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

- 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

Post-ANOVA Comparison of Means

- The analysis of variance tests the hypothesis of equal treatment means
- Assume that residual analysis is satisfactory
- If that hypothesis is rejected, we don't know **which specific means** are different apart from the fact that the smallest and largest means are different, of course!
- Determining which specific means differ following an ANOVA is called the **multiple comparisons problem**
- There are **lots** of ways to do this…see text, Section 3-5, pg. 87

Fisher's Least Significance Difference (LSD) Method

- Comparing all pairs of means
- The procedure uses the t statistic for testing $H_0: \mu_i = \mu_j$

Fisher's Least Significance Difference (LSD) Framework

- Null hypothesis H_0: $\mu_i = \mu_j$ H_a: $\mu_i \neq \mu_j$
- Test Statistics: $$t_0 = \frac{\bar{y}_{i.} - \bar{y}_{j.}}{\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \overset{\text{under } H_0}{\sim} t(N-a)$$

$$N = \sum_{i=1}^{a} n_i$$

- Reject Region: $|\bar{y}_{i.} - \bar{y}_{j.}| > t_{\alpha/2, N-a}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$ (LSD)
  or p-value<α
- Conclusion:

Example 3-8

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Power	3	66871	22290	66.797	2.883e-09	***
Residuals	16	5339	334			

$$LSD = t_{\alpha/2,N-a}\sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} = t_{0.025,20-4}\sqrt{334\left(\frac{1}{5}+\frac{1}{5}\right)} = 24.49$$

Example 3-8

Table 3-1 Etch Rate Data (in Å/min) from the Plasma Etching Experiment

Power (W)	Observations 1	2	3	4	5	Totals	Averages
160	575	542	530	539	570	2756	551.2
180	565	593	590	579	610	2937	587.4
200	600	651	610	637	629	3127	625.4
220	725	700	715	685	710	3535	707.0

$$\bar{y}_{1.} - \bar{y}_{2.} = 551.2 - 587.4 = -36.2*$$

$$\bar{y}_{1.} - \bar{y}_{3.} = 551.2 - 625.4 = -74.2*$$

$$\bar{y}_{1.} - \bar{y}_{4.} = 551.2 - 707.0 = -155.8*$$

$$\bar{y}_{2.} - \bar{y}_{3.} = 587.4 - 625.4 = -38.0*$$

$$\bar{y}_{2.} - \bar{y}_{4.} = 587.4 - 707.0 = -119.6*$$

$$\bar{y}_{3.} - \bar{y}_{4.} = 625.4 - 707.0 = -81.6*$$

Per Comparison error rate vs. Familywise error rate

- In this example, when we set α=0.05 for each comparison, the overall probability of making at least 1 Type I error is equal to approximately 6*0.05=0.3 since the probability of occurrence of one or more events can never exceed the sum of their individual probabilities.
- In this example, the 0.05 alpha for each comparison is called the **per-comparison error rate α_{PC};** and the overall alpha of 0.30 is called the **family-wise error rate (α_{FW}) ----**the probability of making *at least one Type I error* while carrying out a family of comparisons.

Fisher's Least Significance Difference (LSD) Method

- The family-wise error rate may be considerably inflated using LSD technique, as the number of treatments *a* gets larger.
- LSD technique is **NOT RECOMMENDED** *except* when you have three treatment levels.
  - Case 1: the complete null hypothesis is true: $\mu_1 = \mu_2 = \mu_3$. In this case, the probability that you will commit a Type I error with your overall F-test is α and any subsequent Type I errors that might occur will not affect the family-wise error rate. In fact, this is true for any number of means *when the complete null hypothesis is true.*
  - Case 2: If the complete null hypothesis is not true, but a more limited null hypothesis is true, then it must be the case that two of the means are equal and different from the third. For example, it may be that $\mu_1 = \mu_2 < \mu_3$. In this case, it is not possible to make a Type I error when carrying out the omnibus F-test. In this case, there will be only one pairwise comparison for which the null hypothesis is true, and therefore only one opportunity to make a Type I error. And so $\alpha_{FW} = \alpha_{PC}$

The studentized range statistic, q

- It is well known that as sample size (n) increases, so does the magnitude of the sample range.
- Imaging drawing random samples of various size from a normally distributed population with $\sigma=10$. The effect of sample size on *expected value* of the range is clear in the following table

Sample size	Expected Value of Range
2	11
5	23
10	31
20	37
50	45
100	50
200	55
500	61
1000	65

1. The effect of increased sample size on the expected value of the range is most pronounced when the sample size are relatively small.
2. An increase from n=2 to n=5 results in more than double of the expected value of the range (11 to 23)
3. But an increase from n=500 to n=1000 results in a very modest increase in the expected value of the range (from 61 to 65)

The studentized range statistic, q

- When you do a t-test, you are really comparing two scores that are drawn from a normal distribution: the two scores you compare are two sample means, and the normal distribution is the sampling distribution of the mean.
- Note that if you really have 5 samples (and 5 means), then the expected difference between the largest and smallest means (i.e., the range for the set of 5 means) will be much larger than if there are only 2 samples (and 2 means).
- Studentized range statistic

$$q = \frac{\bar{y}_{\max} - \bar{y}_{\min}}{\sqrt{MSE / n}}$$

Tukey's Honestly Significant Difference (HSD) method

- Null hypothesis H_0: $\mu_i = \mu_j$ H_a: $\mu_i \neq \mu_j$
- Test Statistics: $$q_0 = \frac{\bar{y}_{\max} - \bar{y}_{\min}}{\sqrt{MSE/n}} \overset{H_0:\mu_1=\mu_2=\cdots\mu_a}{\sim} q(a, N-a)$$

$$\geq \frac{\bar{y}_{i.} - \bar{y}_{j.}}{\sqrt{MSE/n}}$$

$$N = \sum_{i=1}^{a} n_i$$

- Reject Region: $|\bar{y}_{i.} - \bar{y}_{j.}| > q_\alpha(a, N-a)\sqrt{MSE}$ (HSD)

  or p-value $< \alpha$
- Conclusion:

Tukey's Honestly Significant Difference (HSD) method

- 100(1-α) percent confidence intervals for all pairs of means

$$\bar{y}_{i.} - \bar{y}_{j.} - q_{\alpha}(a, N-a)\sqrt{MSE/n} \leq \mu_i - \mu_j$$
$$\leq \bar{y}_{i.} - \bar{y}_{j.} - q_{\alpha}(a, N-a)\sqrt{MSE/n}$$

- Tukey procedure controls the experimentwise or "familywise" error rate at the selected level α when the sample sizes are equal and at most α when the sample sizes are not equal

Example 3-7

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Power	3	66871	22290	66.797	2.883e-09	***
Residuals	16	5339	334			

$$HSD = q_{\alpha}(a, N-a)\sqrt{MSE} = q_{0.05}(4,16)\sqrt{334/5}$$

$$= 4.05 * \sqrt{334/5} = 33.09$$

Example 3-8

Table 3-1 Etch Rate Data (in Å/min) from the Plasma Etching Experiment

Power (W)	Observations 1	2	3	4	5	Totals	Averages
160	575	542	530	539	570	2756	551.2
180	565	593	590	579	610	2937	587.4
200	600	651	610	637	629	3127	625.4
220	725	700	715	685	710	3535	707.0

$$\bar{y}_{1.} - \bar{y}_{2.} = 551.2 - 587.4 = -36.2*$$

$$\bar{y}_{1.} - \bar{y}_{3.} = 551.2 - 625.4 = -74.2*$$

$$\bar{y}_{1.} - \bar{y}_{4.} = 551.2 - 707.0 = -155.8*$$

$$\bar{y}_{2.} - \bar{y}_{3.} = 587.4 - 625.4 = -38.0*$$

$$\bar{y}_{2.} - \bar{y}_{4.} = 587.4 - 707.0 = -119.6*$$

$$\bar{y}_{3.} - \bar{y}_{4.} = 625.4 - 707.0 = -81.6*$$

R code for Tukey’s HSD

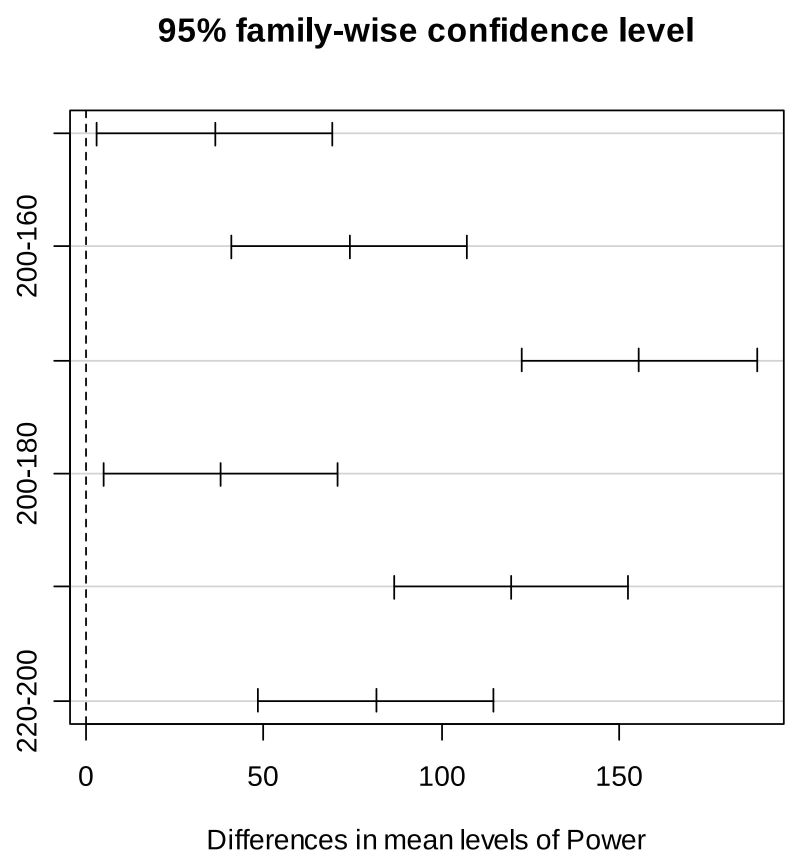
- # Tukey's multiple comparison procedure
- > plasma.mcp=TukeyHSD(plasma.aov)
- > plot(plasma.mcp)
- > plasma.mcp

- Tukey multiple comparisons of means
- 95% family-wise confidence level

- Fit: aov(formula = EtchRate ~ Power)

- $Power

–	diff	lwr	upr	p adj
180-160	36.2	3.145624	69.25438	0.0294279
200-160	74.2	41.145624	107.25438	0.0000455
220-160	155.8	122.745624	188.85438	0.0000000
200-180	38.0	4.945624	71.05438	0.0215995
220-180	119.6	86.545624	152.65438	0.0000001
220-200	81.6	48.545624	114.65438	0.0000146


95% family-wise confidence level
200-160
200-180
220-200
0
50
100
150
Differences in mean levels of Power

Comparing Treatment Means with a Control (Dunnett's test)

- Multiple comparison procedures can be compared to buying insurance. Here, the insurance is against making a claim of a statistically significant result when it is just the result of chance variation. Tukey's HSD is the right amount of insurance when all possible pairwise comparisons are being made in a set of *a* groups. However, sometimes not all comparisons will be made and Tukey's HSD buys too much insurance.
- In many experiments, one of the treatments is a control, and the analyst is interested in comparing each of the other *a-1* treatment means with the control. Thus only *a-1* comparisons are to be made. A procedure for making these comparisons has been developed by Dunnett(1964).

Dunnett's test

- Null hypothesis $H_0: \mu_i = \mu_a$ $H_a: \mu_i \neq \mu_a$
- Test Statistics:

$$d_0 = \frac{\bar{y}_{i.} - \bar{y}_{a.}}{\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_a}\right)}} \overset{H_0:\mu_i=\mu_a}{\sim} d(a-1, N-a)$$

$$N = \sum_{i=1}^{a} n_i$$

- Reject Region: $|\bar{y}_{i.} - \bar{y}_{a.}| > d_\alpha(a-1, N-a)\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_a}\right)}$
  or p-value $<\alpha$
- Conclusion:

Example 3-9

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Power	3	66871	22290	66.797	2.883e-09	***
Residuals	16	5339	334			

$$Dunnett\ d_{\alpha}(a-1, N-a)\sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} = d_{0.05}(3,16)\sqrt{334\left(\frac{1}{5}+\frac{1}{5}\right)}$$

$$= 2.59\sqrt{334\left(\frac{1}{5}+\frac{1}{5}\right)} = 29.92$$

Example 3-8

Table 3-1 Etch Rate Data (in Å/min) from the Plasma Etching Experiment

Power (W)	Observations 1	2	3	4	5	Totals	Averages
160	575	542	530	539	570	2756	551.2
180	565	593	590	579	610	2937	587.4
200	600	651	610	637	629	3127	625.4
220	725	700	715	685	710	3535	707.0

$$\bar{y}_{1.} - \bar{y}_{4.} = 551.2 - 707.0 = -155.8*$$

$$\bar{y}_{2.} - \bar{y}_{4.} = 587.4 - 707.0 = -119.6*$$

$$\bar{y}_{3.} - \bar{y}_{4.} = 625.4 - 707.0 = -81.6*$$

R code for Dunnett’s method

- #In R, go to
- #Packages -> Install package(s) -> find "usa CA1"--> highlight it and click ok-->
- #find "multcomp" in the pop-up window and highlight-> click ok
- >library(multcomp)

- ##### 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)

- #Dunnett's multiple comparison procedure
- >C<-matrix(c(1,0,0,0,1,0,0,0,1,-1,-1,-1),nc=4)
- >simtest(EtchRate~Power,cmatrix=C)

R output for Dunnett's method

```
Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: User-defined Contrasts


Fit: lm(formula = EtchRate ~ Power)

Linear Hypotheses:
       Estimate Std. Error t value p value
1 == 0  -155.80      11.55 -13.485  <1e-05 ***
2 == 0  -119.60      11.55 -10.352  <1e-05 ***
3 == 0   -81.60      11.55  -7.063  <1e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported)
```

Alternative R code for Dunnett's method

```
#Dunnett's multiple comparison procedure
>C<-matrix(c(1,0,0,0,1,0,0,0,1,-1,-1,-1),nc=4)

>summary(glht(plasma.aov, linfct = mcp(Power = C)))
```