

## 4.3: Cell Means Model

### Model 2 - The Cell Means Model

$$Y_{ij} = \mu_i + \epsilon_{ij} \quad (4.3.1)$$

where  $\mu_i$ ,  $i = 1, 2, \dots, T$  are the factor level means. Note that in this model, there is no overall mean being fitted.

The cell means model does not fit an overall mean, but instead fits an individual mean for each of the treatment levels. Let us run this model for the same data assuming that each pair of observations arise from one treatment level, so that  $T$ , the number of treatment levels equals 3. We then have to replace the design matrix in the IML code with:

```
/* The Cell Means Model */
x={
1    0    0,
1    0    0,
0    1    0,
0    1    0,
0    0    1,
0    0    1};
```

Notice that each column represents a specific treatment level and is using indicator coding: 1 for the rows corresponding to the observations receiving the specified treatment level, and 0 for the other rows. It can be seen that  $r = 2$  is the number of replicates for each treatment level. Observe that column 1 generates the mean for treatment level 1, column 2 for treatment level 2, and column 3 for treatment level 3.

To write the cell means model as a GLM, let

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \mathbf{x}_3' \\ \mathbf{x}_4' \\ \mathbf{x}_5' \\ \mathbf{x}_6' \end{bmatrix} \quad (4.3.2)$$

where  $\mathbf{x}_i'$  is the  $i^{th}$  row vector of the design matrix.

The parameter vector  $\boldsymbol{\beta}$  is a 3-dimensional column vector and is defined by

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \quad (4.3.3)$$

The parameter estimates  $\hat{\boldsymbol{\beta}}$  can again be found using the least squares method. One can verify that  $\mu_i = \bar{y}_i$ , the  $i^{th}$  treatment mean, for  $i = 1, 2, 3$ . Using this estimate, the resulting estimated regression equation for the cell means model is,

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \quad (4.3.4)$$

which produces  $\hat{\mathbf{y}}_i = \mathbf{x}_i' \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \end{bmatrix}$ .

We then re-run the program with the new design matrix to get the following output:

Regression Coefficients

Regression Coefficients				1.5
Beta_1				3.5
Beta_0				1.5
Beta_2				5.5
Beta_1				3.5
ANOVA				
Treatment	dF	SS	MS	F
	2	16	8	16
Error	3	1.5	0.5	
Total	5	17.5		

The regression coefficients  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$  are now the means for each treatment level, and in the ANOVA table, we see that the  $SS_{Error}$  is 1.5. This reduction in the  $SS_{Error}$  is the  $SS_{treatment}$ . Notice that the error SS of the Overall Mean model is the sum of the SS values for Treatment and Error term in this model, which means that by not including the treatment effect in that model, its error SS has been unduly inflated.

Adding the optional code given in Section 4.2 to compute additional Internal computations, we can obtain:

xprimex			check			xprimey		
2	0	0	1	0	0			3
0	2	0	0	1	0			7
0	0	2	0	0	1			11
						xprimexinv		
SumY2			CF			0.5	0	0
89.5			73.5			0	0.5	0
						0	0	0.5

Here we can see that  $\mathbf{X}'\mathbf{X}$  now contains diagonal elements that are the  $n_i$  = number of observations for each treatment level mean being computed. In addition, we can verify that  $CF = \sum Y^2 - CF = 16$ , or the working formula equals the treatment  $SS$ .

We can now test for the significance of the treatment by using the General Linear  $F$  test:

$$F = \frac{SSE_{reduced} - SSE_{full}/dfE_{reduced} - dfE_{full}}{SSE_{full}/dfE_{full}} \quad (4.3.5)$$

The Overall Mean model is the "Reduced" model, and the Cell Means model is the "Full" model. From the ANOVA tables, we get:

$$F = \frac{17.5 - 1.5/5 - 3}{1.5/3} = 16 \quad (4.3.6)$$

which can be compared to  $F_{.05,2,3} = 9.55$ .

## Using R

### Steps in R - Cell Means Model

1. Define response variable and design matrix

```
y<-matrix(c(2,1,3,4,6,5), ncol=1)
x<matrix(c(1,0,0,1,0,0,0,1,0,0,1,0,0,0,1,0,0,1),ncol=3,nrow=6,byrow=TRUE)
```

## 2. Regression coefficients

```
beta<-solve(t(x)%*%x)%*(t(x)%*%y)
# beta
#      [,1]
# [1,]  1.5
# [2,]  3.5
# [3,]  5.5
```

## 3. Calculate the entries of the ANOVA Table

```
n<-nrow(y)
p<-ncol(x)
J<-matrix(1,n,n)
ss_tot = (t(y)%*%y) - (1/n)*(t(y)%*%J)%*%y #17.5
ss_trt = t(beta)%*(t(x)%*%y) - (1/n)*(t(y)%*%J)%*%y #16
ss_error = ss_tot - ss_trt #1.5
total_df=n-1 #5
trt_df=p-1 #2
error_df=n-p #3
MS_trt = ss_trt/(p-1) #8
MS_error = ss_error / error_df #0.5
F=MS_trt/MS_error #16
```

## 4. Creating the ANOVA table

```
ANOVA <- data.frame(
  c("", "Treatment", "Error", "Total"),
  c("DF", trt_df, error_df, total_df),
  c("SS", ss_trt, ss_error, ss_tot),
  c("MS", MS_trt, MS_error, ""),
  c("F", F, "", ""),
  stringsAsFactors = FALSE)
names(ANOVA) <- c(" ", " ", " ", " ", "", "", "")
```

## 5. Print the ANOVA table

```
print(ANOVA)
# 1      DF      SS      MS      F
# 2 Treatment  2     16      8     16
# 3 Error      3     1.5    0.5
# 4 Total      5    17.5
```

## 6. Intermediates in the matrix computations

```
xprimex<-t(x)%%x
# xprimex
#      [,1] [,2] [,3]
# [1,]    2    0    0
# [2,]    0    2    0
# [3,]    0    0    2
xprimey<-t(x)%%y
# xprimey
#      [,1]
# [1,]    3
# [2,]    7
# [3,]   11
xprimexinv<-solve(t(x)%%x)
# xprimexinv
#      [,1] [,2] [,3]
# [1,]  0.5  0.0  0.0
# [2,]  0.0  0.5  0.0
# [3,]  0.0  0.0  0.5
check<-xprimexinv)%%xprimex
# check
#      [,1] [,2] [,3]
# [1,]    1    0    0
# [2,]    0    1    0
# [3,]    0    0    1
SumY2<-t(beta)%%(t(x)%%y) #89.5
CF<-(1/n)*(t(y)%%J)%%y #73.5
```

This page titled [4.3: Cell Means Model](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Penn State's Department of Statistics](#).