Lecture notes on Experiment Design & Data Analysis

Design of Engineering Experiments
Part 3 – Analysis of Variance (7)

- Parametric methods in the analysis of variance
- Nonparametric methods in the analysis of variance

How to make up for unequal variance and/or nonnormality

- When normality holds, but variances are not constant: weighted least squares.
- When nonnormality and nonconstant variance: use transformations of Y to make its distribution more nearly normal and to stabilize the variance.
- When departures are too extreme such that transformations do not work: use nonparametric tests in stead of the F test for equal means.

Nonparametric Kruskal-Wallis Test (1)

- When transformations are not successful, a nonparametric procedure can be used. Nonparametric procedures do not depend on the distribution of error terms. Here we assume the *a* distributions form a location family, that is, the p.d.f for the distribution of error terms in the *ith* treatment group is

  $$f_i(.)=f(.-\mu_i),\ i=1,\ldots,a.$$

  Where *f* is a p.d.f. with mean zero. Here the location family means that the shapes of the *a* distributions are the same, only their locations μ_I could be different.

Nonparametric Kruskal-Wallis Test (2)

The nonparametric Kruskal-Wallis Test is as following

1. Get ranks R_{ij} for observations Y_{ij}.: R_{ij} = rank of Y_{ij} among the pool of all observations {i=1,…,a;j=1,…,n). For example, for the data sets: 3,4,2,5,6,4, the corresponding ranks are 2,3.5,1,5,6,3.5.
2. Calculate H statistic as

$$H = \frac{1}{S^2}\left[\sum_{i=1}^{a}\frac{R_{i.}^2}{n_i} - \frac{N(N+1)^2}{4}\right] \overset{\text{under } H_0}{\sim} \text{approximated by } \chi^2(a-1)$$

$$\text{where, } S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{a}\sum_{j=1}^{n_i} R_{ij}^2 - \frac{N(N+1)^2}{4}\right]$$

Example 3-11

Table 3-14 Data and Ranks for the Plasma Etching Experiment in Example 3-1

Power							
160		**180**		**200**		**220**	
y_{1j}	R_{1j}	y_{2j}	R_{2j}	y_{3j}	R_{3j}	y_{4j}	R_{4j}
575	6	565	4	600	10	725	20
542	3	593	9	651	15	700	17
530	1	590	8	610	11.5	715	19
539	2	579	7	637	14	685	16
570	5	610	11.5	629	13	710	18
$R_{i.}$	17		39.5		63.5		90

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{a}\sum_{j=1}^{n_i} R_{ij}^2 - \frac{N(N+1)^2}{4}\right] = \frac{1}{20-1}\left[2869.50 - \frac{20(21)^2}{4}\right] = 34.97$$

$$H = \frac{1}{S^2}\left[\sum_{i=1}^{a}\frac{R_{i.}^2}{n_i} - \frac{N(N+1)^2}{4}\right] = \frac{1}{34.97}\left[2796.30 - 2205\right] = 16.91$$

$$H > \chi^2_{0.01,3} = 11.34 \text{ or P-value} = 0.000738$$

Reject H_0, at 0.01 significance level.

R code and Output

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

> #Specify Power is a categorical factor with 4 levels
> Power<-factor(Power)

> #Kruskal-Wallis test
> kruskal.test(EtchRate~Power)

        Kruskal-Wallis rank sum test

data:  EtchRate by Power
Kruskal-Wallis chi-squared = 16.907, df = 3, p-value = 0.0007386
```