ORIGIN ≡ 0

W. Stein

Interval Estimation in Unbalanced 2-Way ANOVA

Upon fit of a Linear Model to an unbalanced ANOVA design, GLM tests provide overall estimates of Interaction, and Factor A&B Main Effects. This may be followed by multiple simultaneous interval estimations of contrasts involving sample (block) means. Example here continues 2008 Linear Models 19 from Chapter 23 in Kuter et al. (KNNL) A*pplied Linear Statistical Models* 5th Edition.

Cell Means ANOVA Model:

K := READPRN("c:/2008LinearModelsData/GrowthHormoneCM.txt")

< Here the independent variables were explicitly coded into p=6 Indicator variables X1-X6 with each column representing a cell ij.

Variable Assignment:

$Y := K^{\langle 0\rangle}$

$j := 1 .. 6$

$X^{\langle j-1\rangle} := K^{\langle j\rangle}$ **< design matrix**

$N := \text{length}(Y) \quad N = 14$ **< total number of cases**

$r := \text{cols}(X) \quad r = 6$

$p := r$ **< p used previously, r = p to conform with KNNL**

$i := 0 .. N-1$

$ii := 0 .. N-1$

$I := \text{identity}(N) \quad J_{i,ii} := 1$ **< Identity & One Matrix for matrix calculations**

$a := 2 \quad b := 3$ **< levels of factor A & B**

$$K = \begin{pmatrix} 1.4 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2.4 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2.2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2.1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1.7 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1.1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 2.4 & 0 & 0 & 0 & 1 & 0 & 0 \\ 2.5 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1.8 & 0 & 0 & 0 & 0 & 1 & 0 \\ 2 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0.5 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0.9 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1.3 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \qquad X = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Least Squares Estimation of the Regression Parameters:

$$\beta := \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot Y$$

$$\beta = \begin{pmatrix} 2 \\ 1.9 \\ 0.9 \\ 2.4 \\ 2.1 \\ 0.9 \end{pmatrix} \qquad \begin{pmatrix} \beta1(\mu11) \\ \beta2(\mu12) \\ \beta3(\mu13) \\ \beta4(\mu21) \\ \beta5(\mu22) \\ \beta6(\mu23) \end{pmatrix}$$

< KNNL Cell Means Model

Fitted Values & Hat Matrix H:

$Y_h := X \cdot \beta$ **< fitted values Yh**

$H := X \cdot \left(X^T \cdot X\right)^{-1} \cdot X^T$ **< nXn Hat matrix**

$$n := \begin{pmatrix} 3 \\ 2 \\ 2 \\ 1 \\ 3 \\ 3 \end{pmatrix}$$

< counts in each block

$i := 0 .. \text{length}(n) - 1$

Residuals:

$e := Y - Y_h$ **< residuals**

Full Model ANOVA Table for Cell Means Model:

Sum of Squares:		Degrees of Freedom:		Mean Squares:	
$SSTR := Y^T \cdot \left[H - \left(\frac{1}{N}\right) \cdot J\right] \cdot Y$	$SSTR = (4.4743)$	$df_R := p - 1$	$df_R = 5$	$MSTR := \frac{SSTR}{df_R}$	$MSTR = (0.8949)$
$SSE := Y^T \cdot (I - H) \cdot Y$	$SSE = (1.3)$	$df_E := N - p$	$df_E = 8$	$MSE := \frac{SSE}{df_E}$	$MSE = (0.1625)$
$SSTO := Y^T \cdot \left[I - \left(\frac{1}{N}\right) \cdot J\right] \cdot Y$	$SSTO = (5.7743)$	$df_T := N - 1$	$df_T = 13$	$MSTO := \frac{SSTO}{df_T}$	$MSTO = (0.4442)$

Prototype in R:

```
#INTERVAL ESTIMATION IN 2-WAY ANOVA LINEAR MODEL
require(car)      #MUST LOAD {gmodels} PACKAGE FROM CRAN
require(gmodels)  #MUST LOAD {gmodels} PACKAGE FROM CRAN

#READ STRUCTURED DATA TABLE WITH NUMERIC CODED FACTOR
K=read.table("c:/2008LinearModelsData/GrowthHormoneR.txt")
K
attach(K)
Y=Rate
A=factor(GenderA)
B=factor(BoneDevB)
detach(K)

#CONVERTING TO TREATMENT EFFECTS = contr.sum MODEL IN R
#FULL MODEL FROM contr.sum
contrasts(A)=contr.sum
contrasts(B)=contr.sum
FM=lm(Y~A*B)

MM=data.frame(model.matrix(FM))
MM

summary(FM)
Anova(FM,type="III")  #TYPE 3 ANOVA SS

#CROSS TABULATIONS FOR CELL MEANS AND COUNTS IN EACH BLOCK
XX=xtabs(Y~A+B)
XX    #SUM OF Y VALUES IN EACH BLOCK
n=table(A,B)
n    #COUNTS FOR EACH BLOCK
N=length(Y)
N    #TOTAL NUMBER OF OBJECTS
mu=XX/n
mu    #MEANS FOR EACH BLOCK
N=length(Y)
```

```
> summary(FM)
Call:
lm(formula = Y ~ A * B)
Residuals:
       Min         1Q     Median         3Q        Max
-6.000e-01 -2.000e-01 -8.816e-17  2.000e-01  4.000e-01
Coefficients:
              Estimate Std. Error   t value Pr(>|t|)
(Intercept)  1.700e+00  1.164e-01    14.609 4.73e-07 ***
A1          -1.000e-01  1.164e-01    -0.859   0.4152
B1           5.000e-01  1.778e-01     2.813   0.0227 *
B2           3.000e-01  1.576e-01     1.904   0.0934 .
A1:B1       -1.000e-01  1.778e-01    -0.563   0.5891
A1:B2       -1.768e-17  1.576e-01 -1.12e-16   1.0000
```

```
> K
   Rate GenderA BoneDevB Replicate
1   1.4       1        1         1
2   2.4       1        1         2
3   2.2       1        1         3
4   2.1       1        2         1
5   1.7       1        2         2
6   0.7       1        3         1
7   1.1       1        3         2
8   2.4       2        1         1
9   2.5       2        2         1
10  1.8       2        2         2
11  2.0       2        2         3
12  0.5       2        3         1
13  0.9       2        3         2
14  1.3       2        3         3
```

```
> MM
   X.Intercept. A1 B1 B2 A1.B1 A1.B2
1             1  1  1  0     1     0
2             1  1  1  0     1     0
3             1  1  1  0     1     0
4             1  1  0  1     0     1
5             1  1  0  1     0     1
6             1  1 -1 -1    -1    -1
7             1  1 -1 -1    -1    -1
8             1 -1  1  0    -1     0
9             1 -1  0  1     0    -1
10            1 -1  0  1     0    -1
11            1 -1  0  1     0    -1
12            1 -1 -1 -1     1     1
13            1 -1 -1 -1     1     1
14            1 -1 -1 -1     1     1
```

```
> mu    #MEANS FOR EACH BLOCK
   B
A     1   2   3
  1 2.0 1.9 0.9
  2 2.4 2.1 0.9
> N=length(Y)
```

```
> Anova(FM,type="III")  #TYPE 3 ANOVA SS
Anova Table (Type III tests)
Response: Y
            Sum Sq Df  F value    Pr(>F)
(Intercept) 34.680  1 213.4154 4.729e-07 ***
A            0.120  1   0.7385  0.415160
B            4.190  2  12.8914  0.003145 **
A:B          0.075  2   0.2321  0.798034
Residuals    1.300  8
```

Linear Contrast of Means:

$$\beta = \begin{pmatrix} 2 \\ 1.9 \\ 0.9 \\ 2.4 \\ 2.1 \\ 0.9 \end{pmatrix}$$ **< cell means**

$$c := \begin{pmatrix} 0.5 \\ 0 \\ -0.5 \\ 0.5 \\ 0 \\ -0.5 \end{pmatrix}$$ **< Contrast Vector = Coefficients of Linear Combination L**

$L := \left(\beta^T \cdot c\right)_0$ $\quad L = 1.3$ $\quad$ **< $L = c_1 Ybar_1 + c_2 Ybar_2 + c_3 Ybar_3 + ... + c_k Ybar_k$**

Single Test of Contrast Vector:

t-Test for Linear Contrast H_0: L = 0 versus H_1: L <> 0:

Assumption:

ε_{ij} are a random sample ~ $N(0,\sigma^2)$

Restriction:

$$\sum c = 0$$

Hypotheses:

H_0: L = 0 $\quad$ **< Linear Contrast is zero**

H_1: L <> 0

Test Statistic:

$i := 0 .. r - 1 \qquad r = 6$

$$t := \frac{L}{\sqrt{MSE \cdot \left[\sum_i \frac{\left(c_i\right)^2}{n_i}\right]}}$$

$t = (4.3818)$ $\quad$ **< Linear Contrast normalized by Standard Error & and Cell sizes**

Critical Value of the Test:

$\alpha := 0.10$ $\quad$ **< Probability of Type I error must be explicitly set**

$C := qt\left(\frac{\alpha}{2}, N - r\right)$ $\quad C = -1.8595$ $\quad$ **< Note degrees of freedom = (N-r)**

Decision Rule:

IF |t| > |C|, THEN REJECT H_0 OTHERWISE ACCEPT H_0

Probability Value:

$P := \min[2 \cdot pt(t, N - r), 2 \cdot (1 - pt(t, N - r))]$ $\quad P = 0.0023$

Confidence Intervals for Pairwise Single Contrasts:

$$s_L := \sqrt{MSE \cdot \left[\sum_i \frac{\left(c_i\right)^2}{n_i}\right]}$$

$s_{L_0} = 0.2966831$ $\quad$ **< Standard Error of L**

$CI := \left[\left(L - |C| \cdot s_L\right)_0 \quad L \quad \left(L + |C| \cdot s_L\right)_0\right]$

$CI = (0.7483036 \quad 1.3 \quad 1.8516964)$

lower estimate upper

Prototype in R:

```
#PAIRWISE CONTRASTS OF FACTOR LEVEL MEANS
#CONFIDENCE INTERVALS FOR SINGLE CONTRASTS
CS12=fit.contrast(FM,B,c(1,-1,0),conf.int=0.90)
CS13=fit.contrast(FM,B,c(1,0,-1),conf.int=0.90)
CS23=fit.contrast(FM,B,c(0,1,-1),conf.int=0.90)
CS12
CS13
CS23
```

```
> CS12
               Estimate Std. Error   t value  Pr(>|t|)   lower CI  upper CI
B c=( 1 -1 0 )      0.2  0.2966831 0.6741201 0.5192349 -0.3516964 0.7516964
> CS13
               Estimate Std. Error  t value    Pr(>|t|)  lower CI upper CI
B c=( 1 0 -1 )      1.3  0.2966831 4.381780 0.002343192 0.7483036 1.851696
> CS23
               Estimate Std. Error  t value   Pr(>|t|)  lower CI upper CI
B c=( 0 1 -1 )      1.1  0.2602082 4.227383 0.00288663 0.6161303 1.583870
```

^ values match for contrast 1-3 above. Note difference in how contrast is specified here versus by hand above.

NOTE: pairwise.t.test results in R do not match here!

```
#TUKEY CI'S
alpha=0.10
a=nlevels(A)
a
b=nlevels(B)
b
Q=qtukey((1-alpha),(b),(N-a*b)) # 3 CONTRASTS KNNL P961
T=(1/sqrt(2))*Q
D=CS12[1]
lwr=CS12[1]-CS12[2]*T
upr=CS12[1]+CS12[2]*T
CT12=cbind(lwr,D,upr)
D=CS13[1]
lwr=CS13[1]-CS13[2]*T
upr=CS13[1]+CS13[2]*T
CT13=cbind(lwr,D,upr)
D=CS23[1]
lwr=CS23[1]-CS23[2]*T
upr=CS23[1]+CS23[2]*T
CT23=cbind(lwr,D,upr)
CT12
CT13
CT23
```

```
> CT12
            lwr   D       upr
[1,] -0.5078135 0.2 0.9078135

> CT13
           lwr   D      upr
[1,] 0.5921865 1.3 2.007814

> CT23
           lwr   D      upr
[1,] 0.4792065 1.1 1.720794
```

^ verified KNNL p. 963

NOTE: TukeyHSD results in R do not match here!

```
#SCHEFFE CI'S
alpha=0.10
S=sqrt((b-1)*qf(1-alpha,(b-1),N-a*b)) #SCHEFFE MULTIPLIER
S
D=CS12[1]
lwr=CS12[1]-CS12[2]*S
upr=CS12[1]+CS12[2]*S
CF12=cbind(lwr,D,upr)
D=CS13[1]
lwr=CS13[1]-CS13[2]*S
upr=CS13[1]+CS13[2]*S
CF13=cbind(lwr,D,upr)
D=CS23[1]
lwr=CS23[1]-CS23[2]*S
upr=CS23[1]+CS23[2]*S
CF23=cbind(lwr,D,upr)
CF12
CF13
CF23
```

```
> CF12
            lwr   D       upr
[1,] -0.5402962 0.2 0.9402962
> CF13
           lwr   D      upr
[1,] 0.5597038 1.3 2.040296
> CF23
           lwr   D      upr
[1,] 0.4507173 1.1 1.749283
```

Single Degree of Freedom Test:

This test allows one to determine whether a single row or column mean level of one factor is greater or lesser than zero (one-way test) or not equal (two-sided).

t-Test for H_0: Row or Column mean for a Factor is <> or = 0:

Assumption:

ε_{ij} **are a random sample ~ $N(0,\sigma^2)$**

Hypotheses:

H_0: μ of one factor level <> (one-sided) or = 0 (two-sided)

H_1: not H_0

$$\beta = \begin{pmatrix} 2 \\ 1.9 \\ 0.9 \\ 2.4 \\ 2.1 \\ 0.9 \end{pmatrix} \qquad n = \begin{pmatrix} 3 \\ 2 \\ 2 \\ 1 \\ 3 \\ 3 \end{pmatrix}$$

$$\mu := \begin{pmatrix} \text{mean}(\beta_0, \beta_3) \\ \text{mean}(\beta_1, \beta_4) \\ \text{mean}(\beta_2, \beta_5) \end{pmatrix} \qquad \mu = \begin{pmatrix} 2.2 \\ 2 \\ 0.9 \end{pmatrix}$$

< means for each level of factor B

$MSE_0 = 0.1625$

^ from above

$$\text{Var} := \begin{bmatrix} \frac{MSE_0}{a^2} \cdot \left(\frac{1}{n_0} + \frac{1}{n_3}\right) \\ \frac{MSE_0}{a^2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_4}\right) \\ \frac{MSE_0}{a^2} \cdot \left(\frac{1}{n_2} + \frac{1}{n_5}\right) \end{bmatrix} \qquad \text{Var} = \begin{pmatrix} 0.0541667 \\ 0.0338542 \\ 0.0338542 \end{pmatrix}$$

$$SE := \sqrt{\begin{bmatrix} \frac{MSE_0}{a^2} \cdot \left(\frac{1}{n_0} + \frac{1}{n_3}\right) \\ \frac{MSE_0}{a^2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_4}\right) \\ \frac{MSE_0}{a^2} \cdot \left(\frac{1}{n_2} + \frac{1}{n_5}\right) \end{bmatrix}} \qquad SE = \begin{pmatrix} 0.2327373 \\ 0.183995 \\ 0.183995 \end{pmatrix}$$

Test Statistic:

$$t := \frac{\overrightarrow{\mu}}{SE} \qquad t = \begin{pmatrix} 9.4527163 \\ 10.8698595 \\ 4.8914368 \end{pmatrix}$$

Critical Value of the Test:

$\alpha := 0.05$ **< Probability of Type I error must be explicitly set**

$$CV := \begin{pmatrix} qt\left(1 - \frac{\alpha}{2}, N - a \cdot b\right) \\ qt\left(1 - \frac{\alpha}{2}, N - a \cdot b\right) \\ qt(1 - \alpha, N - a \cdot b) \end{pmatrix} \qquad CV = \begin{pmatrix} 2.306 \\ 2.306 \\ 1.8595 \end{pmatrix}$$

< two=sided
< two=sided
< one=sided

< Note degrees of freedom = (N-a·b)

Decision Rule:

IF |t| > |C|, THEN REJECT H_0 OTHERWISE ACCEPT H_0

Probability Value:

$$P := 1 - pt(t, N - a \cdot b) \quad P = \begin{pmatrix} 0.0000065 \\ 0.0000023 \\ 0.0006034 \end{pmatrix}$$

Prototype in R:

```
#SINGLE DEGREE OF FREEDOM TESTS
a
b
alpha=0.05 #TEST LEVEL
MSE=anova(FM)[4,3]
MSE  #FROM anova()
#VARIANCES
V1=(MSE/(a^2))*sum((1/n[1,1])+(1/n[2,1]))
V2=(MSE/(a^2))*sum((1/n[1,2])+(1/n[2,2]))
V3=(MSE/(a^2))*sum((1/n[1,3])+(1/n[2,3]))
VAR=c(V1,V2,V3)
#STANDARD ERROR
sd1=sqrt(V1)
sd2=sqrt(V2)
sd3=sqrt(V3)
SE=c(sd1,sd2,sd3)
#MEANS
mu_dot1=mean(mu[1,1],mu[2,1])
mu_dot2=mean(mu[2,1],mu[2,2])
mu_dot3=mean(mu[1,3],mu[2,3])
MEAN=c(mu_dot1,mu_dot2,mu_dot3)
# t STATISTICS
t1=mu_dot1/sd1
t2=mu_dot2/sd2
t3=mu_dot3/sd3
STAT=c(t1,t2,t3)
#CRITICAL VALUES
CV1=qt(1-alpha/2,N-a*b) #TWO-SIDED
CV2=qt(1-alpha/2,N-a*b) #TWO-SIDED
CV3=qt(1-alpha,N-a*b)   #ONE-SIDED
CV=c(CV1,CV2,CV3)
#PROBABILITIES
P1=1-pt(t1,N-a*b) #TWO-SIDED TEST
P2=1-pt(t2,N-a*b) #TWO-SIDED TEST
P3=1-pt(t3,N-a*b) #ONE-SIDED TEST
PROB=c(P1,P2,P3)
RESULT=cbind(VAR,SE,MEAN,STAT,CV,PROB)
RESULT
```

```
> RESULT
            VAR        SE MEAN      tSTAT       CV         PROB
[1,] 0.05416667 0.2327373  2.2   9.452716 2.306004 6.450224e-06
[2,] 0.03385417 0.1839950  2.0  10.869860 2.306004 2.269398e-06
[3,] 0.03385417 0.1839950  0.9   4.891437 1.859548 6.033818e-04
```

verified KNNL p. 964 for third factor level >