# ANOVA Diagnostics and Remedial Measures

- **ANOVA Model:**

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

where:

1. $Y_{ij}$ is the value of the response variable in the $j$th trial for the $i$th treatment.
2. $\mu_i$ is the population mean (parameter) for the $i$th treatment.
3. $\epsilon_{ij}$ are independent $N(0, \sigma^2)$.
4. $i = 1, \ldots, r$; $j = 1, \ldots, n_i$.

- **Residuals Analysis:**

1. Residuals: $e_{ij} = Y_{ij} - \hat{Y}_{ij}$
2. Semistudentized residuals: $e^*_{ij} = \frac{e_{ij}}{\sqrt{(MSE)}}$
3. Studentized residuals: $r_{ij} = \frac{e_{ij}}{s\{e_{ij}\}}$ where : $s\{e_{ij}\} = \sqrt{\frac{MSE(n_i - 1)}{n_i}}$
4. Studentized deleted residuals: $t_{ij} = e_{ij}\left[\frac{n_T - r - 1}{SSE(1 - \frac{1}{n_i}) - e^2_{ij}}\right]^{\frac{1}{2}}$

- **Diagnosis of Departures from ANOVA Model**. Residual plots can be helpful in diagnosing the following departures from the ANOVA Model:

1. Nonnormality of error terms - Use normal probability plots (qqplots) of residuals and the Shapiro-Wilk test.
2. Nonconstancy of error variance - Plot predicted values versus residuals (residual plot), and/or use the Brown-Forsythe test for equality or error variances.
3. Nonindependence of error terms - Look at residual sequence/index plot.
4. Outliers - Look at the residual plot.
5. Omission of important variables - Use different icons for values/points from different factor levels of an omitted variable.

```
data=read.csv("RustInhibitors.csv",header=T)
attach(data)

result=aov(score~factor(brand))
anova(result)
#               Df Sum Sq Mean Sq F value    Pr(>F)
# factor(brand)  3  15954  5317.8  866.12 < 2.2e-16 ***
# Residuals     36    221     6.1

mse=anova(result)$Mean[2]       # mse=6.139833
sse=anova(result)$Sum[2]        # sse=221.034

e=result$residuals
predicted=result$fitted

plot(predicted,e)
qqnorm(e); qqline(e)

# Shapiro-Wilk Test
shapiro.test(e)

e.star=e/sqrt(mse)          # Semistudentized residuals
data.frame(score,predicted,residuals=e,semistudentized=e.star)

ns=tapply(score,brand,length); nt=sum(ns); r=length(ns)
s.e=sqrt(mse*(ns[brand]-1)/ns[brand])
r=e/s.e                     # Studentized residuals
```

```
# Studentized deleted residual
t=e*sqrt((nt-r-1)/(sse*(1-1/ns[brand])-e^2))

plot(predicted,e.star)
plot(predicted,r)
plot(predicted,t)

qqnorm(t); qqline(t)
```

- **Brown-Forsythe test for equality or error variances.**

  The Brown-Forsythe test statistics is simply the ordinary $F^*$ statistic for testing differences in the treatment means, but based on the absolute deviations $d_{ij} = |Y_{ij} - \tilde{Y}_i|$, where $\tilde{Y}_i$ is the median of the $i$th treatment.

**ANOVA Table**

| Source | DF | Sum of Squares | Mean Square | $F$ |
|---|---|---|---|---|
| Treatments (Between) | $r-1$ | $SSTR = \sum_{i=1}^{r} n_i(\bar{d}_{i\cdot} - \bar{d}_{\cdot\cdot})^2$ | $MSTR = \frac{SSTR}{r-1}$ | $F^* = \frac{MSTR}{MSE}$ |
| Error (Within) | $n_T - r$ | $SSE = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(d_{ij} - \bar{d}_{i\cdot})^2$ | $MSE = \frac{SSE}{n_T - r}$ | |
| Total | $n_T - 1$ | $SSTO = SSTR + SSE$ | | |

```
data=read.csv("RustInhibitors.csv",header=T)
attach(data)

medians=tapply(score,brand,median)
d=abs(score-medians[brand])
result.bf=aov(d~factor(brand))
anova(result.bf)
#               Df Sum Sq Mean Sq F value Pr(>F)
# factor(brand)  3  1.837 0.61233  0.2262 0.8775
# Residuals     36 97.434 2.70650
```

- **Transformations of Response Variable.**

  1. Variance Proportional to $\mu_i : Y' = \sqrt{Y}$
  2. Standard Deviation proportional to $\mu_i : Y' = \log(Y)$
  3. Standard Deviation proportional to $\mu_i^2 : Y' = \frac{1}{Y}$
  4. Box-Cox Procedure: $Y' = Y^\lambda$. Choose $\lambda$ that minimizes the $SSE$.

- **Kruskal-Wallis Non-parametric Test.**

  The Kruskal-Wallis test is a widely used nonparametric test for testing the equality of treatment means. This test is based on the ranks $R_{ij}$ from 1 to $n_T$ and is defined as:

$$X^2_{KW} = \frac{SSTR}{SSTO/(n_T - 1)}$$

```
result.krus=kruskal.test(score,factor(brand))
result.krus
# Kruskal-Wallis chi-squared = 33.7041, df = 3, p-value = 2.288e-07
```