ORIGIN ≡ 0

W. Stein

Linear Combinations and the Contrast Vector/Matrix in ANOVA

Upon rejecting the null hypothesis for the Overall ANOVA F-test, one usually is interested in tests of some Class (Cell or Treatment) means against others. Use of Linear Combinations and Linear Contrasts allow one to formulate hypotheses at will. Example below comes from Chapter 16 in Kuter et al. (KNNL) A*pplied Linear Statistical Models* 5th Edition.

Example:

Cell Means ANOVA Model: **KNNL p.683**

< original variable called "Design" has been dummy coded in this dataset as a set of index variables with: p=4 variables for 4 factor levels.

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

Variable Assignment:

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

$X1 := K^{\langle 1 \rangle}$ $X2 := K^{\langle 2 \rangle}$ $X3 := K^{\langle 3 \rangle}$ $X4 := K^{\langle 4 \rangle}$

N := length(Y) N **< total number of cases**

X := augment(X1, X2, X3, X4) **< design matrix**

r := cols(X) r = 4

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

i := 0 .. N − 1

ii := 0 .. N − 1

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

$$n := \begin{pmatrix} 5 \\ 5 \\ 4 \\ 5 \end{pmatrix}$$

< Cell numbers counted by hand from K

$$K = \begin{pmatrix} 11 & 1 & 0 & 0 & 0 \\ 17 & 1 & 0 & 0 & 0 \\ 16 & 1 & 0 & 0 & 0 \\ 14 & 1 & 0 & 0 & 0 \\ 15 & 1 & 0 & 0 & 0 \\ 12 & 0 & 1 & 0 & 0 \\ 10 & 0 & 1 & 0 & 0 \\ 15 & 0 & 1 & 0 & 0 \\ 19 & 0 & 1 & 0 & 0 \\ 11 & 0 & 1 & 0 & 0 \\ 23 & 0 & 0 & 1 & 0 \\ 20 & 0 & 0 & 1 & 0 \\ 18 & 0 & 0 & 1 & 0 \\ 17 & 0 & 0 & 1 & 0 \\ 27 & 0 & 0 & 0 & 1 \\ 33 & 0 & 0 & 0 & 1 \\ 22 & 0 & 0 & 0 & 1 \\ 26 & 0 & 0 & 0 & 1 \\ 28 & 0 & 0 & 0 & 1 \end{pmatrix} \qquad X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Least Squares Estimation of the Regression Parameters:

$$b := \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot Y \qquad b = \begin{pmatrix} 14.6 \\ 13.4 \\ 19.5 \\ 27.2 \end{pmatrix} \qquad \begin{pmatrix} \mu 1 \\ \mu 2 \\ \mu 3 \\ \mu 4 \end{pmatrix}$$

< estimating μ parameters in the model

^ vector of regresion coefficients = cell (block) means

Note: In the Cell Means ANOVA model, regression coefficients in the Linear Model give the estimate of means for each cell (block).

Fitted Values & Hat Matrix H:

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

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

Residuals:

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

ANOVA Table:

Sum of Squares: Degrees of Freedom: Mean Squares:

$SSTR := Y^T \cdot \left[H - \left(\frac{1}{N}\right) \cdot J\right] \cdot Y$ $SSTR = (588.2211)$ $df_R := p - 1$ $df_R = 3$ $MSTR := \frac{SSTR}{df_R}$ $MSTR = (196.0737)$

$SSE := Y^T \cdot (I - H) \cdot Y$ $SSE = (158.2)$ $df_E := N - p$ $df_E = 15$ $MSE := \frac{SSE}{df_E}$ $MSE = (10.5467)$

$SSTO := Y^T \cdot \left[I - \left(\frac{1}{N}\right) \cdot J\right] \cdot Y$ $SSTO = (746.4211)$ $df_T := N - 1$ $df_T = 18$ $MSTO := \frac{SSTO}{df_T}$ $MSTO = (41.4678)$

Note: The ANOVA table is calculated in exactly the same way as for regression. SSR now becomes SSTR (for treatment), both with df = 3.

A Linear Combination of Means:

$b = \begin{pmatrix} 14.6 \\ 13.4 \\ 19.5 \\ 27.2 \end{pmatrix}$ **< cell means** $c := \begin{pmatrix} 0.5 \\ 0.5 \\ -0.5 \\ -0.5 \end{pmatrix}$ **< Contrast Vector = Coefficients of Linear Combination L**

$L := \left(b^T \cdot c\right)_0$ $L = -9.35$ **<** $L = c_1 Ybar_1 + c_2 Ybar_2 + c_3 Ybar_3 + \ldots + c_k Ybar_k$

Linear Combination with Further Condition as Linear Contrast:

$\sum c = 0$ **< Coefficients of the Linear Combination must add to zero to be a Linear Contrast**

Single Test of Contrast Vector:

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

Cell Means Model:

$Y_{i,j} = \mu_i + \varepsilon_{i,j}$

Factor Effects Model:

$Y_{i,j} = \mu. + \tau_i + \varepsilon_{i,j}$

Equivalent Models, where:
$\mu.$ is the grand mean of all objects
μ_i is mean of each treatment class i (cell or block)
τ_i is the treatment effect = $\mu. - \mu_i$ for each class i.
$\varepsilon_{i,j}$ is the error term specific to each object i,j
i = 1 to r, j = 1 to n_r for each treatment class i

Assumption:

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

For Linear Contrast:

$L = \mu_i c_i$

where:
μ is the vector of Class means
c is some specified contrast vector

Restriction:

$\sum c = 0$

Hypotheses:

H_0: L = 0 < Linear Contrast is zero

H_1: L <> 0

Test Statistic:

$i := 0..r-1 \qquad r = 4$

$$t := \frac{L}{\sqrt{MSE \cdot \left[\sum_i \frac{(c_i)^2}{n_i}\right]}} \qquad t = (-6.2456)$$

< Linear Contrast normalized by Standard Error & and Cell sizes

Critical Value of the Test:

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

$C := qt\left(\frac{\alpha}{2}, N-r\right) \qquad C = -2.1314$ < Note degrees of freedom = (N-r)

Decision Rule:

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

Probability Value:

$P := \min[2 \cdot pt(t, N-r), 2 \cdot (1 - pt(t, N-r))] \qquad P = 1.5675 \times 10^{-5}$

Confidence Interval of the Linear Contrast:

$$s_L := \sqrt{MSE \cdot \left[\sum_i \frac{(c_i)^2}{n_i}\right]} \qquad s_{L_0} = 1.4971$$

< Standard Error of L

$$CI := \left[(L - |C| \cdot s_L)_0 \quad L \quad (L + |C| \cdot s_L)_0 \right] \qquad CI = (-12.5409 \quad -9.35 \quad -6.1591)$$

^ upper limit

confirmed KNNL p. 743 ^ point estimate of L

^ lower limit

Prototype in R:

```
#TREATMENTS MODEL IN R
#READ STRUCTURED DATA TABLE WITH NUMERIC CODED FACTOR
K=read.table("c:/2008LinearModelsData/KentonFoodR.txt")
K
attach(K)

Y=Sales
X=factor(Design) # factor() IN DEFAULT SETTING

FM=lm(Y~X)
model.matrix(FM)
summary(FM)
anova(FM)
```

```
> anova(FM)
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
X          3 588.22  196.07  18.591 2.585e-05 ***
Residuals 15 158.20   10.55
```

```
#CONSTRUCTING CELL MEANS MODEL IN R FROM factor()
contrasts(X,contrasts=F)
CM=lm(Y~X-1)  #FITS CELL MEANS MODEL
model.matrix(CM)
summary(CM)
Ys=scale(Y,scale=F)  #CENTER Y
CMs=lm(Ys~X)
anova(CMs)
```

```
> summary(CM)
Call:
lm(formula = Y ~ X - 1)
Residuals:
   Min     1Q Median     3Q    Max
 -5.20  -1.95  -0.20   1.50   5.80
Coefficients:
   Estimate Std. Error t value Pr(>|t|)
X1   14.600      1.452  10.053 4.66e-08 ***
X2   13.400      1.452   9.226 1.43e-07 ***
X3   19.500      1.624  12.009 4.28e-09 ***
X4   27.200      1.452  18.728 8.16e-12 ***
```

```
> anova(CMs)
Analysis of Variance Table
Response: Ys
          Df Sum Sq Mean Sq F value
Pr(>F)
X          3 588.22  196.07  18.591
2.585e-05 ***
Residuals 15 158.20   10.55
```

```
#MAKING CONTRASTS IN R
require(gmodels)  #MUST LOAD {gmodels} PACKAGE FROM CRAN

#CONTRAST FOR KNNL P. 743
fit.contrast(FM,X,c(0.5,0.5,-0.5,-0.5),conf.int=0.95)
```

```
> fit.contrast(FM,X,c(0.5,0.5,-0.5,-0.5),conf.int=0.95)
                          Estimate Std. Error   t value     Pr(>|t|)  lower CI  upper CI
X c=( 0.5 0.5 -0.5 -0.5 )    -9.35   1.497053 -6.245605 1.567514e-05 -12.54089 -6.159108
```

^ Note inverted order of signs in R's fit.contrast() function.

^ values confirmed KNNL p. 743 and above.

Multiple Tests of Contrast Matrix:

Multiple Linear Combinations of Means:

$$b = \begin{pmatrix} 14.6 \\ 13.4 \\ 19.5 \\ 27.2 \end{pmatrix}$$

< cell means

$$c := \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$$

< Contrast Matrix with columns giving different Linear Combinations as Contrasts

$$L := \left(b^T \cdot c\right) \qquad L^T = \begin{pmatrix} -18.7 \\ -6.5 \\ 8.9 \end{pmatrix}$$

< three linear contrasts

Linear Combination with Further Condition as Linear Contrast:

$$\begin{pmatrix} \sum c^{\langle 0 \rangle} \\ \sum c^{\langle 1 \rangle} \\ \sum c^{\langle 2 \rangle} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

< coefficients of *each* Linear Combination must add to zero to be a Linear Contrast

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

Cell Means Model:

$Y_{i,j} = \mu_i + \varepsilon_{i,j}$

Factor Effects Model:

$Y_{i,j} = \mu. + \tau_i + \varepsilon_{i,j}$

Equivalent Models, where:
$\mu.$ is the grand mean of all objects
μ_i is mean of each treatment class i (cell or block)
τ_i is the treatment effect = $\mu. - \mu_i$ for each class i.
$\varepsilon_{i,j}$ is the error term specific to each object i,j
i = 1 to r, j = 1 to n_r for each treatment class i

Assumption:

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

For Linear Contrast:

$L_k = \mu_i c_i$

where *for each Linear Contrast* k:
μ is the vector of Class means
c is some specified contrsast vector

Restriction:

$$\begin{pmatrix} \sum c^{\langle 0 \rangle} \\ \sum c^{\langle 1 \rangle} \\ \sum c^{\langle 2 \rangle} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Each Hypotheses:

H_0: $L_k = 0$
H_1: L_k <> 0

< *for each Contrast* k:
Linear Contrast is zero

Test Statistic:

$$t_0 := \frac{\left(L^T\right)_0}{\sqrt{MSE_0 \cdot \left[\sum_i \frac{\left[\left(c^{\langle 0 \rangle}\right)_i\right]^2}{n_i}\right]}}$$

$t_0 = -6.2456$

$$t_1 := \frac{\left(L^T\right)_1}{\sqrt{MSE_0 \cdot \left[\sum_i \frac{\left[\left(c^{\langle 1 \rangle}\right)_i\right]^2}{n_i}\right]}}$$ $t_1 = -2.1709$

< Linear Contrasts normalized by Standard Error & and Cell sizes

$$t_2 := \frac{\left(L^T\right)_2}{\sqrt{MSE_0 \cdot \left[\sum_i \frac{\left[\left(c^{\langle 2 \rangle}\right)_i\right]^2}{n_i}\right]}}$$ $t_2 = 2.9725$

$$t := \begin{pmatrix} t_0 \\ t_1 \\ t_2 \end{pmatrix}$$

Critical Value of the Test:

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

$C := qt\left(\frac{\alpha}{2}, N - r\right)$ $C = -2.1314$ **< Note degrees of freedom = (N-r)**

Decision Rule:

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

Probability Value:

$j := 0 .. r - 2$ $r = 4$ **< index for contrasts**

$P_0 := \min\left[2 \cdot pt(t_0, N - r), 2 \cdot (1 - pt(t_0, N - r))\right]$ $P_0 = 0.0000156751$

$P_1 := \min\left[2 \cdot pt(t_1, N - r), 2 \cdot (1 - pt(t_1, N - r))\right]$ $P_1 = 0.0463943972$

$P_2 := \min\left[2 \cdot pt(t_2, N - r), 2 \cdot (1 - pt(t_2, N - r))\right]$ $P_2 = 0.0094890939$

$$P := \begin{pmatrix} P_0 \\ P_1 \\ P_2 \end{pmatrix}$$

Confidence Interval of the Linear Contrast:

$$s_{L0} := \sqrt{MSE_0 \cdot \left[\sum_i \frac{\left[\left(c^{\langle 0 \rangle}\right)_i\right]^2}{n_i}\right]}$$ $s_{L0} = 2.9941$

$$s_{L1} := \sqrt{MSE_0 \cdot \left[\sum_i \frac{\left[\left(c^{\langle 1 \rangle}\right)_i\right]^2}{n_i}\right]}$$ $s_{L1} = 2.9941$

$$s_{L2} := \sqrt{MSE_0 \cdot \left[\sum_i \frac{\left[\left(c^{\langle 2 \rangle}\right)_i\right]^2}{n_i}\right]}$$ $s_{L2} = 2.9941$

$$s_L := \begin{pmatrix} s_{L0} \\ s_{L1} \\ s_{L2} \end{pmatrix}$$

$CI_0 := \left[\left(L^T\right)_0 - |C| \cdot s_{L0} \quad \left(L^T\right)_0 \quad \left(L^T\right)_0 + |C| \cdot s_{L0}\right]$ $CI_0 = (-25.081784 \quad -18.7 \quad -12.318216)$

$CI_1 := \left[\left(L^T\right)_1 - |C| \cdot s_{L1} \quad \left(L^T\right)_1 \quad \left(L^T\right)_1 + |C| \cdot s_{L1}\right]$ $CI_1 = (-12.881784 \quad -6.5 \quad -0.118216)$

$CI_2 := \left[\left(L^T\right)_2 - |C| \cdot s_{L2} \quad \left(L^T\right)_2 \quad \left(L^T\right)_2 + |C| \cdot s_{L2}\right]$ $CI_2 = (2.518216 \quad 8.9 \quad 15.281784)$

^ lower limit ^ upper limit

^ point estimate of L

$$CI := \begin{pmatrix} CI_0 \\ CI_1 \\ CI_2 \end{pmatrix}$$

Prototype in R:

```
#CONTRASTS FOR R TREATMENTS CODING

summary(FM) #GIVES COMPARISON OF DIFFERENCE IN MEANS FROM THE FIRST
#SAME CALCULATION WITH EXPLICIT CONTRASTS
fit.contrast(FM,X,rbind(c(-1,1,0,0),c(-1,0,1,0),c(-1,0,0,1)),conf.int=0.95)
```

```
> summary(FM) #GIVES COMPARISON OF DIFFERENCE IN MEANS
FROM THE FIRST
Call:
lm(formula = Y ~ X)
Residuals:
   Min     1Q Median     3Q    Max
 -5.20  -1.95  -0.20   1.50   5.80
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   14.600      1.452  10.053 4.66e-08 ***
X2            -1.200      2.054  -0.584   0.5677
X3             4.900      2.179   2.249   0.0399 *
X4            12.600      2.054   6.135 1.91e-05 ***
```

```
> #SAME CALCULATION WITH EXPLICIT CONTRASTS
> fit.contrast(FM,X,rbind(c(-1,1,0,0),c(-1,0,1,0),c(-1,0,0,1)),conf.int=0.95)
                  Estimate Std. Error    t value     Pr(>|t|)   lower CI  upper CI
X c=( -1 1 0 0 )      -1.2   2.053939 -0.5842432 5.677402e-01 -5.5778680  3.177868
X c=( -1 0 1 0 )       4.9   2.178532  2.2492214 3.994770e-02  0.2565698  9.543430
X c=( -1 0 0 1 )      12.6   2.053939  6.1345532 1.910149e-05  8.2221320 16.977868
```

Note: R's default summary() report for a Treatments Linear Model gives the mean of Cell 1, followed by differences between means of the other Cells and Cell 1. These are in fact Linear Contrasts of the type specified in fit.contrast{gmodels} Contrast Matrix.

```
#CONTRASTS FOR COMPARING PAIRS OF MEANS IN 2008 LINEAR MODELS 16
fit.contrast(FM,X,rbind(c(1,1,-1,-1),c(1,-1,1,-1),c(1,-1,-1,1)),conf.int=0.95)
```

```
> fit.contrast(FM,X,rbind(c(1,1,-1,-1),c(1,-1,1,-1),c(1,-1,-1,1)),conf.int=0.95)
                   Estimate Std. Error   t value     Pr(>|t|)   lower CI    upper CI
X c=( 1 1 -1 -1 )     -18.7   2.994105 -6.245605 1.567514e-05 -25.081784 -12.3182156
X c=( 1 -1 1 -1 )      -6.5   2.994105 -2.170932 4.639440e-02 -12.881784  -0.1182156
X c=( 1 -1 -1 1 )       8.9   2.994105  2.972507 9.489094e-03   2.518216  15.2817844
```

$$c^T = \begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

$$L^T = \begin{pmatrix} -18.7 \\ -6.5 \\ 8.9 \end{pmatrix} \quad s_L = \begin{pmatrix} 2.994 \\ 2.994 \\ 2.994 \end{pmatrix} \quad t = \begin{pmatrix} -6.2456 \\ -2.1709 \\ 2.9725 \end{pmatrix} \quad P = \begin{pmatrix} 1.568 \times 10^{-5} \\ 4.639 \times 10^{-2} \\ 9.489 \times 10^{-3} \end{pmatrix} \quad CI = \begin{bmatrix} (-25.082 & -18.7 & -12.318) \\ (-12.882 & -6.5 & -0.118) \\ (2.518 & 8.9 & 15.282) \end{bmatrix}$$

^ same as calculated above