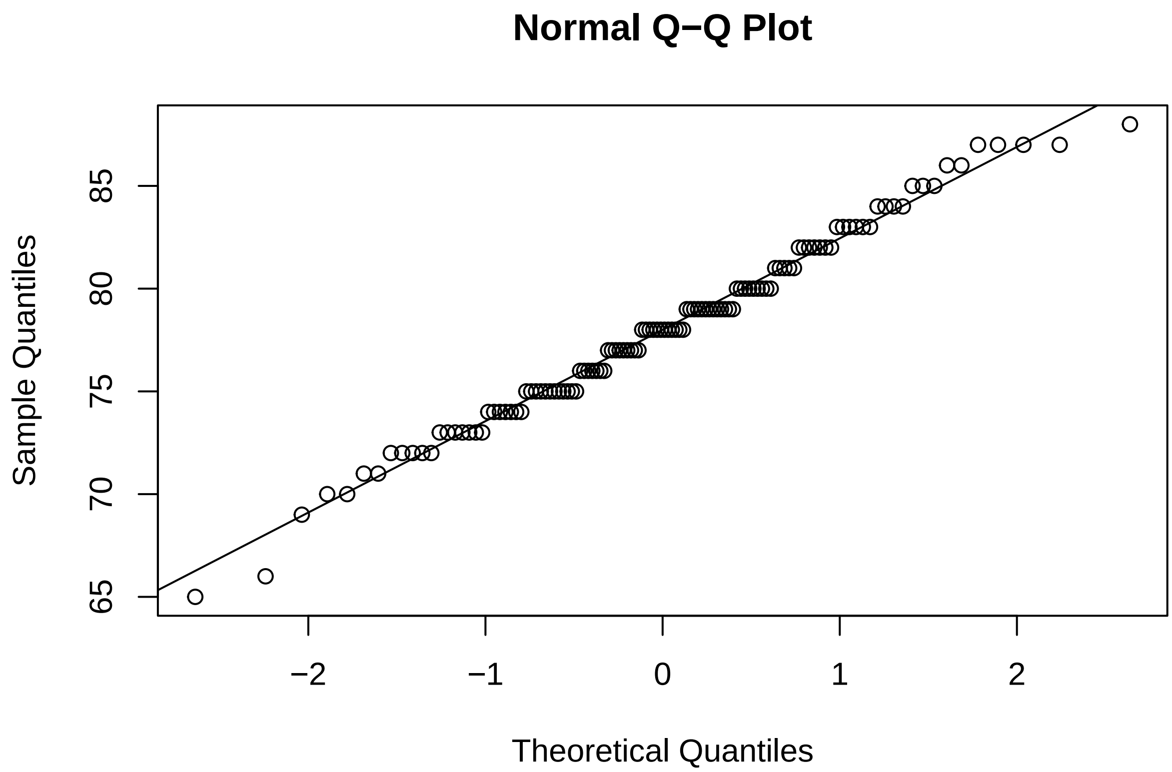
Before, I conduct the ANOVA, I will check if the assumptions for ANOVA are violated.

TESTING THE VALIDITY FOR ANOVA

To test the validity of our results, we need to check our ANOVA assumptions.

The first assumption is normality and independence of our samples. Therefore, I plotted Q-Q plots to check the normality.

```
qqnorm(my_data$tempo)
qqline(my_data$tempo)
```



Since most points fall approximately on the reference line, we can assume normality.

The second assumption is that measurements in each of the sampled populations are equally varied or have equal variances, homoscedasticity.

To check the second assumption, I computed the leveneTest.

```r
options(repos = c(CRAN = "https://cran.rstudio.com/"))
install.packages("car")
```

```
## Installing package into 'C:/Users/lenovo/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'car' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\lenovo\AppData\Local\Temp\Rtmp0YnRvF\downloaded_packages
```

```r
library(car)
```

```
## Warning: package 'car' was built under R version 4.3.3
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.3.3
```

```r
leveneTest(my_data$tempo ~ as.factor(my_data$year))
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value  Pr(>F)
## group   3   2.886 0.03868 *
##       116
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value of the results from levene test is 0.03868, which is less than 0.05. It means that there are differences in variance between the groups. Therefore, the assumption of equal variance is violated. And we cannot perform conventional ANOVA.

Since my data has three different groups and equal variance assumption is violated but the assumption of normal distribution is met, the most appropriate test is Welch's ANOVA.

```r
welch_anova <- oneway.test(tempo ~ year, data = my_data, var.equal = FALSE)
print(welch_anova)
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  tempo and year
## F = 7.0665, num df = 3.000, denom df = 62.494, p-value = 0.0003645
```

Since p-value is less than 0.05, it means that there are significance differences between three groups.

To understand which groups are different, I performed post-hoc tests: Tukey's Honestly Significant Difference (HSD) test

```r
install.packages("TukeyC")
```

```
## Installing package into 'C:/Users/lenovo/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'TukeyC' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\lenovo\AppData\Local\Temp\Rtmp0YnRvF\downloaded_packages
```

```r
library(TukeyC)
```

```
## Warning: package 'TukeyC' was built under R version 4.3.3
```

```r
tukey_result <- TukeyHSD(aov(tempo ~ year, data = my_data))
print(tukey_result)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
```

```
## Fit: aov(formula = tempo ~ year, data = my_data)
##
## $year
##                       diff         lwr       upr     p adj
## Year_10-Year_0   2.833333  0.01705066  5.649616 0.0480331
## Year_15-Year_0   4.966667  2.15038400  7.782949 0.0000644
## Year_5-Year_0    1.500000 -1.31628267  4.316283 0.5091735
## Year_15-Year_10  2.133333 -0.68294934  4.949616 0.2035968
## Year_5-Year_10  -1.333333 -4.14961600  1.482949 0.6065563
## Year_5-Year_15  -3.466667 -6.28294934 -0.650384 0.0092243
```

Adjusted p-values for the comparison between year_0 and year_15, year_10 and year_0, and between year_15 and year_5 are less than 0.05. This indicates significant differences in tempo during those years.

For the rest of the pairs, the adjusted p-values exceed 0.05. This suggests that there are no significant differences in tempo between those paired groups.