end lecture 16

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

2-26-25

where

$$\begin{aligned} \boldsymbol{C\beta} &= \begin{bmatrix} \frac{2}{5} & \frac{3}{5} & -\frac{4}{5} & -\frac{1}{5} \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix} \\ &= \left(\frac{2}{5}\mu_{11} + \frac{3}{5}\mu_{12}\right) - \left(\frac{4}{5}\mu_{21} + \frac{1}{5}\mu_{22}\right). \end{aligned}$$

weighted average reflects the sample sizes in the cells

Null for Time|1 ANOVA test:

$$\frac{2}{5}\mu_{11} + \frac{3}{5}\mu_{12} = \frac{4}{5}\mu_{21} + \frac{1}{5}\mu_{22}$$

the marginal means

Null for Time main effect test:

$$\frac{1}{2}\mu_{11} + \frac{1}{2}\mu_{12} = \frac{1}{2}\mu_{21} + \frac{1}{2}\mu_{22}$$

i.e.

$$\boxed{\bar{\mu}_{1\cdot}} = \boxed{\bar{\mu}_{2\cdot}}$$

LS means

giving same weight to each cell mean when calculating

A Closer Look at the Time$|1$ ANOVA Test

The ANOVA Time$|1$ test is comparing the averages for the two storage times, ignoring storage temperature.

	Storage Temperature		
Storage Time	20°	30°	NOT LSMEANS
3 months	3 5	11 13 15	$\frac{2}{5}\widehat{\mu}_{11} + \frac{3}{5}\widehat{\mu}_{12}$
6 months	5 6 6 7	16	$\frac{4}{5}\widehat{\mu}_{21} + \frac{1}{5}\widehat{\mu}_{22}$

A Closer Look at the Time$|1$ ANOVA Test

The ANOVA Time$|1$ test is comparing the averages for the two storage times, ignoring storage temperature.

Storage Time	Storage Temperature 20°	30°	NOT LSMEANS
3 months	3 5	11 13 15	$\frac{2}{5}\left(\frac{3+5}{2}\right)+\frac{3}{5}\left(\frac{11+13+15}{3}\right)$
6 months	5 6 6 7	16	$\frac{4}{5}\left(\frac{5+6+6+7}{4}\right)+\frac{1}{5}\left(\frac{16}{1}\right)$

The ANOVA Time|1 test is comparing the averages for the two storage times, ignoring storage temperature.

Storage Time	Storage Temperature 20°	30°	NOT LSMEANS
3 months	3 5	11 13 15	$\left(\frac{3+5+11+13+15}{5}\right) = \boxed{9.4}$
6 months	5 6 6 7	16	$\left(\frac{5+6+6+7+16}{5}\right) = \boxed{8.0}$

increase in y (loss)

decrease in loss

these values do not reflect the actual behavior of what we see in the data

The Test for Time Main Effects

The test for time main effects is based on LSMEANS.

Storage Time	Storage Temperature 20°	30°	$\boxed{\text{LSMEANS}}$
3 months	3 5	11 13 15	$\frac{1}{2}\widehat{\mu}_{11} + \frac{1}{2}\widehat{\mu}_{12}$
6 months	5 6 6 7	16	$\frac{1}{2}\widehat{\mu}_{21} + \frac{1}{2}\widehat{\mu}_{22}$

The Test for Time Main Effects

The test for time main effects is based on LSMEANS.

Storage Time	Storage Temperature 20°	30°	LSMEANS
3 months	3 5	11 13 15	$\frac{1}{2}\left(\frac{3+5}{2}\right)+\frac{1}{2}\left(\frac{11+13+15}{3}\right)$
6 months	5 6 6 7	16	$\frac{1}{2}\left(\frac{5+6+6+7}{4}\right)+\frac{1}{2}\left(\frac{16}{1}\right)$

The Test for Time Main Effects

The test for time main effects is based on a comparison of LSMEANS.

Storage Time	Storage Temperature 20°	30°	LSMEANS
3 months	3 5	11 13 15	$\frac{1}{2}4 + \frac{1}{2}13 = \underline{8.5}$
6 months	5 6 6 7	16	$\frac{1}{2}6 + \frac{1}{2}16 = \underline{11.0}$

marginal means reflect what we see in the data as storage time increases!

```
> fractions((p3-p2)%*%x)
   b20:a3 b30:a3 b20:a6 b30:a6
1    9/25  -9/25   6/25  -6/25
2    9/25  -9/25   6/25  -6/25
3   -6/25   6/25  -4/25   4/25
4   -6/25   6/25  -4/25   4/25
5   -6/25   6/25  -4/25   4/25
6    3/25  -3/25   2/25  -2/25
7    3/25  -3/25   2/25  -2/25
8    3/25  -3/25   2/25  -2/25
9    3/25  -3/25   2/25  -2/25
10 -12/25  12/25  -8/25   8/25

> fractions((25/15)*(p3-p2)%*%x)[1,]
b20:a3 b30:a3 b20:a6 b30:a6
   3/5   -3/5    2/5   -2/5
```

Temp|1,Time ANOVA Test

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

where

$$\begin{aligned}\boldsymbol{C\beta} &= \begin{bmatrix} \frac{3}{5} & -\frac{3}{5} & \frac{2}{5} & -\frac{2}{5} \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix} \\ &= \left(\frac{3}{5}\mu_{11} + \frac{2}{5}\mu_{21}\right) - \left(\frac{3}{5}\mu_{12} + \frac{2}{5}\mu_{22}\right).\end{aligned}$$

weighted average

This is not the test for a storage temperature main effect.

Time×Temp|1,Time,Temp ANOVA Test

```
> fractions((p4-p3)%*%x)
   b20:a3 b30:a3 b20:a6 b30:a6
1    6/25  -6/25  -6/25   6/25
2    6/25  -6/25  -6/25   6/25
3   -4/25   4/25   4/25  -4/25
4   -4/25   4/25   4/25  -4/25
5   -4/25   4/25   4/25  -4/25
6   -3/25   3/25   3/25  -3/25
7   -3/25   3/25   3/25  -3/25
8   -3/25   3/25   3/25  -3/25
9   -3/25   3/25   3/25  -3/25
10  12/25 -12/25 -12/25  12/25

> fractions((25/6)*(p4-p3)%*%x)[1,]
b20:a3 b30:a3 b20:a6 b30:a6
     1     -1     -1      1
```

test for interaction will be the same (yield same results) regardless of balance/imbalance in our data

bc: - highest order term in our model
- it enters the model last

Time$\times$Temp$|1$,Time,Temp ANOVA Test

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

where

$$\begin{aligned}\boldsymbol{C}\boldsymbol{\beta} &= [1 \quad -1 \quad -1 \quad 1]\begin{bmatrix}\mu_{11}\\ \mu_{12}\\ \mu_{21}\\ \mu_{22}\end{bmatrix}\\ &= \mu_{11} - \mu_{12} - \mu_{21} + \mu_{22}.\end{aligned}$$

This is the test for Time $\times$ Temp interaction.

We could consider a different sequence of progressively more complex models for the response mean that lead up to our full cell means model.

enter temp before entering time

1. $E(y_{ijk}) = \mu$
2. $E(y_{ijk}) = \mu + \beta_j$
3. $E(y_{ijk}) = \mu + \alpha_i + \beta_j$
4. $E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij} \Longleftrightarrow E(y_{ijk}) = \mu_{ij}$


```
> 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
```

```
> x2=cbind(x1,model.matrix(~0+b))
> x2
       b20 b30
1   1    1    0
2   1    1    0
3   1    0    1
4   1    0    1
5   1    0    1
6   1    1    0
7   1    1    0
8   1    1    0
9   1    1    0
10  1    0    1
```

```
> x3=cbind(x2,model.matrix(~0+a))
> x3
       b20 b30 a3 a6
1   1    1   0  1  0
2   1    1   0  1  0
3   1    0   1  1  0
4   1    0   1  1  0
5   1    0   1  1  0
6   1    1   0  0  1
7   1    1   0  0  1
8   1    1   0  0  1
9   1    1   0  0  1
10  1    0   1  0  1
```

```
> x4=model.matrix(~0+b:a)
> x4
   b20:a3 b30:a3 b20:a6 b30:a6
1       1      0      0      0
2       1      0      0      0
3       0      1      0      0
4       0      1      0      0
5       0      1      0      0
6       0      0      1      0
7       0      0      1      0
8       0      0      1      0
9       0      0      1      0
10      0      0      0      1
```

```
> 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(1,10))
```

} these are different from before due to the change in the order of temp & time!

ANOVA Table

look different from before

Source	Sum of Squares	DF
Temp$|$1	$\boldsymbol{y}^\top(\boldsymbol{P}_2 - \boldsymbol{P}_1)\boldsymbol{y}$	$2 - 1 = 1$
Time$|$1, Temp	$\boldsymbol{y}^\top(\boldsymbol{P}_3 - \boldsymbol{P}_2)\boldsymbol{y}$	$3 - 2 = 1$
Temp $\times$ Time$|$1, Temp, Time	$\boldsymbol{y}^\top(\boldsymbol{P}_4 - \boldsymbol{P}_3)\boldsymbol{y}$	$4 - 3 = 1$
Error	$\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_4)\boldsymbol{y}$	$10 - 4 = 6$
C. Total	$\boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{y}$	$10 - 1 = 9$

```
> 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(
+ "Temp|1",
+ "Time|1,Temp",
+ "Temp x Time|1,Temp,Time",
+ "Error",
+ "C. Total")
```

```
> data.frame(Source,SumOfSquares)
                    Source SumOfSquares
1                   Temp|1    170.01667
2              Time|1,Temp     11.60333
3 Temp x Time|1,Temp,Time      0.48000
4                    Error     12.00000
5                 C. Total    194.10000
>
> anova(lm(y~temp+time+temp:time,data=d))

Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
temp       1 170.017 170.017 85.0083 9.185e-05 ***
time       1  11.603  11.603  5.8017   0.05267 .
temp:time  1   0.480   0.480  0.2400   0.64160
Residuals  6  12.000   2.000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

evidence in favor of a "time effect" looks stronger

Still not the test for the time main effect!

```
> x=x4
> fractions((p2-p1)%*%x)
   b20:a3 b30:a3 b20:a6 b30:a6
1    2/15  -3/10   4/15  -1/10
2    2/15  -3/10   4/15  -1/10
3    -1/5   9/20   -2/5   3/20
4    -1/5   9/20   -2/5   3/20
5    -1/5   9/20   -2/5   3/20
6    2/15  -3/10   4/15  -1/10
7    2/15  -3/10   4/15  -1/10
8    2/15  -3/10   4/15  -1/10
9    2/15  -3/10   4/15  -1/10
10   -1/5   9/20   -2/5   3/20

> fractions((30/12)*(p2-p1)%*%x)[1,]
b20:a3 b30:a3 b20:a6 b30:a6
   1/3   -3/4    2/3   -1/4
```

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

where

$$\begin{aligned}\boldsymbol{C\beta} &= \begin{bmatrix}\frac{1}{3} & -\frac{3}{4} & \frac{2}{3} & -\frac{1}{4}\end{bmatrix}\begin{bmatrix}\mu_{11}\\ \mu_{12}\\ \mu_{21}\\ \mu_{22}\end{bmatrix}\\ &= \left(\frac{1}{3}\mu_{11} + \frac{2}{3}\mu_{21}\right) - \left(\frac{3}{4}\mu_{12} + \frac{1}{4}\mu_{22}\right).\end{aligned}$$

Time|1,Temp ANOVA Test

```
> fractions((p3-p2)%*%x)
   b20:a3  b30:a3  b20:a6  b30:a6
1    32/75    6/25  -32/75   -6/25
2    32/75    6/25  -32/75   -6/25
3     4/25   9/100   -4/25  -9/100
4     4/25   9/100   -4/25  -9/100
5     4/25   9/100   -4/25  -9/100
6   -16/75   -3/25   16/75    3/25
7   -16/75   -3/25   16/75    3/25
8   -16/75   -3/25   16/75    3/25
9   -16/75   -3/25   16/75    3/25
10  -12/25 -27/100   12/25  27/100

> fractions((3/2)*(p3-p2)%*%x)[1,]
b20:a3 b30:a3 b20:a6 b30:a6
 16/25   9/25 -16/25  -9/25
```

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

where

$$\begin{aligned} \boldsymbol{C}\boldsymbol{\beta} &= \begin{bmatrix} \frac{16}{25} & \frac{9}{25} & -\frac{16}{25} & -\frac{9}{25} \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix} \\ &= \left(\frac{16}{25}\mu_{11} + \frac{9}{25}\mu_{12}\right) - \left(\frac{16}{25}\mu_{21} + \frac{9}{25}\mu_{22}\right). \end{aligned}$$

Temp×Time|1,Temp,Time ANOVA Test

```
> fractions((p4-p3)%*%x)
   b20:a3 b30:a3 b20:a6 b30:a6
1    6/25  -6/25  -6/25   6/25
2    6/25  -6/25  -6/25   6/25
3   -4/25   4/25   4/25  -4/25
4   -4/25   4/25   4/25  -4/25
5   -4/25   4/25   4/25  -4/25
6   -3/25   3/25   3/25  -3/25
7   -3/25   3/25   3/25  -3/25
8   -3/25   3/25   3/25  -3/25
9   -3/25   3/25   3/25  -3/25
10  12/25 -12/25 -12/25  12/25

> fractions((25/6)*(p4-p3)%*%x)[1,]
b20:a3 b30:a3 b20:a6 b30:a6
     1     -1     -1      1
```

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

where

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

```
> 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)
+ }
```

```
> o=lm(y~0+temp:time)
>
> #Test for time main effect
>
> C=matrix(c(
+ .5,.5,-.5,-.5
+ ),nrow=1,byrow=T)
>
> test(o,C)
$Fstat
[1] 6

$pvalue
[1] 0.04982526
```

Some weak evidence in favor of a time main effect

```
> #ANOVA Test for time|1
>
> C=matrix(c(
+ 2/5,3/5,-4/5,-1/5
+ ),nrow=1,byrow=T)
>
> test(o,C)
$Fstat
[1] 2.45

$pvalue
[1] 0.1685623
```

time is entered first

```
> #ANOVA Test for time|1,temp
>
> C=matrix(c(
+ 16/25,9/25,-16/25,-9/25
+ ),nrow=1,byrow=T)
>
> test(o,C)
$Fstat
[1] 5.801667

$pvalue
[1] 0.05266955
```

time is entered second

```
> #Test for temp main effect
>
> C=matrix(c(
+ .5,-.5,.5,-.5
+ ),nrow=1,byrow=T)
>
> test(o,C)
$Fstat
[1] 86.64

$pvalue
[1] 8.704602e-05
```

```
> #ANOVA Test for temp|1
>
> C=matrix(c(
+ 1/3,-3/4,2/3,-1/4
+ ),nrow=1,byrow=T)
>
> test(o,C)
$Fstat
[1] 85.00833

$pvalue
[1] 9.185462e-05
```

```
> #ANOVA Test for temp|1,time
>
> C=matrix(c(
+ 3/5,-3/5,2/5,-2/5
+ ),nrow=1,byrow=T)
>
> test(o,C)
$Fstat
[1] 88.36

$pvalue
[1] 8.233372e-05
```

```
> #Test for interactions
>
> C=matrix(c(
+ 1,-1,-1,1
+ ),nrow=1,byrow=T)
>
> test(o,C)
$Fstat
[1] 0.24

$pvalue
[1] 0.6416021
```

consistent

throughout

reduction in SSE due to factor A

Source	Type I (Sequential)	Type II (account for all terms that do not involve the factor under consideration)	Type III (given all other terms in the model)
A	$SS(A\|1)$ intercept	$SS(A\|1,B)$	$SS(A\|\boxed{1,B,AB})$
B	$SS(B\|1,A)$	$SS(B\|1,A)$	$SS(B\|1,A,AB)$
AB	$SS(AB\|1,A,B)$	$SS(AB\|1,A,B)$	$SS(AB\|1,A,B)$
Error	SSE	SSE	SSE
C. Total	$SSTotal$	?	?

Different Types of Sums of Squares for Three Factors

Type I	Type II
$SS(A\|1)$	$SS(A\|1, B, C, BC)$
$SS(B\|1, A)$	$SS(B\|1, A, C, AC)$
$SS(C\|1, A, B)$	$SS(C\|1, A, B, AB)$
$SS(AB\|1, A, B, C)$	$SS(AB\|1, A, B, C, AC, BC)$
$SS(AC\|1, A, B, C, AB)$	$SS(AC\|1, A, B, C, AB, BC)$
$SS(BC\|1, A, B, C, AB, AC)$	$SS(BC\|1, A, B, C, AB, AC)$
$SS(ABC\|1, A, B, C, AB, AC, BC)$	$SS(ABC\|1, A, B, C, AB, AC, BC)$
SSE	SSE
$SSTotal$	?

no interaction involving B: AB
BC

unlike Type III SS, we do not account for ABC interaction

Different Types of Sums of Squares for Three Factors

Type III

$SS(A\|1, B, C, AB, AC, BC, ABC)$
$SS(B\|1, A, C, AB, AC, BC, ABC)$
$SS(C\|1, A, B, AB, AC, BC, ABC)$
$SS(AB\|1, A, B, C, AC, BC, ABC)$
$SS(AC\|1, A, B, C, AB, BC, ABC)$
$SS(BC\|1, A, B, C, AB, AC, ABC)$
$SS(ABC\|1, A, B, C, AB, AC, BC)$
SSE

?

Sums of Squares for Balanced Data

For balanced data, the three types of sums of squares are identical: Type I = Type II = Type III.

This equality in not obvious (at least to most normal humans), but it is true. We will not attempt to prove this in 510.

The ANOVA F-tests in *the* ANOVA table can be used to test for factor main effects and interactions.

Sums of Squares for Unbalanced Data

For unbalanced data, the types of sums of squares differ.

Type I sums of squares always add to the total sum of squares, even when data are unbalanced.

Type II and III sums of squares do not add to anything special when data are unbalanced.

The ANOVA F-tests in the Type III ANOVA table can be used to test for factor main effects and interactions.

Type I and II ANOVA F-tests do not, in general, test for factor main effects or interactions (except for the F-test for the highest-order interactions, which is the same for all three types).

```
proc glm;
  class time temp;
  model y=time temp time*temp / ss1 ss2 ss3;
run;

The GLM Procedure

   Class Level Information

Class           Levels    Values
time                 2    3 6
temp                 2    20 30

Number of Observations Read          10
Number of Observations Used          10
```

```
Dependent Variable: y

                                        Sum of
Source                      DF         Squares     Mean Square    F Value    Pr > F

Model                        3     182.1000000      60.7000000      30.35    0.0005

Error                        6      12.0000000       2.0000000

Corrected Total              9     194.1000000


R-Square     Coeff Var      Root MSE        y Mean

0.938176      16.25533      1.414214      8.700000
```

Source	DF	Type I SS	Mean Square	F Value	Pr > F
time	1	4.9000000	4.9000000	2.45	$\boxed{0.1686}$
temp	1	176.7200000	176.7200000	88.36	<.0001
time*temp	1	0.4800000	0.4800000	0.24	0.6416

Source	DF	Type II SS	Mean Square	F Value	Pr > F
time	1	11.6033333	11.6033333	5.80	$\boxed{0.0527}$
temp	1	176.7200000	176.7200000	88.36	<.0001
time*temp	1	0.4800000	0.4800000	0.24	0.6416

Source	DF	Type III SS	Mean Square	F Value	Pr > F
time	1	12.0000000	12.0000000	6.00	$\boxed{0.0498}$
temp	1	173.2800000	173.2800000	86.64	<.0001
time*temp	1	0.4800000	0.4800000	0.24	0.6416

end lecture 17
2-28-25

Type IV Sums of Squares

In addition to computing Type I, II, and III sums of squares, SAS can compute Type IV sums of squares.

Type IV sums of squares are only relevant for factorial designs with missing cells.

When cells are missing, I recommend determining the linear combinations of the estimable cell means that are of scientific interest, and then conducting the corresponding tests as tests of $\mathrm{H}_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}$.