Main Effects in the R Additive Model Formulation

No Diet main effect $\iff \alpha_2 = 0$

this differs from the set-up using the less-than-full rank model matrix

end lecture 10 2-12-25

No Drug main effects $\iff \beta_2 = \beta_3 = 0$

	Drug 1	Drug 2	Drug 3	
Diet 1	μ	$\mu + \beta_2$	$\mu + \beta_3$	$\mu + \frac{\beta_2+\beta_3}{3}$
Diet 2	$\mu + \alpha_2$	$\mu + \alpha_2 + \beta_2$	$\mu + \alpha_2 + \beta_3$	$\mu + \alpha_2 + \frac{\beta_2+\beta_3}{3}$
	$\mu + \frac{\alpha_2}{2}$	$\mu + \frac{\alpha_2}{2} + \beta_2$	$\mu + \frac{\alpha_2}{2} + \beta_3$	$\mu + \frac{\alpha_2}{2} + \frac{\beta_2+\beta_3}{3}$

all the same when $\beta_2 = \beta_3 = 0$

H_0: No Diet Main Effect ($\alpha_2 = 0$ in R)

$$c^T = \begin{array}{c} \begin{array}{cccc} & \alpha_2 & & \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \end{array} \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$$

H_0: No Drug Main Effects ($\beta_2 = \beta_3 = 0$ in R)

$$C = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

LSMEANS are still the OLS estimators of the quantities in the margins below.

	Drug 1	Drug 2	Drug 3	
Diet 1	μ	$\mu + \beta_2$	$\mu + \beta_3$	$\mu + \frac{\beta_2+\beta_3}{3}$
Diet 2	$\mu + \alpha_2$	$\mu + \alpha_2 + \beta_2$	$\mu + \alpha_2 + \beta_3$	$\mu + \alpha_2 + \frac{\beta_2+\beta_3}{3}$
	$\mu + \frac{\alpha_2}{2}$	$\mu + \frac{\alpha_2}{2} + \beta_2$	$\mu + \frac{\alpha_2}{2} + \beta_3$	$\mu + \frac{\alpha_2}{2} + \frac{\beta_2+\beta_3}{3}$

OLS

OLS

OLS

LSMEANS for the Additive Model (continued)

For example, the LSMEAN for Diet 1 is

$$\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}} = \left[1, 0, \frac{1}{3}, \frac{1}{3}\right] \underbrace{\begin{bmatrix} \widehat{\mu} \\ \widehat{\alpha}_2 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{bmatrix}}_{\widehat{\beta}} = \underbrace{\widehat{\mu} + \frac{\widehat{\beta}_2 + \widehat{\beta}_3}{3}}_{\text{estimated diet 1 marginal mean}},$$

after plugging in data

where $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$.

```
> o=lm(weightgain~diet+drug,data=d)
> model.matrix(o)
   (Intercept) diet2 drug2 drug3
1            1     0     0     0
2            1     0     0     0
3            1     0     1     0
4            1     0     1     0
5            1     0     0     1
6            1     0     0     1
7            1     1     0     0
8            1     1     0     0
9            1     1     1     0
10           1     1     1     0
11           1     1     0     1
12           1     1     0     1
```

μ α_2 β_2 β_3

full-rank model matrix X_R

$\mu_{23} = \mu + \alpha_2 + \beta_3$

```
> #betahat vector:
>
> coef(o)
(Intercept)       diet2       drug2       drug3
  41.616667   -5.033333   -2.100000   -2.550000
```

$\hat{\mu}$ $\hat{\alpha}_2$ $\hat{\beta}_2$ $\hat{\beta}_3$

R: $\widehat{\text{Var}}(\hat{\beta})$ and Error Degrees of Freedom

```
> #Estimated variance of betahat:
>
> vcov(o)
            (Intercept)         diet2         drug2         drug3
(Intercept)   0.8186111 -4.093056e-01 -6.139583e-01 -6.139583e-01
diet2        -0.4093056  8.186111e-01 -6.759159e-17 -6.759159e-17
drug2        -0.6139583 -6.759159e-17  1.227917e+00  6.139583e-01
drug3        -0.6139583 -6.759159e-17  6.139583e-01  1.227917e+00

> #The degrees of freedom for error:
>
> o$df
[1] 8
```

$\widehat{Var}(\hat{\beta})$

$\widehat{Var}(\hat{\mu}) = \widehat{Var}(\hat{\alpha}_2)$

$\widehat{Var}(\hat{\beta}_2) = \widehat{Var}(\hat{\beta}_3)$

R: A Function for Point and Interval Estimation

Output associated with lm function

```
> estimate=function(lmout,C,a=0.05)
+ {
+   b=coef(lmout)
+   V=vcov(lmout)
+   df=lmout$df
+   Cb=C%*%b
+   se=sqrt(diag(C%*%V%*%t(C)))
+   tval=qt(1-a/2,df)
+   low=Cb-tval*se
+   up=Cb+tval*se
+   m=cbind(C,Cb,se,low,up)
+   dimnames(m)[[2]]=c(paste("c",1:ncol(C),sep=""),
+                "estimate","se",
+                paste(100*(1-a),"% Conf.",sep=""),
+                "limits")
+   m
+ }
```

$\hat{\beta}$

$\hat{Var}(\hat{\beta})$

$C\hat{\beta}$

$SE(C\hat{\beta})$

$C\hat{\beta} \pm$ Critical value $*$ $SE(C\hat{\beta})$

R: Entering a C Matrix

```
> C=matrix(c(
+ 1, 0, 1/3, 1/3,
+ 1, 1, 1/3, 1/3,
+ 1, 1/2, 0, 0,
+ 1, 1/2, 1, 0,
+ 1, 1/2, 0, 1,
+ 0, -1, 0, 0,
+ 0, 0, -1,  0,
+ 0, 0,  0, -1,
+ 0, 0,  1, -1
+ ),ncol=4,byrow=T)
```

LSmean for diet 1 (OLS)
LSmean for diet 2
LS mean for drug 1
drug 2
drug 3
diet 1 vs. diet 2 *
drug 1 vs. drug 2 *
drug 1 vs. drug 3 *
drug 2 vs. drug 3 *

* these effects are the same regardless of what level the other factor is bc we have no interactions

R: The C Matrix

```
> C
     [,1] [,2]       [,3]       [,4]
 [1,]    1  0.0  0.3333333  0.3333333
 [2,]    1  1.0  0.3333333  0.3333333
 [3,]    1  0.5  0.0000000  0.0000000
 [4,]    1  0.5  1.0000000  0.0000000
 [5,]    1  0.5  0.0000000  1.0000000
 [6,]    0 -1.0  0.0000000  0.0000000
 [7,]    0  0.0 -1.0000000  0.0000000
 [8,]    0  0.0  0.0000000 -1.0000000
 [9,]    0  0.0  1.0000000 -1.0000000
```

R: Interpreting the C Matrix

```
>
> #With this choice of C, you get estimates and
> #confidence intervals for the following:
>
> #Row 1: lsmean for diet 1
> #Row 2: lsmean for diet 2
> #Row 3: lsmean for drug 1
> #Row 4: lsmean for drug 2
> #Row 5: lsmean for drug 3
> #Row 6: diet 1 - diet 2 effect
> #Row 7: drug 1 - drug 2 effect
> #Row 8: drug 1 - drug 3 effect
> #Row 9: drug 2 - drug 3 effect
```

Simple effects

R: Results

estimates

estimated standard error

```
> estimate(o,C)
     c1   c2         c3         c4  estimate        se
[1,]  1  0.0  0.3333333  0.3333333 40.066667 0.6397699
[2,]  1  1.0  0.3333333  0.3333333 35.033333 0.6397699
[3,]  1  0.5  0.0000000  0.0000000 39.100000 0.7835549
[4,]  1  0.5  1.0000000  0.0000000 37.000000 0.7835549
[5,]  1  0.5  0.0000000  1.0000000 36.550000 0.7835549
[6,]  0 -1.0  0.0000000  0.0000000  5.033333 0.9047713
[7,]  0  0.0 -1.0000000  0.0000000  2.100000 1.1081140
[8,]  0  0.0  0.0000000 -1.0000000  2.550000 1.1081140
[9,]  0  0.0  1.0000000 -1.0000000  0.450000 1.1081140

        95% Conf.    limits
[1,]  38.591354577 41.541979
[2,]  33.558021243 36.508645
[3,]  37.293119084 40.906881
[4,]  35.193119084 38.806881
[5,]  34.743119084 38.356881
[6,]  2.946926967  7.119740
[7,] -0.455315497  4.655315
[8,] -0.005315497  5.105315
[9,] -2.105315497  3.005315
```

CIs don't contain zero

CIs contain zero

R: Results

```
> estimate(o,C)[,-(1:4)]
      estimate        se    95% Conf.    limits
 [1,] 40.066667 0.6397699 38.591354577 41.541979
 [2,] 35.033333 0.6397699 33.558021243 36.508645
 [3,] 39.100000 0.7835549 37.293119084 40.906881
 [4,] 37.000000 0.7835549 35.193119084 38.806881
 [5,] 36.550000 0.7835549 34.743119084 38.356881
 [6,]  5.033333 0.9047713  2.946926967  7.119740
 [7,]  2.100000 1.1081140 -0.455315497  4.655315
 [8,]  2.550000 1.1081140 -0.005315497  5.105315
 [9,]  0.450000 1.1081140 -2.105315497  3.005315
```

R: Function for Testing $H_0 : \boldsymbol{C\beta} = \boldsymbol{d}$

```
> test=function(lmout,C,d=0){
+   b=coef(lmout)
+   V=vcov(lmout)
+   dfn=nrow(C)
+   dfd=lmout$df
+   Cb.d=C%*%b-d
+   Fstat=drop(
         t(Cb.d)%*%solve(C%*%V%*%t(C))%*%Cb.d/dfn)
+   pvalue=1-pf(Fstat,dfn,dfd)
+   list(Fstat=Fstat,pvalue=pvalue)
+ }
```

$$F = \frac{(C\hat{\beta}_R - \underset{\sim}{d})^T (C(X_R^T X_R)^- C^T)^{-1} (C\hat{\beta}_R - \underset{\sim}{d}) / q}{\hat{\sigma}^2}$$

Fitting the Additive Model in R

from earlier: $H_0: \alpha_2 = 0$

```
> #Test for diet main effect:
> C=matrix(c(
+ 0, 1, 0, 0
+ ), nrow=1, byrow=T)
> C
     [,1] [,2] [,3] [,4]
[1,]    0    1    0    0
> test(o,C)
$Fstat
[1] 30.94808

$pvalue
[1] 0.0005327312
```

$C^T = (0 \quad 1 \quad 0 \quad 0)$

fairly strong evidence in favor of a main diet effect $\Rightarrow$ marginal means for diet 1 and 2 differ significantly

```
> #Test for drug main effects:
> C=matrix(c(
+ 0, 0, -1,  0,
+ 0, 0,  0, -1
+ ), nrow=2, byrow=T)
> C
     [,1] [,2] [,3] [,4]
[1,]    0    0   -1    0
[2,]    0    0    0   -1
> test(o,C)
$Fstat
[1] 3.017306

$pvalue
[1] 0.1055743
```

* however, CIs involving drug 1 have a lower bound close to zero ...

weak evidence in favor of a main drug effect - which matches the conclusion based on CIs on slide 34 *

Can we estimate average movie ratings and individual customer ratings? yes under an additive model but no under a cell-means model

		movie 1	movie 2	movie 3
customer	1	4	1	?
	2	?	3	5
	3	?	?	3
	4	3	1	?

Can we guess ratings for customer/movie combinations not in the dataset?

$$y_{ij} = \text{customer } i\text{'s rating of movie } j$$

1 through 5

Which movie is best?

$$y_{ij} = \mu + c_i + m_j + \epsilon_{ij}$$

end lecture 11 2-14-25

