Analysis of Factor Level Means

- **Tukey Multiple Comparison Procedure:**

  The Tukey mulpitle comparison confidence limits for all pairwise comparisons $D = \mu_i - \mu_{i'}$ with family confidence coefficient of at least $1 - \alpha$ are as follows:

$$\hat{D} \pm Ts\{\hat{D}\}$$

  Frequently two treatments or factor levels are to be compared by estimating the difference D between two factor level means, say, μ_i and $\mu_{i'}$:

$$D = \mu_i - \mu_{i'}$$

  1. $\hat{D} = \bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}$

  2. $\sigma^2\{\hat{D}\} = \sigma^2\{\bar{Y}_{i\cdot}\} + \sigma^2\{\bar{Y}_{i'\cdot}\} = \sigma^2(\frac{1}{n_i} + \frac{1}{n_{i'}})$

  3. $s^2\{\hat{D}\} = MSE(\frac{1}{n_i} + \frac{1}{n_{i'}})$

  4. $T = \frac{1}{\sqrt{2}} q(1 - \alpha; r, n_T - r)$

  Example. In a study of the effectiveness of different rust inhibitors, four brands (A, B, C, D) were tested. Altogether, 40 experimental units were randomly assigned to the four brands, with 10 units assigned to each brand. The data for this experiment are stored in the file 'RustInhibitors.csv' and can be found in our course website. The higher the 'score' value, the more effective is the rust inhibitor. Determine if there is a difference between the four brands.

```
data=read.csv("RustInhibitors.csv",header=T)
attach(data)
data[1:4,]

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

plot(result$fit,result$res)
qqnorm(result$res)
qqline(result$res)
shapiro.test(result$res)

par(mfrow=c(2,2))
plot(result)

means=tapply(score,brand,mean)
#     1     2     3     4
# 43.14 89.44 67.95 40.47
ns=tapply(score,brand,length)
#  1  2  3  4
# 10 10 10 10

# 95% C.I. for mean_1-mean_2 (Rust Inhibitors Example).
lwr.21=(means[2]-means[1])-(1/sqrt(2))*qtukey(.95,4,46)*sqrt(MSE*(1/ns[1]+1/ns[2]))
upr.21=(means[2]-means[1])+(1/sqrt(2))*qtukey(.95,4,46)*sqrt(MSE*(1/ns[1]+1/ns[2]))
c(lwr.21,upr.21)
# 43.34627 49.25373

# Tukey Procedure (Good for all pair-wise comparisons)
brand=factor(brand)
result=aov(score~brand)
TukeyHSD(result,"brand",level=.95)
```

- **Scheffe Multiple Comparison Procedure:**

The Scheffe multiple comparison applies for analysis of variance models when the family of interest is the set of all possible contrasts among the factor level means:

$$L = \sum_{i=1}^{r} c_i \mu_i \qquad \text{where } \sum_{i=1}^{r} c_i = 0$$

The Scheffe confidence intervals for the family of contrasts L are of the form:

$$\hat{L} \pm S s\{\hat{L}\}$$

where

1. $s^2\{\hat{L}\} = MSE \sum_{i=1}^{r} \frac{c_i^2}{n_i}$

2. $S^2 = (r-1)F(1-\alpha; r-1, n_T - r)$

Example. In the Kenton Food example, interest centered on estimating the following four contrasts with family confidence coefficient of .90:

1. Comparison of the mean sales for the two 3-color designs:

2. Comparison of the mean sales for the 3-color and 5-color designs:

3. Comparison of the mean sales for designs with and without cartoons:

4. Comparison of the mean sales for design 1 with average sales for all 4 designs:

```
data2=read.csv("KentonFood.csv",header=T)
attach(data2)
data2[1:4,]

result=aov(sales~factor(design))
anova(result)
#                Df Sum Sq Mean Sq F value    Pr(>F)
# factor(design)  3 588.22 196.074  18.591 2.585e-05 ***
# Residuals      15 158.20  10.547

MSE=anova(result)$Mean[2]       # MSE=10.54667
means=tapply(sales,design,mean)
#    1    2    3    4
# 14.6 13.4 19.5 27.2
ns=tapply(sales,design,length)
# 1 2 3 4
# 5 5 4 5

# 95% C.I. for contrast 2 (Kenton Food Example).
c2=c(1/2,1/2,-1/2,-1/2)
L2=sum(c2*means)
r=length(means)
nt=length(sales)
S2=(r-1)*qf(.90,r-1,nt-r)
lwr.L2=L2-sqrt(S2*MSE*sum(c2^2/ns))     # -49.25373
upr.L2=L2+sqrt(S2*MSE*sum(c2^2/ns))     # -43.34627
```

- **Bonferroni Multiple Comparison Procedure:**

  The Bonferroni multiple comparison applies for analysis of variance models when the family of interest is a particular set of pairwise comparisons, contrasts, or linear combinations of factor level means that is specified by the user in advance of the data analysis:

  The Bonferroni confidence intervals for the family of g linear combinations L are of the form:

$$\hat{L} \pm Bs\{\hat{L}\}$$

  where

  1. $s^2\{\hat{L}\} = MSE \sum_{i=1}^{r} \frac{c_i^2}{n_i}$

  2. $B = t(1 - \alpha/2g; n_T - r)$

```
data2=read.csv("KentonFood.csv",header=T)
attach(data2)

result2=aov(sales~factor(design))
anova(result2)
MSE=anova(result2)$Mean[2]

means=tapply(sales,design,mean)
ns=tapply(sales,design,length)

# 97.5% C.I. for contrasts L_1 and L_2 (Kenton Food Example).
c1=c(1/2,1/2,-1/2,-1/2); c2=c(1/2,-1/2,+1/2,-1/2)

L1=sum(means*c1); L2=sum(means*c2)
r=length(means); nt=length(sales); g=2

B=qt(1-0.025/(2*g),nt-r)
lwr.1=L1-B*sqrt(MSE*sum(c1^2/ns))
upr.1=L1+B*sqrt(MSE*sum(c1^2/ns))
lwr.2=L2-B*sqrt(MSE*sum(c2^2/ns))
upr.2=L2+B*sqrt(MSE*sum(c2^2/ns))

lwr=c(lwr.1,lwr.2); upr=c(upr.1,upr.2)
lwr=round(lwr,digit=3); upr=round(upr,digit=3)  # Rounding to 3 decimal places
Contrast=c("L1","L2")
data.frame(Contrast,lwr, upr)
  Contrast     lwr    upr
1       L1 -13.597 -5.103
2       L2  -7.497  0.997
```

- **Homework.** Do 17.10, 17.13, and 17.15, on pages 768-770.