# Week 5

## PSTAT 122C

## 4/24/2022

```
library(knitr) # create tables in rmd
x<-c("A","A","B","B","B","C","C","C","D","D")
y<-c(12,18,14,12,13,19,17,21,24,30)
data <- data.frame(x, y)
aov <- aov(y~x, data)
summary(aov)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## x             3    258   86.00   11.22 0.00713 **
## Residuals     6     46    7.67
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
opts <- options(knitr.kable.NA = "")
sum_aov = summary(aov)
kable(sum_aov[[1]])
```

|           | Df | Sum Sq | Mean Sq  | F value  | Pr(>F)    |
|-----------|----|--------|----------|----------|-----------|
| x         | 3  | 258    | 86.000000| 11.21739 | 0.0071349 |
| Residuals | 6  | 46     | 7.666667 |          |           |

The table does not have a line for $SS_T$, but a full ANOVA table should include $SS_T$.

### LSD method

We can use the $t$ test for testing all pairs of mean parameters $H_0 : \mu_i = \mu_j$.

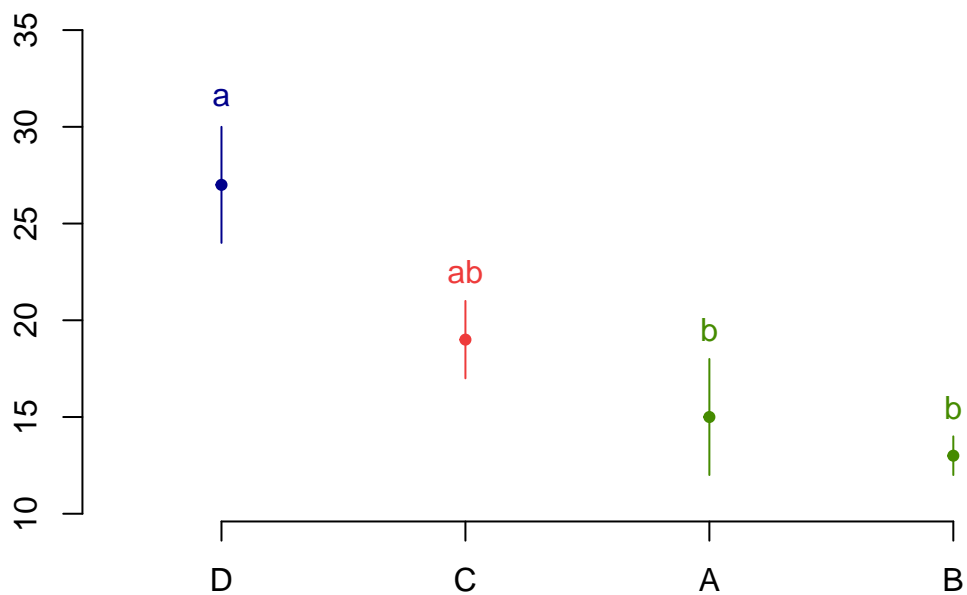$$t_0 = \frac{\bar{y}_{i.} - \bar{y}_{j.}}{\sqrt{MS_e(\frac{1}{n_i} + \frac{1}{n_j})}} \overset{H_0}{\sim} t_{N-a}$$

The pair of means $\mu_i$ and $\mu_j$ would be declared significantly different if $|\bar{y}_{i.} - \bar{y}_{j.}| > t_{\alpha/2, N-a}\sqrt{MS_e(\frac{1}{n_i} + \frac{1}{n_j})}$.

```
library(agricolae)
colnames(data) <- c("X","Y")
data
```

```
##    X  Y
## 1  A 12
## 2  A 18
## 3  B 14
## 4  B 12
## 5  B 13
## 6  C 19
## 7  C 17
## 8  C 21
## 9  D 24
## 10 D 30
```

```
model <- aov(Y~X, data)
out <- LSD.test(model,"X", p.adj="bonferroni")
plot(out)
```

**Groups and Range**



```
print(out)
```

```
## $statistics
##     MSerror Df Mean       CV
##   7.666667  6   18 15.38264
##
## $parameters
##        test  p.ajusted name.t ntr alpha
##   Fisher-LSD bonferroni      X   4  0.05
```

```
## 
## $means
##   Y       std r       LCL      UCL Min Max  Q25 Q50  Q75
## A 15 4.242641 2 10.209216 19.79078  12  18 13.5  15 16.5
## B 13 1.000000 3  9.088341 16.91166  12  14 12.5  13 13.5
## C 19 2.000000 3 15.088341 22.91166  17  21 18.0  19 20.0
## D 27 4.242641 2 22.209216 31.79078  24  30 25.5  27 28.5
## 
## $comparison
## NULL
## 
## $groups
##   Y groups
## D 27      a
## C 19     ab
## A 15      b
## B 13      b
## 
## attr(,"class")
## [1] "group"
```

```r
### compare level A and B
qt(0.975, df=6)*sqrt(7.666667*(1/2+1/3)) # > 2
```

```
## [1] 6.184876
```

```r
### compare level A and D
qt(0.975, df=6)*sqrt(7.666667*(1/2+1/2)) # < 12
```
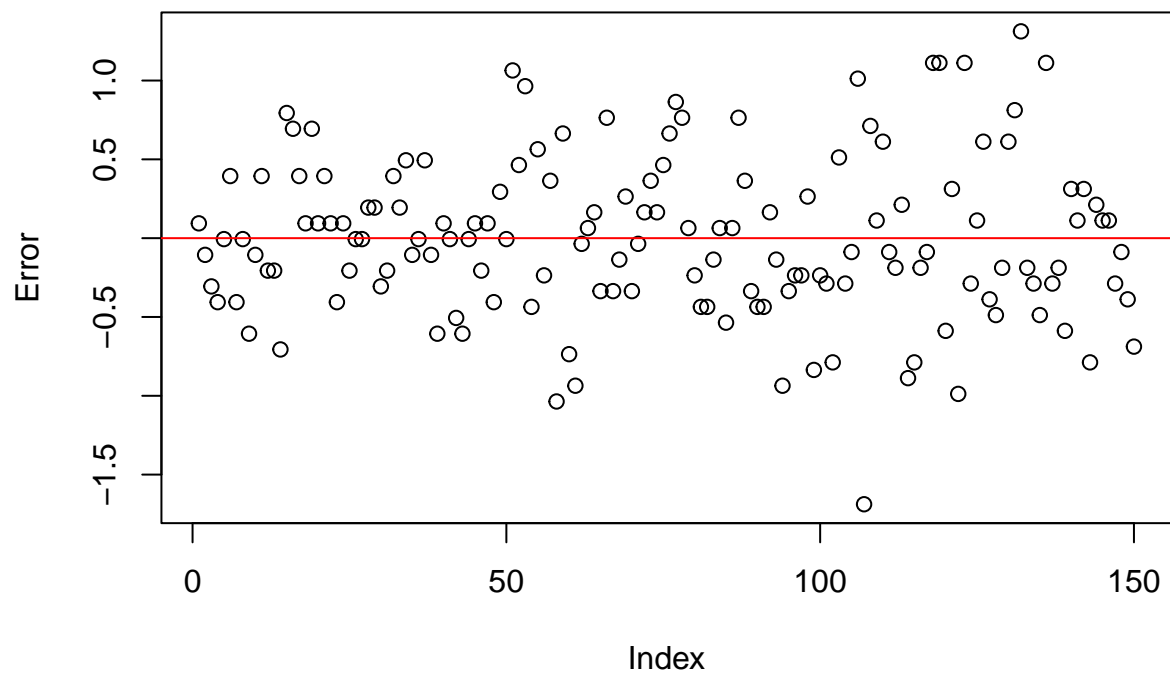
```
## [1] 6.775192
```

**Residuals**

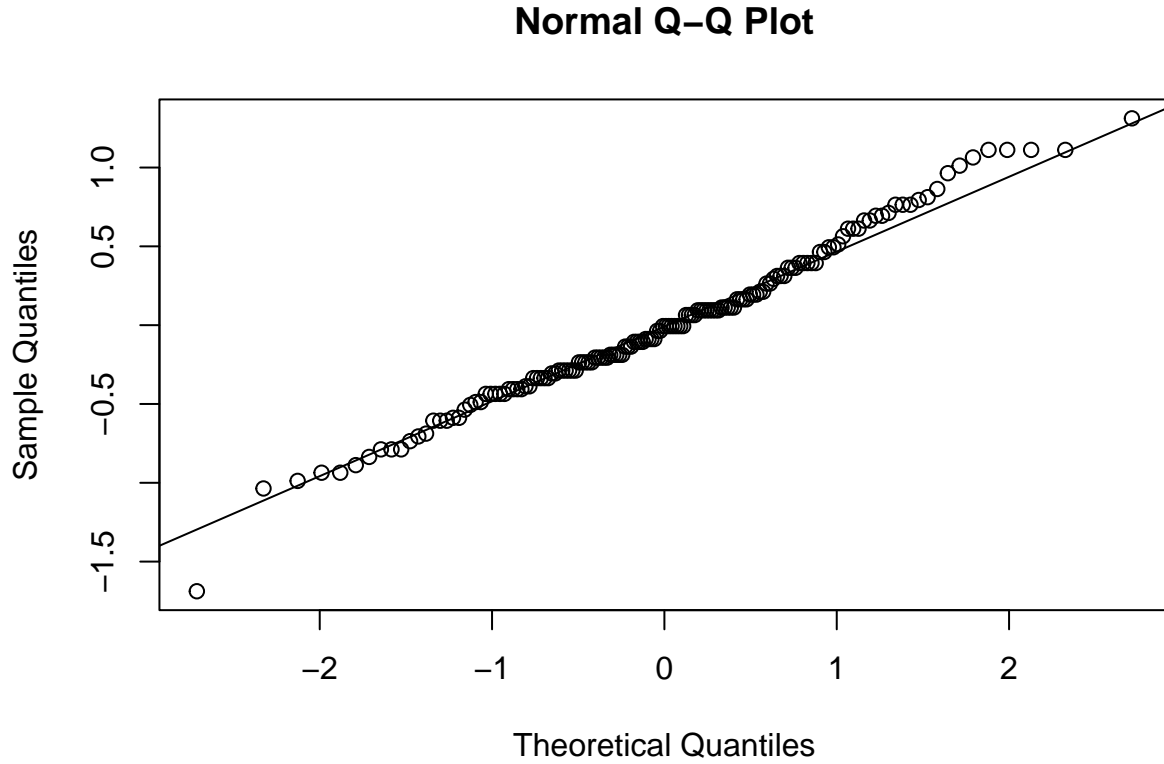$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$
$$\epsilon_{ij} \sim N(0, \sigma^2)$$

The estimates of means are $\hat{\mu}_i = \bar{y}_{i\cdot}$, so errors (residuals) are $e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i\cdot}$. Residual analysis is not trivial for model analysis.

```r
data(iris)
aov1 = aov(Sepal.Length ~ Species, iris)
res = aov1$residuals

plot(res, ylab="Error")
abline(h=0, col="red")
```

```
qqnorm(res)
qqline(res)
```

## Normal Q–Q Plot



**Review the random effects model**

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$
$$\epsilon_{ij} \sim N(0, \sigma^2)$$
$$\tau_i \sim N(0, \sigma_\tau^2)$$

We have to see the treatment effect as a random variable and $\tau_i$ is independent of $\epsilon_{ij}$. The variance of any observation is $(\sigma^2 + \sigma_\tau^2)$. The $H_0$ in the previous model is no longer meaningful, and the hypothesis we need to be concerned with is

$$H_0 : \quad \sigma_\tau^2 = 0$$
$$H_1 : \quad \sigma_\tau^2 > 0$$

One can prove that we still have $SS_e = SS_T - SS_A$. Under the null hypothesis, a statisitc $F_0 = \frac{SS_A/(a-1)}{SS_e/(n-a)} = \frac{MS_A}{MS_e} \sim F_{a-1,N-a}$, so we reject $H_0$ if $F_0 > F_{a-1,N-a}(\alpha)$

**Model Parameters**

$$\hat{\sigma}^2 = MS_e$$
$$\hat{\sigma}_\tau^2 = \frac{MS_A - MS_e}{n} \quad (balanced)$$
$$\hat{\sigma}_\tau^2 = \frac{MS_A - MS_e}{n_0}, \quad where \quad n_0 = \frac{1}{a-1}(\sum_{i=1}^{a} n_i - \frac{\sum_{i=1}^{a} n_i^2}{\sum_{i=1}^{a} n_i}) \quad (unbalanced).$$

**Note** It's possible that the estimation $\hat{\sigma}_\tau^2 < 0$. If this happens, take $\hat{\sigma}_\tau^2$ to be 0.