ANOVA Table for Our Two-Factor Example

by how much does SSE decrease when adding diet into model?

end lecture 15 2-24-25

Source	Sum of Squares	DF
Diets$\|1$	$\boldsymbol{y}^\top(\boldsymbol{P}_2 - \boldsymbol{P}_1)\boldsymbol{y}$	$2 - 1 = 1$
Drugs$\|1$, Diets	$\boldsymbol{y}^\top(\boldsymbol{P}_3 - \boldsymbol{P}_2)\boldsymbol{y}$	$4 - 2 = 2$
Diets $\times$ Drugs$\|1$, Diets, Drugs	$\boldsymbol{y}^\top(\boldsymbol{P}_4 - \boldsymbol{P}_3)\boldsymbol{y}$	$6 - 4 = 2$
Error	$\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_4)\boldsymbol{y}$	$12 - 6 = 6$
C. Total	$\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{y}$	$12 - 1 = 11$

given intercept & diet, how much more variability can we explain by adding drug?

ANOVA Table for Our Two-Factor Example

sequential SS : conditioning on factors that previously entered the model

Source	Sum of Squares	DF
Diets	$\boldsymbol{y}^\top(\boldsymbol{P}_2 - \boldsymbol{P}_1)\boldsymbol{y}$	$2 - 1 = 1$
Drugs	$\boldsymbol{y}^\top(\boldsymbol{P}_3 - \boldsymbol{P}_2)\boldsymbol{y}$	$4 - 2 = 2$
Diets $\times$ Drugs	$\boldsymbol{y}^\top(\boldsymbol{P}_4 - \boldsymbol{P}_3)\boldsymbol{y}$	$6 - 4 = 2$
Error	$\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_4)\boldsymbol{y}$	$12 - 6 = 6$
C. Total	$\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{y}$	$12 - 1 = 11$

The Diet-Drug Dataset

```
> d
   diet drug weightgain
1     1    1       41.3
2     1    1       43.7
3     1    2       40.9
4     1    2       39.2
5     1    3       37.4
6     1    3       37.9
7     2    1       36.8
8     2    1       34.6
9     2    2       33.6
10    2    2       34.3
11    2    3       35.8
12    2    3       35.1
```

R Code and Output for Two-Factor ANOVA

```
> d$diet=factor(d$diet)
> d$drug=factor(d$drug)
>
> a=d$diet
> b=d$drug
> y=d$weightgain
```

R Code and Output for Two-Factor ANOVA

```
> x1=matrix(1,nrow=nrow(d),ncol=1)
> x1
      [,1]
 [1,]    1
 [2,]    1
 [3,]    1
 [4,]    1
 [5,]    1
 [6,]    1
 [7,]    1
 [8,]    1
 [9,]    1
[10,]    1
[11,]    1
[12,]    1
```

R Code and Output for Two-Factor ANOVA

diet added

```
> x2=cbind(x1,model.matrix(~0+a))
> x2
   x1 a1 a2
1   1  1  0
2   1  1  0
3   1  1  0
4   1  1  0
5   1  1  0
6   1  1  0
7   1  0  1
8   1  0  1
9   1  0  1
10  1  0  1
11  1  0  1
12  1  0  1
```


R Code and Output for Two-Factor ANOVA

```
> x3=cbind(x2,model.matrix(~0+b))
> x3
   x1 a1 a2 b1 b2 b3
1   1  1  0  1  0  0
2   1  1  0  1  0  0
3   1  1  0  0  1  0
4   1  1  0  0  1  0
5   1  1  0  0  0  1
6   1  1  0  0  0  1
7   1  0  1  1  0  0
8   1  0  1  1  0  0
9   1  0  1  0  1  0
10  1  0  1  0  1  0
11  1  0  1  0  0  1
12  1  0  1  0  0  1
```

```
> x4=model.matrix(~0+b:a)
> x4
   b1:a1 b2:a1 b3:a1 b1:a2 b2:a2 b3:a2
1      1     0     0     0     0     0
2      1     0     0     0     0     0
3      0     1     0     0     0     0
4      0     1     0     0     0     0
5      0     0     1     0     0     0
6      0     0     1     0     0     0
7      0     0     0     1     0     0
8      0     0     0     1     0     0
9      0     0     0     0     1     0
10     0     0     0     0     1     0
11     0     0     0     0     0     1
12     0     0     0     0     0     1
```

R Code and Output for Two-Factor ANOVA

```
> library(MASS)
> proj=function(x){
+   x%*%ginv(t(x)%*%x)%*%t(x)
+ }
>
> p1=proj(x1)
> p2=proj(x2)
> p3=proj(x3)
> p4=proj(x4)
> I=diag(rep(1,12))
```

R Code and Output for Two-Factor ANOVA

for diet

$y^T(P_{j+1} - P_j)y$

```
> SumOfSquares=c(
+ t(y)%*%(p2-p1)%*%y,
+ t(y)%*%(p3-p2)%*%y,
+ t(y)%*%(p4-p3)%*%y,
+ t(y)%*%(I-p4)%*%y,
+ t(y)%*%(I-p1)%*%y)
>
> Source=c(
+ "Diet|1",
+ "Drug|1,Diet",
+ "Diet x Drug|1,Diet,Drug",
+ "Error",
+ "C. Total")
```

for diet + drug

diet + drug + interaction

SSTo

SSE

R Code and Output for Two-Factor ANOVA

```
> data.frame(Source,SumOfSquares)
                     Source SumOfSquares
1                    Diet|1     76.00333
2              Drug|1,Diet     14.82000
3 Diet x Drug|1,Diet,Drug     12.28667
4                     Error      7.36000
5                  C. Total    110.47000
```

R Code and Output for Two-Factor ANOVA

```
> o=lm(weightgain~diet+drug+diet:drug,data=d)
>
> anova(o)
Analysis of Variance Table

Response: weightgain
          Df Sum Sq Mean Sq F value    Pr(>F)
diet       1 76.003  76.003 61.9592 0.0002226 ***
drug       2 14.820   7.410  6.0408 0.0365383 *
diet:drug  2 12.287   6.143  5.0082 0.0525735 .
Residuals  6  7.360   1.227
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

sequential SS from the previous slide

Strong diet effect

$= \frac{6.143}{1.227}$

keep small sample size in mind

What do the F-tests in this ANOVA table test?

$H_0: \mu_{cp} = 0 \quad$ vs. $\quad H_a: \mu_{cp} \neq 0$

Recall the null hypothesis for F_j is true if and only if

$$\boldsymbol{\beta}^\top \boldsymbol{X}^\top (\boldsymbol{P}_{j+1} - \boldsymbol{P}_j) \boldsymbol{X}\boldsymbol{\beta} = 0.$$

We have the following equivalent conditions

$$\begin{aligned}
\boldsymbol{\beta}^\top \boldsymbol{X}^\top (\boldsymbol{P}_{j+1} - \boldsymbol{P}_j) \boldsymbol{X}\boldsymbol{\beta} = 0 \iff & \boldsymbol{\beta}^\top \boldsymbol{X}^\top (\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)^\top (\boldsymbol{P}_{j+1} - \boldsymbol{P}_j) \boldsymbol{X}\boldsymbol{\beta} = 0 \\
\iff & \| (\boldsymbol{P}_{j+1} - \boldsymbol{P}_j) \boldsymbol{X}\boldsymbol{\beta} \|^2 = 0 \\
\iff & (\boldsymbol{P}_{j+1} - \boldsymbol{P}_j) \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0} \\
\iff & \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{0},
\end{aligned}$$

where $\boldsymbol{C}$ is any full-row-rank matrix with the same row space as $(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{X}$.

What do the F-tests in this ANOVA table test?

Let's take a look at $\boxed{(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{X}}$ for each test in the ANOVA table.

When computing $(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{X}$, we can use any model matrix $\boldsymbol{X}$ that specifies one unrestricted treatment mean for each of the six treatments.

The entries in any rows of $(\boldsymbol{P}_{j+1} - \boldsymbol{P}_j)\boldsymbol{X}$ are coefficients defining linear combinations of the elements of the parameter vector $\boldsymbol{\beta}$ that corresponds to the chosen model matrix $\boldsymbol{X}$.

Our Choice for $\boldsymbol{X}$ and $\boldsymbol{\beta}$

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix}$$

ANOVA Diet Test

main diet effect ⇒ diet has 2 levels ⇒ df = 1

```
> x=x4
> fractions((p2-p1)%*%x)
   b1:a1 b2:a1 b3:a1 b1:a2 b2:a2 b3:a2
1    1/6   1/6   1/6  -1/6  -1/6  -1/6
2    1/6   1/6   1/6  -1/6  -1/6  -1/6
3    1/6   1/6   1/6  -1/6  -1/6  -1/6
4    1/6   1/6   1/6  -1/6  -1/6  -1/6
5    1/6   1/6   1/6  -1/6  -1/6  -1/6
6    1/6   1/6   1/6  -1/6  -1/6  -1/6
7   -1/6  -1/6  -1/6   1/6   1/6   1/6
8   -1/6  -1/6  -1/6   1/6   1/6   1/6
9   -1/6  -1/6  -1/6   1/6   1/6   1/6
10  -1/6  -1/6  -1/6   1/6   1/6   1/6
11  -1/6  -1/6  -1/6   1/6   1/6   1/6
12  -1/6  -1/6  -1/6   1/6   1/6   1/6
```

C row rank = 1

$*1$

$*(-1)$

ANOVA Diet Test

$$(\boldsymbol{P}_2 - \boldsymbol{P}_1)\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0} \iff \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{0},$$

where

$$\boldsymbol{C}\boldsymbol{\beta} = \begin{bmatrix} \underbrace{\frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3}}_{\bar{\mu}_{1.}} & - & \underbrace{-\frac{1}{3} \quad -\frac{1}{3} \quad -\frac{1}{3}}_{\bar{\mu}_{2.}} \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix} = \bar{\mu}_{1\cdot} - \bar{\mu}_{2\cdot}.$$

drug has 3 levels ⇓

C needs to have a row rank = 2

```
> fractions((p3-p2)%*%x)
   b1:a1 b2:a1 b3:a1 b1:a2 b2:a2 b3:a2
1    1/3  -1/6  -1/6   1/3  -1/6  -1/6
2    1/3  -1/6  -1/6   1/3  -1/6  -1/6
3   -1/6   1/3  -1/6  -1/6   1/3  -1/6
4   -1/6   1/3  -1/6  -1/6   1/3  -1/6
5   -1/6  -1/6   1/3  -1/6  -1/6   1/3
6   -1/6  -1/6   1/3  -1/6  -1/6   1/3
7    1/3  -1/6  -1/6   1/3  -1/6  -1/6
8    1/3  -1/6  -1/6   1/3  -1/6  -1/6
9   -1/6   1/3  -1/6  -1/6   1/3  -1/6
10  -1/6   1/3  -1/6  -1/6   1/3  -1/6
11  -1/6  -1/6   1/3  -1/6  -1/6   1/3
12  -1/6  -1/6   1/3  -1/6  -1/6   1/3
```

ANOVA Drug Test

```
> p3p2x=(p3-p2)%*%x
>
> fractions(p3p2x[1,]-p3p2x[3,])
b1:a1 b2:a1 b3:a1 b1:a2 b2:a2 b3:a2
  1/2  -1/2     0   1/2  -1/2     0
>
> fractions(p3p2x[1,]-p3p2x[5,])
b1:a1 b2:a1 b3:a1 b1:a2 b2:a2 b3:a2
  1/2     0  -1/2   1/2     0  -1/2
```

drug 1 vs. drug 2

drug 1 vs. drug 3

$$(\boldsymbol{P}_3 - \boldsymbol{P}_2)\boldsymbol{X\beta} = \boldsymbol{0} \Longleftrightarrow \boldsymbol{C\beta} = \boldsymbol{0},$$

where

$$\begin{aligned}\boldsymbol{C\beta} &= \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{2} & 0 & -\frac{1}{2} & \frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix} \\ &= \begin{bmatrix} \bar{\mu}_{\cdot 1} - \bar{\mu}_{\cdot 2} \\ \bar{\mu}_{\cdot 1} - \bar{\mu}_{\cdot 3} \end{bmatrix}. \end{aligned}$$

} comparing marginal means associated with drug

ANOVA Test for Diet × Drug Interactions

need 2 independent rows
lin.

```
> fractions((p4-p3)%*%x)
   b1:a1 b2:a1 b3:a1 b1:a2 b2:a2 b3:a2
1   1/3  -1/6  -1/6  -1/3   1/6   1/6
2   1/3  -1/6  -1/6  -1/3   1/6   1/6
3  -1/6   1/3  -1/6   1/6  -1/3   1/6
4  -1/6   1/3  -1/6   1/6  -1/3   1/6
5  -1/6  -1/6   1/3   1/6   1/6  -1/3
6  -1/6  -1/6   1/3   1/6   1/6  -1/3
7  -1/3   1/6   1/6   1/3  -1/6  -1/6
8  -1/3   1/6   1/6   1/3  -1/6  -1/6
9   1/6  -1/3   1/6  -1/6   1/3  -1/6
10  1/6  -1/3   1/6  -1/6   1/3  -1/6
11  1/6   1/6  -1/3  -1/6  -1/6   1/3
12  1/6   1/6  -1/3  -1/6  -1/6   1/3
```

ANOVA Test for Diet × Drug Interactions

```
> p4p3x=(p4-p3)%*%x
>
> fractions(2*(p4p3x[1,]-p4p3x[3,]))
b1:a1 b2:a1 b3:a1 b1:a2 b2:a2 b3:a2
    1    -1     0    -1     1     0
>
> fractions(2*(p4p3x[1,]-p4p3x[5,]))
b1:a1 b2:a1 b3:a1 b1:a2 b2:a2 b3:a2
    1     0    -1    -1     0     1
>
```


$$(\boldsymbol{P}_4 - \boldsymbol{P}_3)\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0} \iff \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{0},$$

where

$$\begin{aligned}\boldsymbol{C}\boldsymbol{\beta} &= \begin{bmatrix} 1 & -1 & 0 & -1 & 1 & 0 \\ 1 & 0 & -1 & -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix} \\ &= \begin{bmatrix} \mu_{11} - \mu_{12} - \mu_{21} + \mu_{22} \\ \mu_{11} - \mu_{13} - \mu_{21} + \mu_{23} \end{bmatrix}.\end{aligned}$$

The diet-drug experiment is *balanced* in the sense that every treatment (defined by a diet-drug combination) has the same number of experimental units.

Each experimental unit provided a single response measurement (weight gain), so the resulting diet-drug dataset is *balanced* in the sense that each treatment has the same number of independent, constant variance observations of the response.

Due to this balance, the tests for diets, drugs, and diets $\times$ drugs in the ANOVA table turn out to be exactly the same as the tests for diet main effect, drug main effects, and diet $\times$ drug interactions we discussed previously as tests of $\boldsymbol{C\beta} = \mathbf{0}$.