Stat 445/545: Analysis of Variance and Experimental Design

Chapter 18: ANOVA Diagnostics and Remedial Measures

Instructor: Yan Lu

THE UNIVERSITY OF NEW MEXICO

$$Y_{ij} = \mu_i + \varepsilon_{ij},\ \ i = 1, 2, \cdots, r; j = 1, 2, \cdots n_i$$

where $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$
Questions:

- Error Assumption: Are errors independent, normal random variables with common variance?
- Nonindependence of error terms
- Outlier Detection: Are there outliers i.e a response that is vastly different from other responses?
- Predictor Range: Are there one or more important predictor variables that have been omitted from the model?

Type of Residuals

- Raw residual: $e_{ij} = Y_{ij} - \bar{Y}_{i.}$
- Semi-studentized residual:

$$e_{ij}^* = \frac{e_{ij}}{\sqrt{MSE}}$$

  MSE is an approximation of the standard deviation of e_i

- Studentized Residual:

$$r_{ij} = \frac{e_{ij}}{s(e_{ij})} = \frac{e_{ij}}{\sqrt{MSE(1 - h_{ij})}}$$

  where $h_{ij,ij}$ is the diagonal element of hat matrix
  $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and $\hat{\mathbf{Y}} = \mathbf{HY}$
  —-for ANOVA model, the leverage of Y_{ij} is given by

$$h_{ij,ij} = \frac{1}{n_i}, \text{ therefore } s(e_{ij}) = \sqrt{MSE(n_i - 1)/n_i}$$

$$
\begin{aligned}
e_{ij} &= Y_{ij} - \bar{Y}_{i.} \\
&= Y_{ij} - \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i} \\
&= \frac{n_i - 1}{n_i} Y_{ij} - \frac{\sum_{k \neq j} Y_{ik}}{n_i}
\end{aligned}
$$

$$
\begin{aligned}
V(e_{ij}) &= \left(\frac{n_i - 1}{n_i}\right)^2 \sigma^2 + \frac{n_i - 1}{n_i^2} \sigma^2 \\
&= \frac{(n_i - 1)^2}{n_i^2} \sigma^2 + \frac{n_i - 1}{n_i^2} \sigma^2 \\
&= \frac{n_i - 1}{n_i} \sigma^2
\end{aligned}
$$

- Studentized deleted residual (rstudent):
  ——t_{ij} is the studentized residual with Y_{ij} deleted from the analysis
  ——Let $\bar{Y}_{i.(j)}$ be the mean of Y's in ith level with jth observation deleted
  ——deleted residual

$$e_{ij(j)} = Y_{ij} - \bar{Y}_{i.(j)}$$

$$t_{ij} = \frac{e_{ij(j)}}{s(e_{ij(j)})}$$

After some algebra,

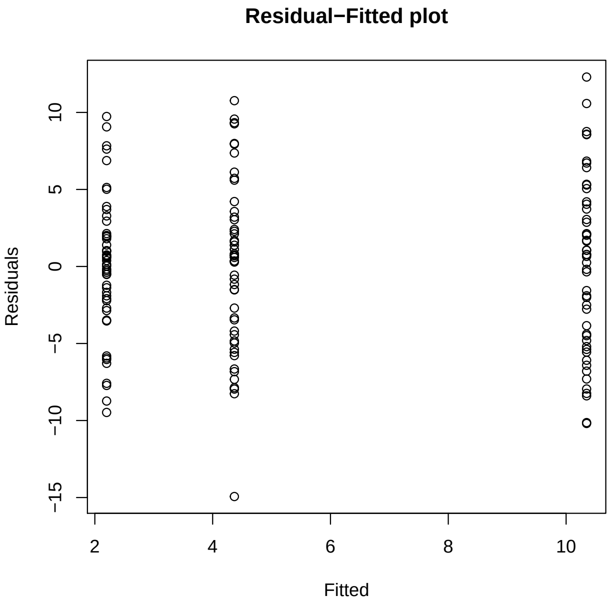
$$t_{ij} = e_{ij} \left[\frac{n - r - 1}{SSE\left(1 - \frac{1}{n_i}\right) - e_{ij}^2} \right]^{1/2}$$

Diagnosis of Departures from ANOVA Model

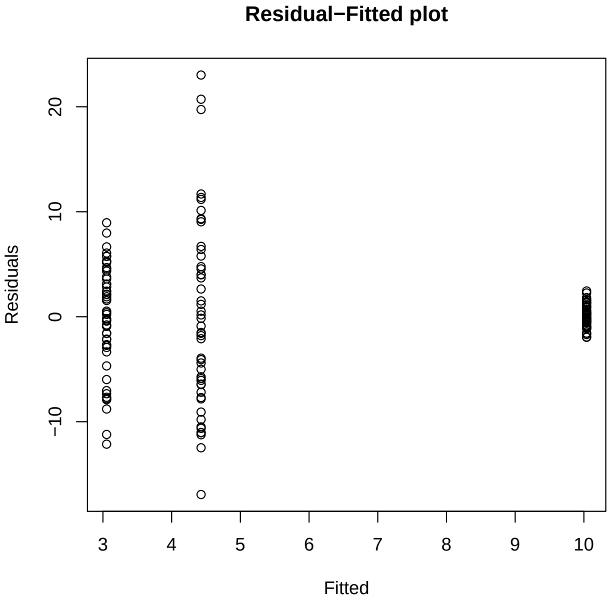
- Nonconstancy of error variance
- Nonindependence of error terms
- Outliers
- Omission of important explanatory variables
- Nonnormality of error terms

1. Checking for non-constant error variance
    - Plot Residuals v.s Fitted Values
      ——When the group sizes are not large and do not differ greatly, the appropriateness of this assumption can be studied by using the residuals, semistudentized residuals, or studentized residuals.
      ——When the group sizes differ greatly, studentized residuals should be used in these plots.
    - If error variance is constant, these plots have about the same extent of scatter of the residuals aorund zero $e = 0$ for each factor level.
      ——$\sum_{j=1}^{n_i} e_{ij} = 0$
    - If there is a problem with the constancy of variance assumption, e_i's will show some trend.

```
> ## Generate Data
> mu1 <- 10
> mu2 <- 3
> mu3<- 5
> eps1 <- rnorm(50, mean=0, sd=5)  #constant variance
> eps2 <- rnorm(50, mean=0, sd=5)
> eps3 <- rnorm(50, mean=0, sd=5)
> y1 <- mu1 + eps1
> y2 <- mu2 + eps2
> y3 <- mu3 + eps3
> y<-c(y1,y2,y3)
> group<-c(rep(1,50),rep(2,50),rep(3,50))
> ex.data<-data.frame(y,group)
> myfit = aov(y ~ factor(group), data=ex.data)
> plot(myfit$fit,myfit$resid,xlab="Fitted",
+      ylab="Residuals",main="Residual-Fitted plot")
>
```


Residual−Fitted plot
Residuals
10
5
0
−5
−10
−15
2
4
6
8
10
Fitted

```
> ## Generate Data
> mu1 <- 10
> mu2 <- 3
> mu3<- 5
> eps1 <- rnorm(50, mean=0, sd=1)  #unequal variance
> eps2 <- rnorm(50, mean=0, sd=5)
> eps3 <- rnorm(50, mean=0, sd=10)
> y1 <- mu1 + eps1
> y2 <- mu2 + eps2
> y3 <- mu3 + eps3
> y<-c(y1,y2,y3)
> group<-c(rep(1,50),rep(2,50),rep(3,50))
> ex.data<-data.frame(y,group)
> myfit2 = aov(y ~ factor(group), data=ex.data)
> plot(myfit2$fit,myfit2$resid,xlab="Fitted",
+      ylab="Residuals",main="Residual-Fitted plot")
>
```


Residual−Fitted plot
Residuals
Fitted
20
10
0
−10
3
4
5
6
7
8
9
10

2. Checking for independence of Error terms
Whenever data are obtained in a time sequence, a residual sequence plot should be prepared to examine if the error terms are serially correlated to the sequence in which the observations are obtained.

- Plot e_i v.s time sequence (order in which data is collected)
- Independence would result in a random scatter of points.
- Non independence would result in a trend in the plot

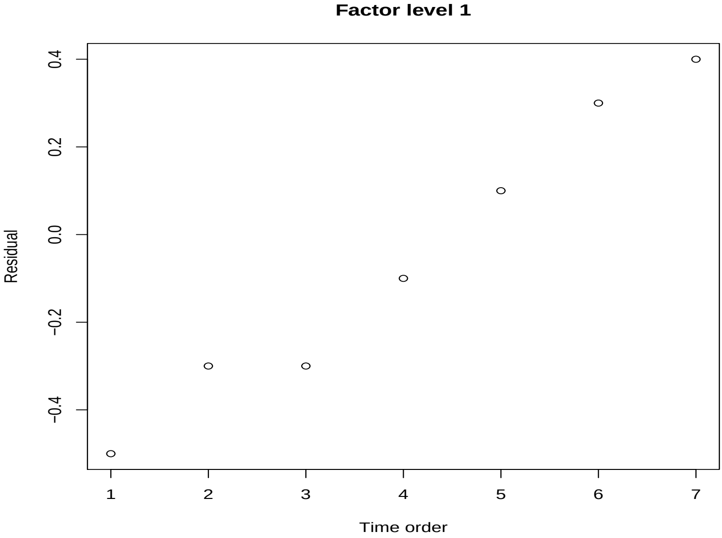
Example: an experiment on group interactions.

- Three different treatments were applied, and the group interactions were recorded on videotapes.
- Seven replications were made for each treatment.
- Aterward, the experimenter measured the number of interactions by viewing the tapes in randomized order.
- Figure 18.3 strongly suggests that the experimenter discerned a larger number of interactions as more experience in viewing the tapes was gained.

  ——As a result, the residuals in Figure 18.3 appear to be serially correlated.

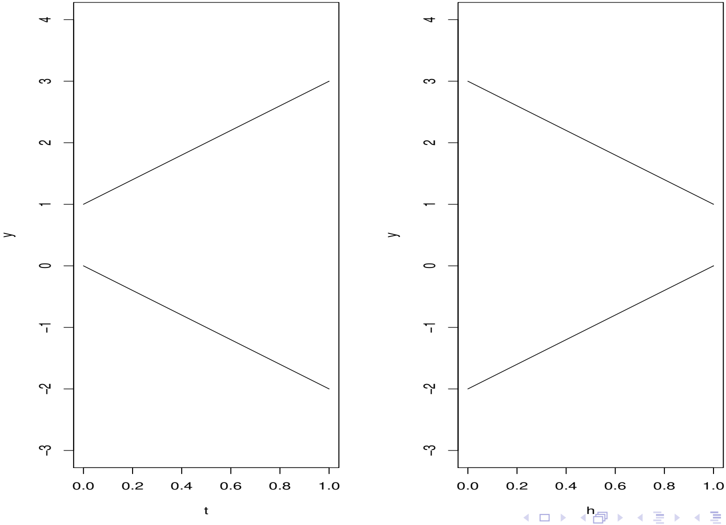
  ——In this instance, an inclusion in the model of a linear term for the time effect might be sufficient to assure independence of the error terms in the revised model.

Figure: 18.3 for level 1



- Time-related effects may also lead to increases or decreases in the error variance over time.
  —-For instance, an experimenter may make more precise measurements over time.
  —-On the other hand, an experimenter may make less precise measurements over time when it is close to midnight.

Figure: A residual plot that displays an increasing variance looks roughly like a horn opening to the right, and a residual plot indicating a decreasing variance is a horn opening to the left.



3. Checking for Outliers
   - Outliers are observations whose values are far from “typical” values in the sample
   - Plot studentized deleted residuals vs fitted value,
   - Outliers (y observations) can easily be spotted on a residual plot, especially if studentized residuals are used. Look for residuals that are far from the main set of residual values.
   - It is wise practice to discard outlying observations only if they can be identified as being due to such specific casuses as instrumentation malfunctioning, miscalculation, oberver measurement blunder, or recording error.

4. Omission of Important Explanatory Variables

- Residual analysis may also be used to study whether or not the signle-factor ANOVA model is an adequate model.
  Example: In a learning experiment involving three motivational treatments, the residual shown in Figure 18.5 show no unusual overall pattern.
  But the treatment effects do fiffer according to gender, with residual for female subjects are below 0, and residuals for male subjects are above 0.
  We may need to add gender as an explanatory variable.

5. Checking for normality
   - Whether the distribution of the errors is far enough away from normal. Look at the residuals' distribution. Use a normal quantile plot (qqplot) and a histogram plot.
   - Normal quantile plot: A graph of the residuals versus the expected order statistics of the standard normal distribution. It plots quantiles of the data versus quantiles of a distribution. The Q-Q plot may be constructed using raw, studentized or jack-knifed residuals

     —-If the observations come from a normal distribution we would expect the observed order statistics to be reasonably close to the expected order statistics. We should get approximately a straight line

     —-In general, Q-Q plots showing curvature indicate skew distributions, with downward concavity corresponding to negative skewness (long tail to the left) and upward concavity indicating positive skewness. On the other hand, S-shaped Q-Q plots indicate heavy tails, or an excess of extreme values, relative to the normal distribution.

For the simulated data with equal variance, the following normal qq plot shows approximately a straight line.

```
> par(mfrow=c(2,2)) # optional layout
> plot(myfit) # diagnostic plots
```

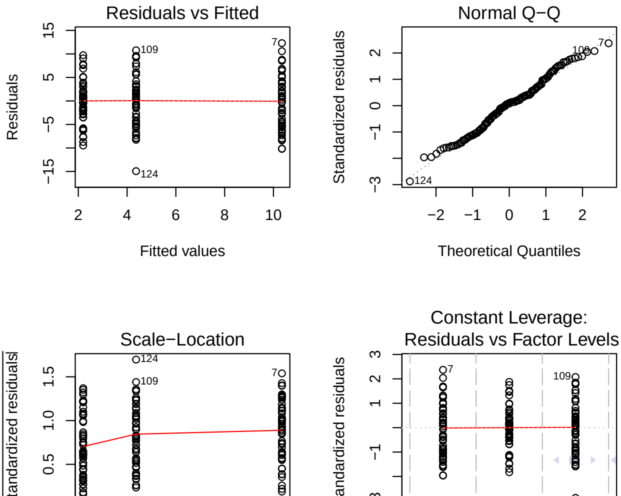
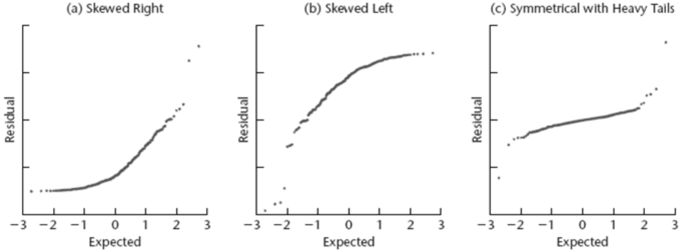


Figure: Normal probability plots when error term distribution is not normal

FIGURE 3.9 **Normal Probability Plots when Error Term Distribution Is Not Normal.**



(a) Skewed Right

(b) Skewed Left

(c) Symmetrical with Heavy Tails

Robustness:

A statistical procedure is robust if it is not greatly affected by a violation of the assumptions

- Nonnormality: F test is robust with respect to nonnormality, the larger the n is, the more robust it is. Don't worry about moderate departures from normality
- Unequal variances: F test is robust if all n_i's are equal, balanced designs are good. But F test is not robust for unbalanced design
  ——If big n_i's go with big variances, then F^* is small, P-value from test $>$ true p value, don't reject H_0 often enough
  ——If big n_i's go with small variance, then F^* is large, P-value from test $<$ true p-value, reject more often than we should

$$\begin{aligned} MSE &= \frac{\sum_i \sum_j (n_i - 1) S_i^2}{n - r} \\ F^* &= \frac{MSTR}{MSE} \end{aligned}$$

- Nonindependence: F test is not robust, if most common observations are positively correlated, MSE too small, MSTR/MSE too large, p-value too small, reject too often
——solutions: randomize as much as possible, collect data in a way that independence assumption is justified
——or use method for analysis of dependent data

Some Tests for Checking Assumptions:
Brown-Forsythe Test (modified Levene's test)

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_r^2$$

- BF test is robust from departures from normality, which often occur together with unequal variances. It doesn't require equal sample sizes.
- Test procedure:
  (1) Let $\tilde{Y}_i =$ median of Y_{ij}, for example, $\tilde{Y}_1$ is the median of $Y_{11}, Y_{12}, \cdots, Y_{1n_1}$
  (2) Define
  $$d_{ij} = |Y_{ij} - \tilde{Y}_i|$$
  absolute deviation between observation Y_{ij} and median for the observations in that level.

(3) Use d_{ij} to perform the one-way ANOVA

$$F^*_{BF} = \frac{MSTR}{MSE}$$

where

$$MSTR = \frac{\sum_{i=1}^{r} n_i(\bar{d}_{i.} - \bar{d}_{..})^2}{n-1}$$

$$MSE = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i}(d_{ij} - \bar{d}_{i.})^2}{n-r}$$

(4) Reject $H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_r^2$ if $F^*_{BF} > F(1-\alpha; r-1, n-r)$

Example: The ABF Electronics Corporation performed an experiment to evaluate five types of flux for use in soldering printed circuit boards. A major concern of the firm's reliability engineers was the strentgh of the soldered joints. To test the five types of flux,

- 40 printed circuit boards were selected at random,
- Each of the five flux types was randomly assigned to 8 of the 40 circuit boards and an electornic switch was slodered to each board using the designated flux type.
- The force (in pounds) required to break a joint, termed the pull strength, is the response of interest.

```
> abt<-read.table(file=
+ "~/Desktop/jenn/teaching/stat445545/data/CH18TA02.txt",
+                     col.names=c("force","type","obs"))
> nrow(abt)

[1] 40

> abt[1:10,]

   force type obs
1  14.87    1   1
2  16.81    1   2
3  15.83    1   3
4  15.47    1   4
5  13.60    1   5
6  14.76    1   6
7  17.40    1   7
8  14.62    1   8
9  18.43    2   1
10 18.76    2   2
```

```
> abt$type<-factor(abt$type)
> attach(abt)
> tapply(force,type,mean) #group means
       1        2        3        4        5
15.42000 18.52750 15.00375  9.74125 12.34000
> med<-tapply(force,type,median) #group medians
> med
     1      2      3      4      5
15.170 18.595 15.255 10.010 12.105
> tapply(force,type,var) #group variance
        1         2         3         4         5
1.5305143 1.5699357 6.1833982 0.6668411 0.5920000
> abt$temp<-c(rep(med[1],8),rep(med[2],8),rep(med[3],8),
+             rep(med[4],8),rep(med[5],8))
> abt$dij<-abs(abt$force - abt$temp)
> myfit<-aov(abt$dij~type)
> summary(myfit)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
type         4  9.348   2.337   2.936 0.0341 *
Residuals   35 27.861   0.796
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```

```
> library(car)
> leveneTest(force~type)
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  4  2.9358 0.03414 *
      35
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```

The absolute deviations for the first observation is

$$d_{11} = |Y_{11} - \tilde{Y}_1| = |14.87 - 15.17| = 0.300$$

The $F^*_{BF} = 2.94, F(0.95; 4, 35) = 2.64$
Since $F^*_{BF} = 2.94 > 2.64$, we coclude H_α, that the error terms do not have constant variance.

Tests for Normality

- correlation test, based approximately on the coefficient of correlation between the ordered residuals and their expected values under normality H_0 : error terms are normally distributed
  Reject H_0, if correlation $<$ critical value given in Table B.6
- Lilliefors's test
- Shapiro-Wilk Test

```
> shapiro.test(myfit$resid)

        Shapiro-Wilk normality test


data:  myfit$resid
W = 0.96613, p-value = 0.2696
```

In the ABT example, p-value for normality test is 0.2696, we do not reject the normality assumption.

Test for outliers

Rstudent t_{ij} is the studentized deleted residual with Y_{ij} being deleted from the analysis

$$d_{ij} = Y_{ij} - \bar{Y}_{.(j)}$$

$$t_{ij} = \frac{d_{ij}}{se(d_{ij})} = e_{ij}\left[\frac{(n-r-1)n_i}{SSE(n_i-1)-n_i e_{ij}^2}\right]^{1/2}$$

$$H_0: \text{ observation } Y_{ij} \text{ is not an outlier}$$

Under H_0, $t_{ij} \sim t_{n-r-1}$

- Examine the largest absolute t_{ij},
- The appropriate Bonferroni critical value at α level test is $t\left(1-\frac{\alpha}{2n}, n-r-1\right)$.
- Rejects H_0 if

$$\max|(t_{ij})| \geq t\left(1-\frac{\alpha}{2n}, n-r-1\right).$$

```
> rresid<-rstudent(myfit)
> rresid[1:24]
```

```
         1          2          3          4          5
-0.7834677  0.8138268 -0.3519790 -0.7834677  0.7289769 -0.6
         8          9         10         11         12
-0.4828940 -0.9053050 -0.9053050  0.7168938 -0.4828940  0.3
        15         16         17         18         19
 0.6085359  1.5553304 -0.2319493  1.3118722  1.6757417  0.2
        22         23         24
-1.7057342  2.9083377 -2.0340581
```

```
> outlierTest(myfit)
```

```
No Studentized residuals with Bonferonni p < 0.05
Largest |rstudent|:
   rstudent unadjusted p-value Bonferonni p
23 2.908338          0.0063599      0.25439
```

Summary for Diagnostics

- Plot, plot, plot, always check the plots first. Although graphic analysis of residuals is only an informal method of analysis, in many cases it suffices for examining the aptness of a model. ——Use the significance tests if you are uncertain what to conclude after examining the plots. Tests are not a replacement for the plots, but a supplement to them.
- Plot residuals (t_{ij} for checking outliers) vs. fitted values (check for constant variance, outliers,important predictor variables)
- qqplot and histogram of residuals (check normality)
- Plot residual vs. time sequqence (check independence)
- In practice, several types of departures may occur together.

Transformation

When both the model assumptions of constancy of the error variance and normality of the error distributions are violated, a transformation of the response variable is often useful.
Simple guides to finding a transformation
If $\sigma_i \approx c(\mu_i)^a$ use transformation

$$Y'_{ij} = \begin{cases} Y_{ij}^{1-a} & \text{if } a \neq 1 \\ log(Y_{ij}) & \text{if } a = 1 \end{cases}$$

Table: Variance stabilizing transformations

Mean, Variance Relationship	Transformation
$\sigma_i \propto \sqrt{\mu_i}$	$\sqrt{Y_{ij}}$ or $\sqrt{Y_{ij}} + \sqrt{Y_{ij}+1}$
$\sigma_i \propto \mu_i$	$\log(Y_{ij})$
$\sigma_i \propto \mu_i^2$	$1/Y_{ij}$
Response is a proportion	$2\sin^{-1}(\sqrt{y_h})$

- Square root transformation is often found when Y is a count, for example, number of attempts by a subject before the correct solution is found.
- Log transformaiton is often found when Y is the amount, for example, revenue of a company.
- use s_i to estimate σ_i, use $\bar{Y}_{i.}$ to estimate μ_i, plot $s_i^2/\bar{Y}_{i.}$, $s_i/\bar{Y}_{i.}$ or $s_i/\bar{Y}_{i.}^2$ for each level, approximate constancy of the three statistics over all factor levels would suggest the corresponding tranformation as useful for stabilizing the error variance and making the error distributions more nearly normal.

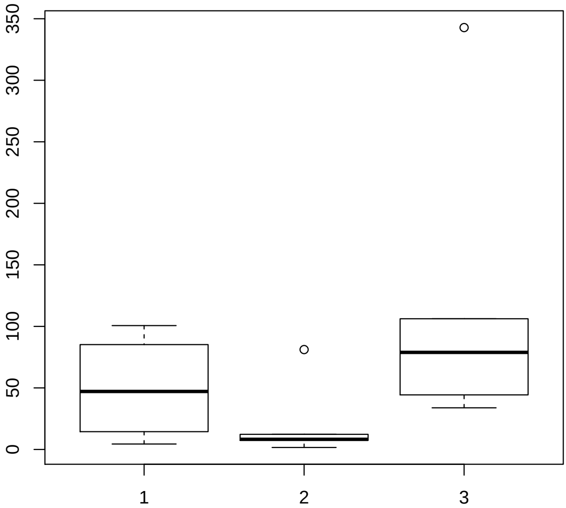
—-or plot $log(s_i)$ versus $log(\bar{Y}_{i.})$, if slope is β, try transformation with power of Y of $1-\beta$

$$\sigma_i = k\mu_i^{1-a}$$

$$log(\sigma_i) = log(k) + (1-a)log(\mu_i).$$

Example: Servo-data, Inc, operates mainframe computers at three different locations. The computers are identical as to make and model, but are subject to different degrees of voltage fluctuation in the power lines serving the respective installations. Y_{ij}: the lengths of time beteen computer failures for the three locations, for five failure intervals each.

```
> servo<-read.table(file=
+ "~/Desktop/jenn/teaching/stat445545/data/CH18TA05.txt",
+                     col.names=c("time","location","obs"))
> nrow(servo)

[1] 15

> servo[1:6,]

    time location obs
1   4.41        1   1
2 100.65        1   2
3  14.45        1   3
4  47.13        1   4
5  85.21        1   5
6   8.24        2   1

> servo$location<-factor(servo$location)
> attach(servo)
> boxplot(time~location)
```

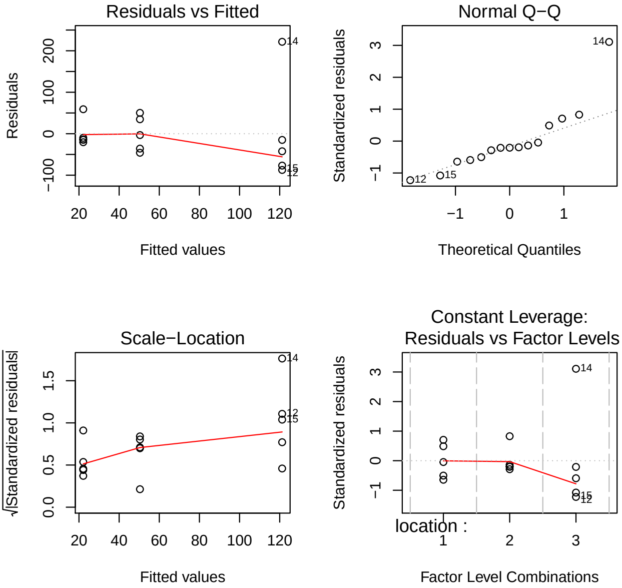

0
50
100
150
200
250
300
350
1
2
3

```
> m<-tapply(time,location,mean) #group means
> v<-tapply(time,location,var) #group variance
> ratio1<-v/m
> ratio1

        1         2         3
 35.51206  49.86237 133.38597

> ratio2<-sqrt(v)/m
> ratio2

        1         2         3
0.8396571 1.5010521 1.0490337

> ratio3<-sqrt(v)/m^2
> ratio3

          1           2           3
0.016669785 0.067828835 0.008654822
```

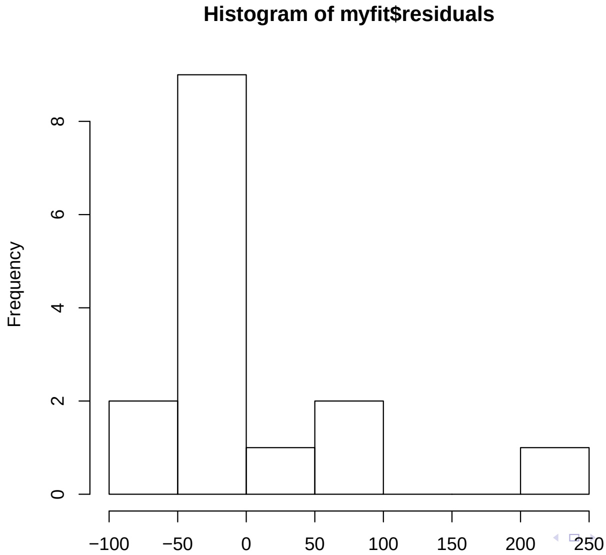
```
> myfit<-aov(time~location)
> summary(myfit)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
location     2  26053   13027    2.05  0.171
Residuals   12  76239    6353
```

```
> par(mfrow=c(2,2))
> plot(myfit)
```


Residuals vs Fitted
Residuals
−100
0
100
200
20
40
60
80
100
120
Fitted values
14
15
12
Normal Q−Q
Standardized residuals
−1
0
1
2
3
−1
0
1
Theoretical Quantiles
14
12
15
Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
20
40
60
80
100
120
Fitted values
14
12
15
Constant Leverage:
Residuals vs Factor Levels
Standardized residuals
−1
0
1
2
3
location :
1
2
3
Factor Level Combinations
14
15
12

```
> par(mfrow=c(1,1))
> hist(myfit$residuals)
```



- Residual versus fitted value plot looks like a horn opening to the right, with more residuals below 0;
- QQ plot shows upward concavity, indicating skewness to the right.

May try transformations to fix the problems.

Table: Statistics based on Servo-Data

i	$\frac{s_i^2}{\bar{Y}_{i.}}$	$\frac{s_i}{\bar{Y}_{i.}}$	$\frac{s_i}{\bar{Y}_{i.}^2}$
1	35.5	0.84	0.017
2	49.9	1.50	0.068
3	133.4	1.05	0.009

The relationship $s_i/\bar{Y}_{i.}$ is the most stable, hence the logarithmic transformation may be helpful here.

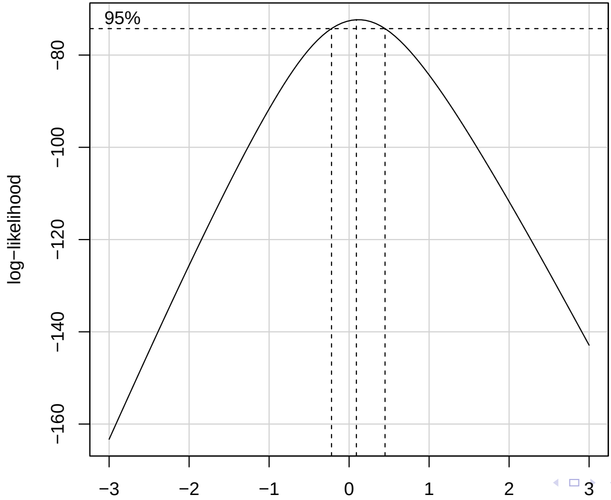
Box-Cox Transformation

Commonly used transformation is the power transformation (Box-Cox transformation)

$$y^* = \begin{cases} y^\lambda & \lambda \neq 0 \\ \ln y \quad \text{if} & \lambda = 0. \end{cases}$$

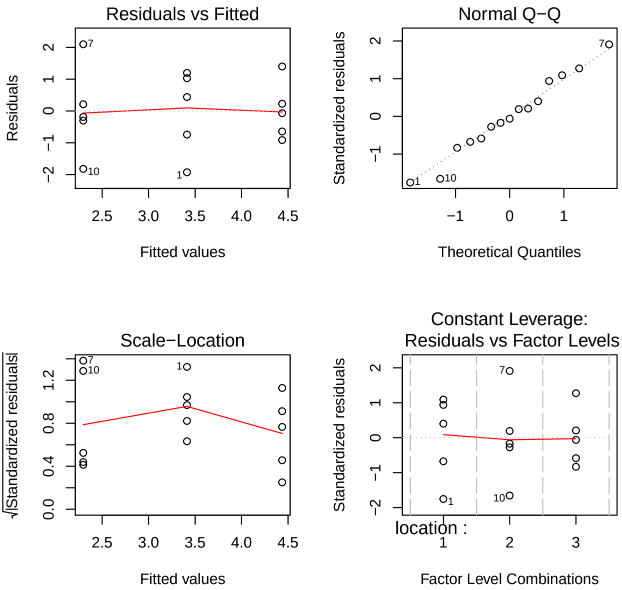
Power transformation relies on the method of maximum likelihood estimation to estimate the exponent λ that provides data that are approximately normal in distribution and have approximate constant variance across the levels.

```
> #box-cox transformation
> boxCox(myfit, lambda = seq(-3, 3, length = 10))
```



- $\lambda = 1$ is not included in the 95% confidence interval,
- $\lambda = 0$ is included in the 95% confidence interval, go ahead with log transformation

```
> time2<-log(time)
> tapply(time2,location,mean) #group means after transforma

       1        2        3
3.412849 2.297029 4.436669

> myfit2<-aov(time2~location)
> summary(myfit2)

            Df Sum Sq Mean Sq F value Pr(>F)
location     2  11.45   5.726   3.789  0.053 .
Residuals   12  18.14   1.511
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '

> par(mfrow=c(2,2))
> plot(myfit2)
```


Residuals vs Fitted
Residuals
Fitted values
2.5
3.0
3.5
4.0
4.5
−2
−1
0
1
2
7
10
1
Normal Q−Q
Standardized residuals
Theoretical Quantiles
−1
0
1
2
7
1
10
Scale−Location
√|Standardized residuals|
Fitted values
0.0
0.4
0.8
1.2
7
10
1
Constant Leverage:
Residuals vs Factor Levels
Standardized residuals
Factor Level Combinations
location :
1
2
3
7
1
10

Analyze transformed data:
Testing $H_0 : \mu_1 = \mu_2 = \mu_3$

$$F^* = \frac{MSTR}{MSE} = \frac{5.7264}{1.5112} = 3.789$$

For $\alpha = 0.05, p - value = 0.053$, we do not reject equal mean hypothesis.
If we take $\alpha = 0.10, F(0.90; 2, 12) = 2.81$, since $F^* = 3.789 > 2.81$, we conclude H_α, that the three means are not equal.

Multiple comparisons:
The transformed means for the three groups are 3.412849, 2.297029, and 4.436669, respectively.
Conduct Boneferroni pairwise comparison at 0.10 level,

$$s^2(\hat{D}) = MSE\left(\frac{1}{n_i} + \frac{1}{n_i}\right) = 1.511\left(\frac{1}{5} + \frac{1}{5}\right) = 0.6044$$

So that $s(\hat{D}) = 0.775$

$$B = t(1 - 0.05/3; n - r) = t(0.9833; 12) = 2.402.$$

The resulting 90 percent Bonferroni pairwise confidence intervals are

$$-2.984 \le \mu_2 - \mu_1 \le 0.752$$
$$-0.884 \le \mu_3 - \mu_1 \le 2.892$$
$$0.272 \le \mu_3 - \mu_2 \le 4.008$$

Therefore, at 90% confidence, we conclude that location 3 has longer average time computer failures than location 2.

Check model assumptions:

- After transformation, the residual versus fitted value plot looks randomly scattered around 0 for each level,
- Normal qq plot appears to be more reasonably close to a straight line.
- We don't have information on time order for collecting data, therefore, independence assumption can't be assessed by residual plots.
- No outlier is detected from residual plots.