Design of Engineering Experiments
Part 3 – Analysis of Variance (5)

- Comparing more that two factor levels…**the analysis of variance**
  - ANOVA decomposition of total variability
  - Statistical testing & analysis
  - Checking assumptions, model validity
  - Post-ANOVA testing of means
  - Box-Cox method
- **Sample size** determination

Selecting a Transformation (1)

- The inequality of variance problem occurs relatively often in practice, often in conjunction with a nonnormal response variable.
- e.g. a count of defects or particles, proportion data, a response variable following skewed distribution (one “tail” of the response distribution is longer than the other)

Selecting a Transformation (2)

- Generally, **transformation of the response variable** are used for three purposes:
  - **stabilizing the variance of the response variable**
  - **Making the distribution of the response variable closer to the normal distribution**
  - **Improving the fit of the model to the data**

  An effective transformation will simultaneously accomplish more than one of these objectives.

Selecting a Transformation (3)

- There are two methods for selecting the **form** of the transformation
  - Empirical selection
  - Box-Cox method

Empirical Selection of a Transformation (1), text 82-83

- Let $E(y)=\mu$, and $\sigma_y \propto \mu^{\alpha}$
- Suppose that the transformation is a power of the original data, say $y^*=y^{\lambda}$
- It can be shown that $\sigma_{y^*} \propto \mu^{\lambda+\alpha-1}$

Clearly, if we set $\lambda=1-\alpha$, the variance of y* is constant

Relationship Between σ_y and μ	α	$\lambda=1-\alpha$	Transformation	Comment
$\sigma_y \propto$ constant	0	1	No transformation	
$\sigma_y \propto \mu^{1/2}$	½	½	Square root	Possion (count) data
$\sigma_y \propto \mu$	1	0	Log	
$\sigma_y \propto \mu^{3/2}$	3/2	-1/2	Reciprocal square root	
$\sigma_y \propto \mu^{2}$	2	-1	Reciprocal	

Transformation often have little effect unless the ratio y_{max}/ y_{min} is larger than 2 or 3

Empirical Selection of a Transformation (2)

- When there is replication, we can empirically estimate α.

$\rightarrow \sigma_{y_i} \propto \mu_i^{\alpha}$, i.e. $\sigma_{y_i} = \theta \mu_i^{\alpha}$, were θ is a constant

$\rightarrow$ Taking logs, we obtain

$$\log \sigma_{y_i} = \log \theta + \alpha \log \mu_i$$

$\rightarrow$ Therefore, a plot of $\log \sigma_{y_i}$ versus $\log \mu_i$

would be a straight line with slope α.

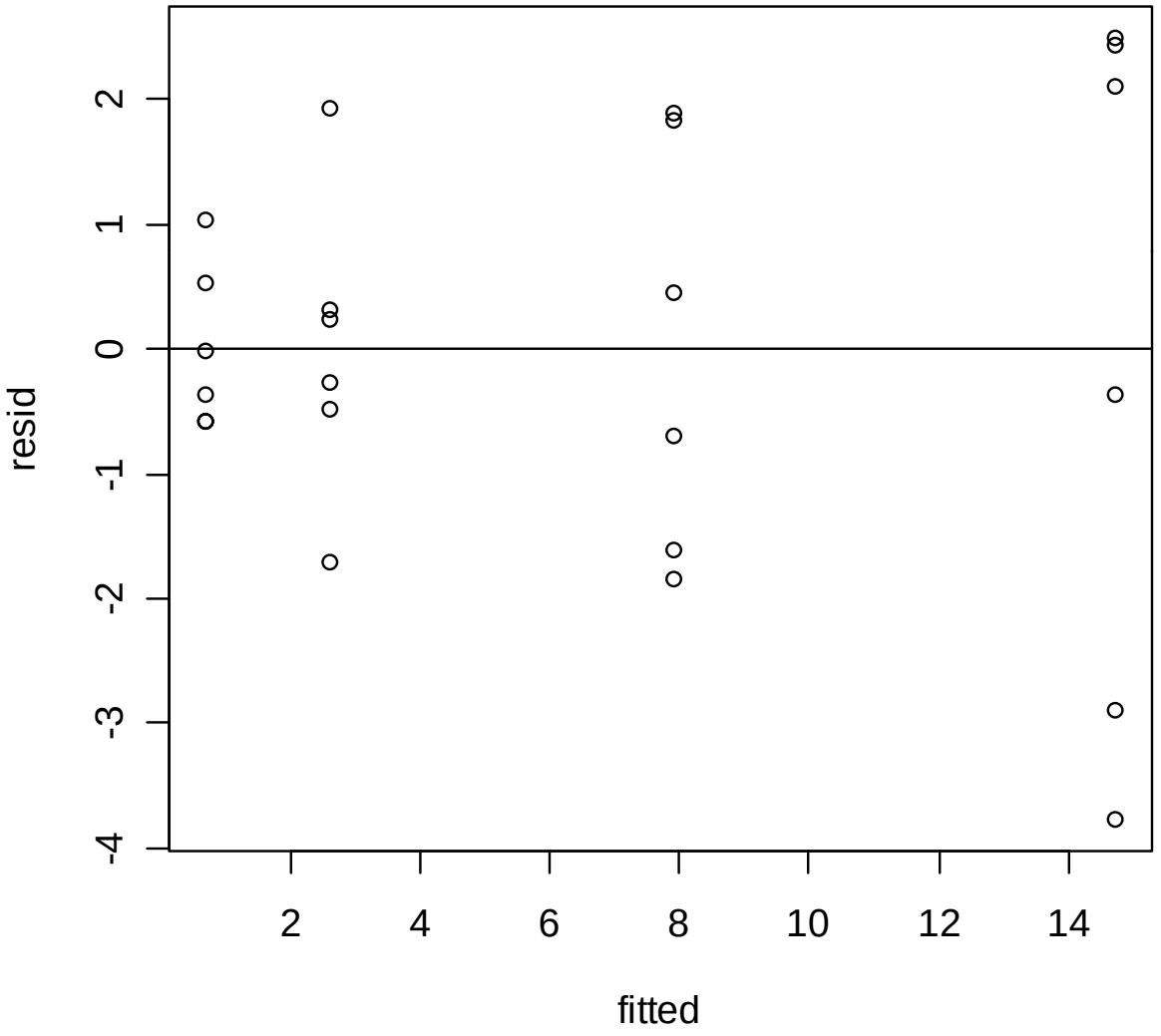
$\rightarrow$ Typically, the standard deviation S_i and the average $\bar{y}_{i.}$

of the ith treatment are used to estimate σ_{y_i} and μ_i .

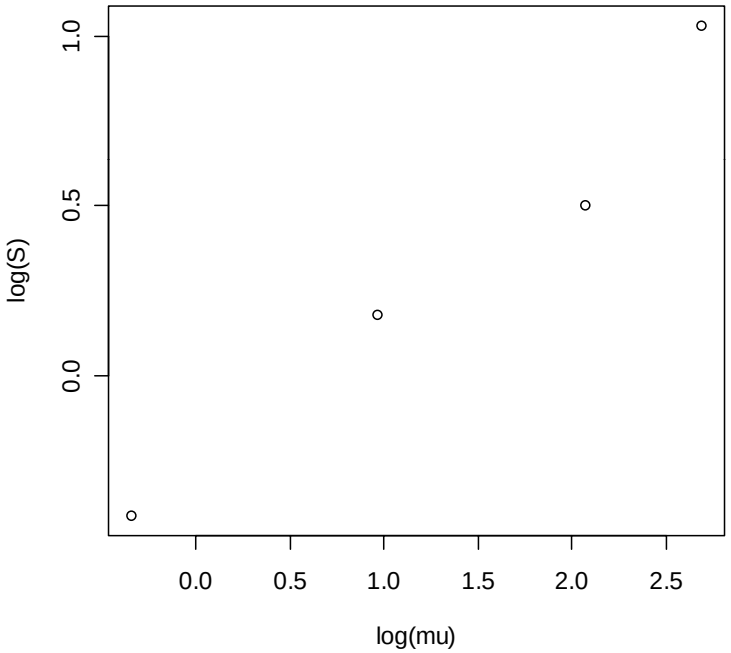
Example 3-5

- A civil engineer is interested in determine whether four different methods of estimating flood flow frequency produce equivalent estimates of peak discharge when applied to the same watershed. Each procedure is used six times on the watershed.
- The plot of residuals versus fitted values has outward-opening funnel shape indicating the constant variance assumption is not satified
- The modified Levene's test reject the null hypothesis of equal variances/

Example 3-5



Example 3-5



Call:
lm(formula = log(S) ~ log(mu))

Coefficients:

(Intercept)	log(mu)
-0.2781	0.4465

Square Root Transformation is desired

R-code

```
##### One Way ANOVA # Read in Discharge Experiment data
>discharge=read.table("Discharge.txt", header=T)
>attach(discharge)

>#obtain mean and standard deviation for each method
>S<-tapply(Peak,Method,sd)
>mu<-tapply(Peak,Method,mean)

#obtain the slope value
>plot(log(mu),log(S))
>lm(log(S)~log(mu))
```

Box-Cox Method (text 560-562)

(1)Transform data based on $y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda}-1}{\lambda \dot{y}^{\lambda-1}} & \lambda \neq 0 \\ \dot{y} \ln y & \lambda = 0 \end{cases}$

where $\dot{y} = \ln^{-1}[(1/n)\sum \ln y]$

(2) A standard analysis of variance on $y^{(\lambda)}$

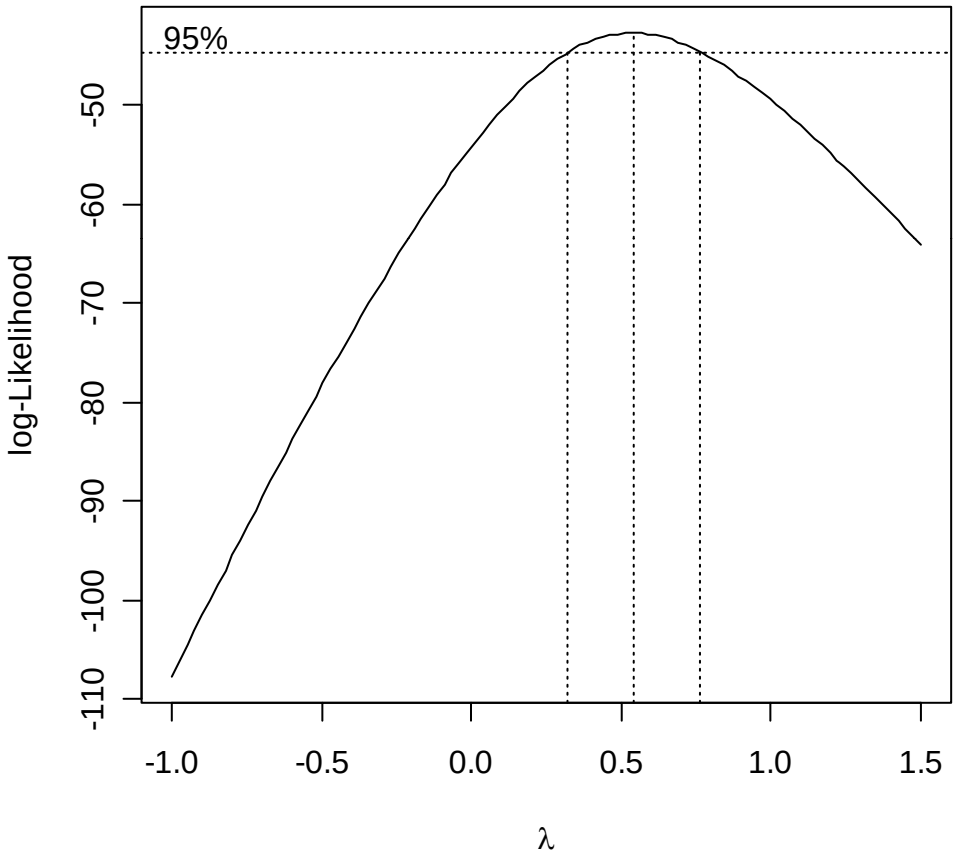
(3) obtain $SS_E(\lambda)$

(4) repeat (1) - (3) for 10 to 20 values of λ.

(5) plot λ versus $SS_E(\lambda)$.

(6) locate λ where $SS_E(\lambda)$ is minimum.

Example 3-5



R code

- #boxcox transformation

```
>library(MASS)

>boxcox(discharge.aov,lambda=seq(-1,1.5,1/10))
```