STARTED : Mon 3/28/22 (Week 10, Lecture 25)

& note: we started on pg 23 of the H.O. @ ≈ 35 min mark.

HANDOUT # 8

CRD WITH FACTORIAL TREATMENT STRUCTURE - SPECIAL SITUATIONS -

I. Augmented Factorial Experiment:

1. Factorial Experiment in Which a Control (Standard) Is Not Crossed with Factors
2. Experiment is Often Incorrectly Analyzed
3. Example

II. Factorial Experiment with Missing Treatments

1. Some of the Factor Combinations Are Not Observed in Experiment
2. Which Sum of Squares is Correct?
3. Using Contrasts to Test Incomplete Hypotheses for Interactions-Main Effects
4. Example

- **Supplemental Reading: Design & ANOVA Book** - Section 7.6.3

START Thu section used 3/30 (week 10, lecture 26) at ≈ 29 min mark.

I. CR Factorial Experiment Augmented With a Control

Often experiments involving factors with one factor having a qualitative levels and a second factor having b quantitative levels with one of the levels being a 0 **level**, that is, a standard, control, or placebo level, will result in an inappropriate analysis. This comes about when the treatments consisting of the cross of the levels of the qualitative factor with the 0 level of the quantitative factor does not yield unique treatments. Thus, instead of having $t = ab$ unique treatments, there are only $t = a(b-1)+1$ treatments. Consider the following examples.

- **Example 1:** There are three new drugs (D_1, D_2, D_3) proposed for the treatment of *Spring Fever*. The drugs are to be tested on 600 high school seniors at four dose levels each: $0, 10, 20, 30$ mg. Fifty students will be randomly assigned to each of the 12 treatments:

  $D_1 - 0, \quad D_2 - 0, \quad D_3 - 0; \quad D_1 - 10, \quad D_2 - 10, \quad D_3 - 10;$

  $D_1 - 20, \quad D_2 - 20, \quad D_3 - 20; \quad D_1 - 30, \quad D_2 - 30, \quad D_3 - 30.$

  The effectiveness of the drug-dose combinations are then measured on each of the 600 students. The experiment was then analyzed as a CRD with a 3×4 factorial treatment structure with $r = 50$ reps per treatment.

  Problems with the analysis? The treatments $D_1 - 0, D_2 - 0, D_3 - 0$ are identical, not 3 unique treatments. Thus, there are only $t = (3)(3)+1$ unique treatments in the experiment with 50 reps for 9 of the treatments and 150 reps for the Placebo (0 dose level).

a=4

- **Example 2:** A marketing firm has four new home cleaning products (P_1, P_2, P_3, P_4) that the firm will be promoting for a large corporation through direct mail. The marketing firm is interested in evaluating several approaches for introducing new products to the consumer. The firm will offer a number of incentives to induce the consumer to purchase the product. The incentives consist of varying amounts of a free sample of the product $(0, 1, 5, 10, 15)$ ounces of the product. A package containing either $0, 1, 5, 10, 15$ ounces of one of the four products is sent to 100 households. The number of households purchasing each of the products was recorded. Thus, it would appear that we have a CRD with a 4×5 factorial treatment and 100 reps per treatment.

b=5

  Problems with the analysis? No, in this case, the treatments $P_1 - 0, P_2 - 0, P_3 - 0, P_4 - 0$ are not identical. These four treatments are evaluating the difference in the ability of marketing each of the products without any incentive.

- **Example 3:** A researcher is evaluating a new chemical for the control of a disease that infects grasses used in household yards. The chemical will be evaluated at 5 application levels (0, 10, 20, 30, 40) pounds per 1000 square feet on three types of grasses (G_1, G_2, G_3). Each of the specified amounts of the chemical will be applied to 7 plots planted with one of the three types of grass which have been artificially infected with the disease. The effectiveness of the various application levels is measured for each of the three types of grasses. The researcher evaluates the experiment as a CRD with a 3×5 factorial treatment structure with $r = 7$ reps per treatment.

  Problems with the analysis? Are the three treatments consisting of a 0 amount of chemical the same for all three grasses?

  This is a valid factorial design because the natural resistance of the grass to the disease may be different for the three types of grass.

  G_1 at dose 0; G_2 at dose 0; G_3 at dose 0; are three distinct treatments. Essentially three Controls.

- **Example 4:** An animal scientist is studying the effect of alcohol on the brain in pregnant sheep. Female sheep are given a prescribed amount of alcohol during one or both of two periods: Period 1-During the first 50 days of pregnancy; Period 2-During the 100-150 days of pregnancy. Each sheep is examined in one of two regions of their brain (RA or RB). The examination is such that only one region can be examined on each sheep. There are 20 sheep in a Control group which do not receive the alcohol treatment but there brains are examined. There are 10 sheep randomly assigned to the eight treatment groups Alcohol(Yes or No), Alcohol Period (1-50 or 100-150), Brain Region (RA, RB).

  Problems with the analysis?

  This is a valid 2x2x2 factorial experiment with 10 reps for each of the 8 treatments:

Treatment	1	2	3	4	5	6	7	8
Alcohol - P1	Y	Y	N	N	Y	Y	N	N
Alcohol - P2	N	N	Y	Y	Y	Y	N	N
Region	RA	RB	RA	RB	RA	RB	RA	RB

General Approach to Analysis

Suppose we have a CRD with two factors:

1. Factor F_1 having a quantitative levels with a "0" level which is designated as a Control, Standard or Placebo

2. Factor F_2 having b qualitative levels

3. Suppose the treatments consisting of combining the 0 level of F_1 with the b levels of F_2 are identical. Thus, there are only $t = (a-1)b + 1$ unique treatments

4. Randomly assign n_1 EU's to the Control Treatment and n_2 EU's to the remaining $(a-1)b$ treatments.

5. This is then a CRD experiment with a $(a-1) \times b$ factorial treatment structure with an additional Control Treatment. There are n_1 reps of the control and n_2 reps of the remaining $(a-1)b$ treatments.

The analysis of the above experiment can be done using the effects model using the following steps:

Steps to Analyze Augmented Factorial Experiment

- Step 1: Analysis of a Quantitative Levels (Ignore Qualitative Factor F_2)
  (a) Fit the Model: $y_{ij} = \mu + \tau_i + e_{ij}$
    - All of the data is used in fitting this model.
  (b) This yields SS_{F_1} with $df = a - 1$.

- Step 2: Analysis of Factorial Structure (Ignore Control Treatment)
  (a) Fit the Model: $y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + e_{ijk}$
    - Do not use the data from the control treatment in fitting this model.
  (b) This yields SS_{F_2} with $df = b - 1$ and
  $SS_{F_1*F_2}$ with $df = (b-1)(a-2)$.

- Step 3: Analysis as a CR Design with $t = (a-1)b + 1$ Treatments (Ignore Factorial Structure)
  (a) Fit the Cell Means Model: $y_{ij} = \mu_i + e_{ij}$
    - All of the data is used in fitting this model.
  (b) This yields SSE with $df_E = DF_{TOT} - DF_{F_1} - DF_{F_2} - DF_{F_1*F_2}$

The analysis is then summarized in an ANOVA table:

Source	df	Sum of Squares	Mean Square	F-Statistic	p-value
F_1	a-1	SS_{F_1}	$\frac{SS_{F_1}}{(a-1)}$	$\frac{MS_{F_1}}{MSE}$	$1 - G_{a-1,df_E}(\frac{MS_{F_1}}{MSE})$
F_2	b-1	SS_{F_2}	$\frac{SS_{F_2}}{(b-1)}$	$\frac{MS_{F_2}}{MSE}$	$1 - G_{b-1,df_E}(\frac{MS_{F_2}}{MSE})$
$F_1 * F_2$	(a-2)(b-1)	$SS_{F_1*F_2}$	$\frac{SS_{F_1*F_2}}{(a-2)(b-1)}$	$\frac{MS_{F_1*F_2}}{MSE}$	$1 - G_{(a-2)(b-1),df_E}(\frac{MS_{F_1*F_2}}{MSE})$
Error	df_E	SSE	MSE		
Total	$n_1 + n_2(a-1)b - 1$				

A total of three models were fit to obtain the above AOV table.

Example: Controlling Eelworms Using Soil Fumigants (Cochran-Cox 1957)

The object of the experiment was to measure the effectiveness of 4 soil fumigants in reducing the numbers of eelworms in the soil. The fumigants were chlorodinitrobenzene (CN), carbon disulphide jelly (CS), cymag (CM), and seekay (CK). Each fumigant was tested at two doses (D1, D2). A control treatment (C) of no fumigant was also used in the experiment. Thus, we have a 4x2 factorial experiment, four fumigants each at two dose levels, along with a control treatment resulting in a total of nine treatments. Four sections of land were located and divided into 12 plots with 4 plots randomly assigned to the control and 1 plot randomly assigned to each of the eight treatments. The extra plots for the treatment enabled the researchers to have an accurate standard against which to measure the performance of the fumigants.

The response variable is the number of eelworm cysts per 400 grams of soil after application of the treatment and the crop being harvested. We will only consider the second count in our analysis. The data is given in Table 1.

Table 1. Numbers of Eelworm Cysts Per 400 Grams of Soil

	Treatment											
	CONTROL				CK		CN		CM		CS	
Dose	0	0	0	0	D1	D2	D1	D2	D1	D2	D1	D2
	466	219	421	708	256	283	398	304	386	379	194	372
	590	137	356	212	236	142	176	199	332	308	221	166
	505	363	563	338	268	408	415	365	222	561	433	311
	352	254	106	268	132	292	454	298	114	92	80	281

Analysis:

This type of experiment has often been incorrectly analyzed as a 3x4 factorial experiment:

4 Fumigants: CK, CM, CN, CS

3 Dose Levels: 0, D1, D2

Thus, we would have $t = (3)(4) = 12$ treatments. However, there are only 9 different treatments: a 2x4 factorial plus a 'Control' treatment. The 0 Dose Level of the 4 Fumigants is identical for all four Fumigants, and thus is a single treatment, not 4 distinct treatments. The analysis must be divided into three separate analyzes in order to obtain an assessment of the factorial structure along with a comparison of the 8 treatments with the control.

Incorrect Analysis I

```
OPTION LS=75 PS=55 NOCENTER NODATE;
TITLE 'QUALITATIVE FACTOR WITH A QUANITATIVE FACTOR WITH A 0 LEVEL';
DATA AUGFAC;
INPUT  DOSE $ FUM $ Y @@;
LABEL Y='EELWORM COUNT';
CARDS;
0 CK   466 0 CN   219 0 CM   421 0 CS   708
1 CK   256 2 CK   283 1 CN   398 2 CN   304
1 CM   386 2 CM   379 1 CS   194 2 CS   372
0 CK   590 0 CN   137 0 CM   356 0 CS   212
1 CK   236 2 CK   142 1 CN   176 2 CN   199
1 CM   332 2 CM   308 1 CS   221 2 CS   166
0 CK   505 0 CN   363 0 CM   563 0 CS   338
1 CK   268 2 CK   408 1 CN   415 2 CN   365
1 CM   222 2 CM   561 1 CS   433 2 CS   311
0 CK   352 0 CN   254 0 CM   106 0 CS   268
1 CK   132 2 CK   292 1 CN   454 2 CN   298
1 CM   114 2 CM    92 1 CS    80 2 CS   281
RUN;
PROC PRINT;RUN;
*INCORRECT ANALYSIS;
PROC GLM;
 CLASS DOSE FUM;
 MODEL   Y = DOSE FUM DOSE*FUM;
RUN;
  LSMEANS DOSE FUM DOSE*FUM/STDERR;

RUN;
```

OUTPUT FROM SAS:

```
QUALITATIVE FACTOR WITH A QUANITATIVE FACTOR WITH A 0 LEVEL            107

The GLM Procedure

      Class Level Information

Class         Levels    Values

DOSE               3    0 1 2

FUM                4    CK CM CN CS


Number of Observations Read          48
Number of Observations Used          48
```

INCORRECT ANALYSIS I

The GLM Procedure

Dependent Variable: Y EELWORM COUNT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	245810.6667	22346.4242	1.20	0.3188
Error	36	667668.0000	18546.3333		
Corrected Total	47	913478.6667			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DOSE	2	78650.5417	39325.2708	2.12	0.1347
FUM	3	7938.0000	2646.0000	0.14	0.9337
DOSE*FUM	6	159222.1250	26537.0208	1.43	0.2300

Least Squares Means

DOSE	Y LSMEAN	Standard Error	Pr > |t|
0	366.125000	34.046231	<.0001
1	269.812500	34.046231	<.0001
2	297.562500	34.046231	<.0001

FUM	Y LSMEAN	Standard Error	Pr > |t|
CK	327.500000	39.313201	<.0001
CM	320.000000	39.313201	<.0001
CN	298.500000	39.313201	<.0001
CS	298.666667	39.313201	<.0001

DOSE	FUM	Y LSMEAN	Standard Error	Pr > |t|
0	CK	478.250000	68.092462	<.0001
0	CM	361.500000	68.092462	<.0001
0	CN	243.250000	68.092462	0.0010
0	CS	381.500000	68.092462	<.0001
1	CK	223.000000	68.092462	0.0023
1	CM	263.500000	68.092462	0.0004
1	CN	360.750000	68.092462	<.0001
1	CS	232.000000	68.092462	0.0016
2	CK	281.250000	68.092462	0.0002
2	CM	335.000000	68.092462	<.0001
2	CN	291.500000	68.092462	0.0001
2	CS	282.500000	68.092462	0.0002

Incorrect Analysis II

```
OPTION LS=75 PS=55 NOCENTER NODATE;
TITLE 'QUALITATIVE FACTOR WITH A QUANITATIVE FACTOR WITH A 0 LEVEL';
DATA AUGFAC;
INPUT  DOSE $ FUM $ TRT $ Y @@;
LABEL Y='EELWORM COUNT';
CARDS;
0 C  C0    466 0 C C0    219 0 C C0   421 0 C C0   708
1 CK CKD1  256 2 CK CKD2  283 1 CN CND1 398 2 CN CND2 304
1 CM CMD1  386 2 CM CMD2 379 1 CS CSD1 194 2 CS CSD2 372
0 C C0   590 0 C C0    137 0 C C0   356 0 C C0   212
1 CK CKD1  236 2 CK CKD2  142 1 CN CND1 176 2 CN CND2 199
1 CM CMD1  332 2 CM CMD2 308 1 CS CSD1 221 2 CS CSD2 166
0 C C0   505 0 C C0    363 0 C C0   563 0 C C0   338
1 CK CKD1  268 2 CK CKD2  408 1 CN CND1 415 2 CN CND2 365
1 CM CMD1  222 2 CM CMD2 561 1 CS CSD1 433 2 CS CSD2 311
0 C C0   352 0 C C0    254 0 C C0   106 0 C C0   268
1 CK CKD1  132 2 CK CKD2  292 1 CN CND1 454 2 CN CND2 298
1 CM CMD1  114 2 CM CMD2  92 1 CS CSD1  80 2 CS CSD2 281
RUN;

PROC GLM;
 CLASS DOSE FUM;
 MODEL   Y = DOSE FUM DOSE*FUM;
RUN;
  LSMEANS DOSE FUM DOSE*FUM/STDERR;
RUN;
```

OUTPUT FROM SAS:

```
INCORRECT ANALYSIS OF A QUALITATIVE FACTOR WITH A QUANITATIVE FACTOR WITH A 0 LEVEL

The GLM Procedure

       Class Level Information

Class         Levels    Values

DOSE               3    0 1 2

FUM                5    C CK CM CN CS

Number of Observations Read          48
Number of Observations Used          48
```

Dependent Variable: Y EELWORM COUNT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	134098.4167	16762.3021	0.84	0.5746
Error	39	779380.2500	19984.1090		
Corrected Total	47	913478.6667			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DOSE	1	6160.50000	6160.50000	0.31	0.5819
FUM	3	29906.12500	9968.70833	0.50	0.6853
DOSE*FUM	3	25541.75000	8513.91667	0.43	0.7354

Least Squares Means

DOSE	Y LSMEAN
0	Non-est
1	Non-est
2	Non-est

FUM	Y LSMEAN
C	Non-est
CK	Non-est
CM	Non-est
CN	Non-est
CS	Non-est

DOSE	FUM	Y LSMEAN	Standard Error	Pr > |t|
0	C	366.125000	35.341290	<.0001
1	CK	223.000000	70.682581	0.0031
1	CM	263.500000	70.682581	0.0006
1	CN	360.750000	70.682581	<.0001
1	CS	232.000000	70.682581	0.0022
2	CK	281.250000	70.682581	0.0003
2	CM	335.000000	70.682581	<.0001
2	CN	291.500000	70.682581	0.0002
2	CS	282.500000	70.682581	0.0003

Correct Analysis: Using SAS Effects Models

Step 1: Analysis of 3 Dose Levels (Ignore Different Fumigants)

(a) Model: $y_{ij} = \mu + \tau_i + e_{ij}$

(b) This yields $SS_{DOSE} = 78650.5417$ with DF=3-1=2.

Step 2: Analysis of Factorial Structure (Ignore Control Treatment)

(a) Model: $y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + e_{ijk}$

(b) This yields $SS_{FUM} = 29906.1250$ with df=4-1=3

$SS_{FUM*DOSE} = 25541.7500$ with df=(4-1)(3-1-1)=3.

Step 3: Analysis as a CR Design with 9 Treatments (Ignore Factorial Structure)

(a) Model: $y_{ij} = \mu + \alpha_i + e_{ij}$

(b) This yields $SS_E = 779380.2500$ with

$df = DF_{TOT} - DF_{DOSE} - DF_{FUM} - DF_{FUM*DOSE} = 47 - 2 - 3 - 3 = 39$

The analysis is then summarized in an ANOVA table:

Source	df	Sum of Squares	Mean Square	F-Statistic	p-value
Dose	2	78650.5417	39325.27	1.97	0.1534
Fumigant	3	29906.1250	9968.71	0.50	0.6845
Dose*Fumigant	3	25541.7500	8513.92	0.43	0.7354
Error	39	779380.2500	19984.11		
Total	47				

```
* augmentedFac.sas
ods html; ods graphics on;
OPTION LS=75 PS=55 NOCENTER NODATE;
TITLE 'QUALITATIVE FACTOR WITH A QUANITATIVE FACTOR WITH A 0 LEVEL';
DATA AUGFAC;
INPUT  DOSE $ FUM $ TRT $ Y @@;
LABEL Y='EELWORM COUNT';
CARDS;
0 C  C0    466 0 C C0    219 0 C C0   421 0 C C0   708
1 CK CKD1  256 2 CK CKD2  283 1 CN CND1 398 2 CN CND2 304
1 CM CMD1  386 2 CM CMD2 379 1 CS CSD1 194 2 CS CSD2 372
0 C C0   590 0 C C0    137 0 C C0   356 0 C C0   212
1 CK CKD1  236 2 CK CKD2  142 1 CN CND1 176 2 CN CND2 199
1 CM CMD1  332 2 CM CMD2 308 1 CS CSD1 221 2 CS CSD2 166
0 C C0   505 0 C C0    363 0 C C0   563 0 C C0   338
1 CK CKD1  268 2 CK CKD2  408 1 CN CND1 415 2 CN CND2 365
1 CM CMD1  222 2 CM CMD2 561 1 CS CSD1 433 2 CS CSD2 311
0 C C0   352 0 C C0    254 0 C C0   106 0 C C0   268
1 CK CKD1  132 2 CK CKD2  292 1 CN CND1 454 2 CN CND2 298
1 CM CMD1  114 2 CM CMD2  92 1 CS CSD1  80 2 CS CSD2 281
RUN;
TITLE STEP 1: ANALYSIS WITH 2 DOSES PLUS CONTROL;
PROC GLM;
CLASS DOSE;
MODEL Y = DOSE;
RUN;
TITLE STEP 2: ANALYSIS AS A 2X4 FACTORIAL WITHOUT CONTROL;
DATA SUBSET; SET AUGFAC; IF DOSE > 0;
IF FUM='CK' THEN FUMPLOT = 1;IF FUM='CM' THEN FUMPLOT = 2;
IF FUM='CN' THEN FUMPLOT =3;IF FUM='CS' THEN FUMPLOT =4;
RUN;
PROC PRINT;
RUN;
PROC GLM DATA=SUBSET;
 CLASS DOSE FUM;
 MODEL   Y = DOSE FUM DOSE*FUM;
RUN;
TITLE  FUM BY DOSE INTERACTION PLOT;
proc glimmix data=SUBSET;
CLASS DOSE FUM;
 MODEL   Y = DOSE FUM DOSE*FUM;
lsmeans DOSE*FUM /  plot = meanplot cl;
RUN;
TITLE STEP 3: ANALYSIS AS A CR WITH 9 TREATMENTS;
PROC GLM DATA=AUGFAC;
  CLASS TRT;
  MODEL   Y  = TRT;
  LSMEANS TRT/STDERR;
  MEANS TRT/DUNNETT( 'C0');
RUN;
ods graphics off; ods html close;
```

STEP 1: ANALYSIS WITH 2 DOSES PLUS CONTROL

Class	Levels	Values
DOSE	3	0 1 2

Number of observations in data set = 48

Dependent Variable: Y EELWORM COUNT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	78650.5417	39325.2708	2.12	0.1319
Error	45	834828.1250	18551.7361		
Corrected Total	47	913478.6667			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DOSE	2	78650.5417	39325.2708	2.12	0.1319

STEP 2: ANALYSIS AS A 2X4 FACTORIAL WITHOUT CONTROL

Class	Levels	Values
DOSE	2	1 2
FUM	4	CK CM CN CS

Number of observations in data set = 32

Dependent Variable: Y EELWORM COUNT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	61608.3750	8801.1964	0.60	0.7503
Error	24	352482.5000	14686.7708		
Corrected Total	31	414090.8750			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DOSE	1	6160.5000	6160.5000	0.42	0.5234
FUM	3	29906.1250	9968.7083	0.68	0.5736
DOSE*FUM	3	25541.7500	8513.9167	0.58	0.6340

STEP 3: ANALYSIS AS A CR WITH 9 TREATMENTS

Class	Levels	Values
TRT	9	C0 CKD1 CKD2 CMD1 CMD2 CND1 CND2 CSD1 CSD2

Number of observations in data set = 48

Dependent Variable: Y EELWORM COUNT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	134098.417	16762.302	0.84	0.5746
Error	39	779380.250	19984.109		
Corrected Total	47	913478.667			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	8	134098.417	16762.302	0.84	0.5746

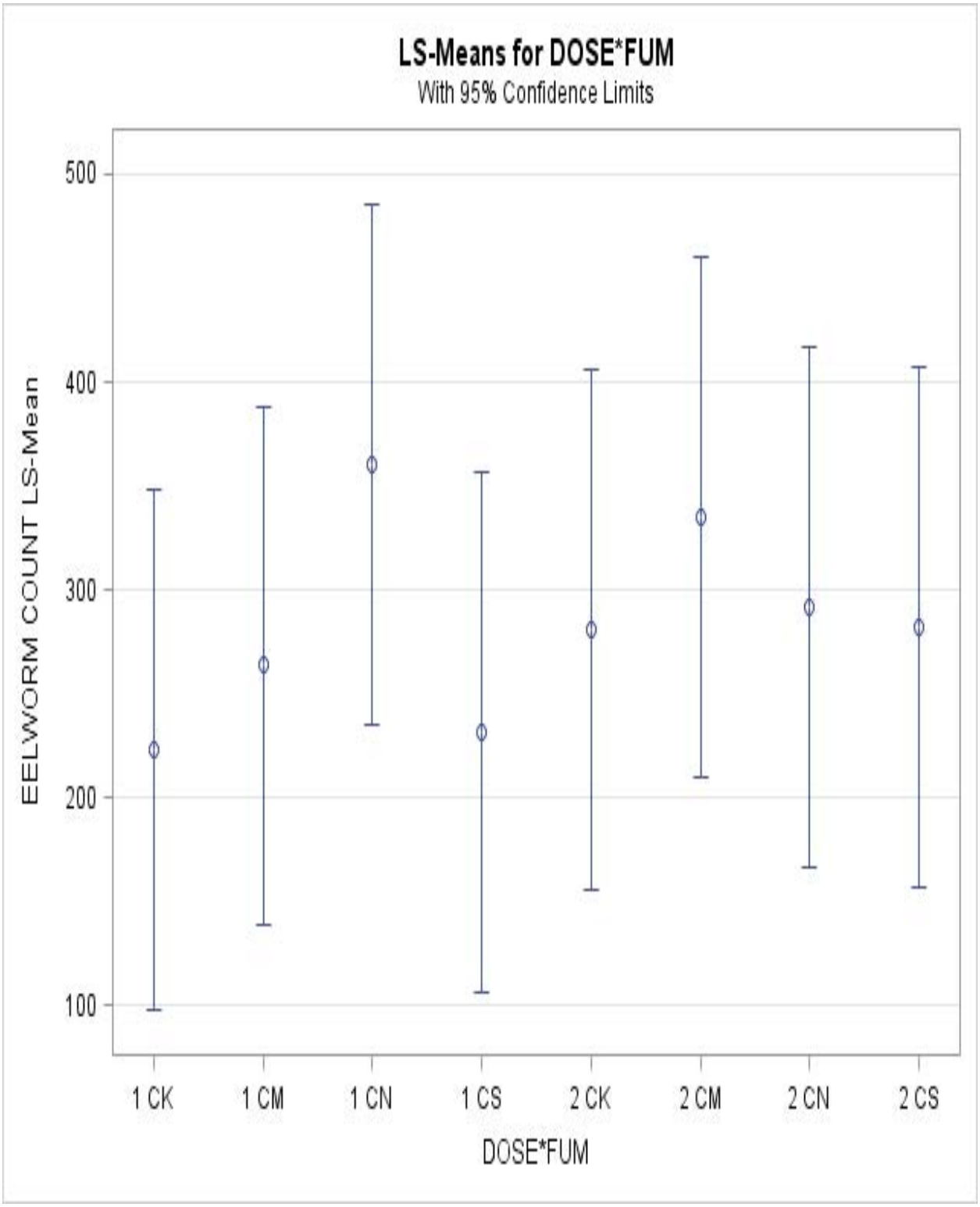
Least Squares Means

TRT	Y LSMEAN	Std Err LSMEAN	Pr > |T| H0:LSMEAN=0
C0	366.125000	35.341290	0.0001
CKD1	223.000000	70.682581	0.0031
CKD2	281.250000	70.682581	0.0003
CMD1	263.500000	70.682581	0.0006
CMD2	335.000000	70.682581	0.0001
CND1	360.750000	70.682581	0.0001
CND2	291.500000	70.682581	0.0002
CSD1	232.000000	70.682581	0.0022
CSD2	282.500000	70.682581	0.0003

Dunnett's T tests for variable: Y Alpha= 0.05 Confidence= 0.95 df= 39 MSE= 19984.11

Critical Value of Dunnett's T= 2.858

TRT Comparison	Simultaneous Lower Confidence Limit	Difference Between Means	Simultaneous Upper Confidence Limit
CND1 - C0	-231.25	-5.38	220.50
CMD2 - C0	-257.00	-31.13	194.75
CND2 - C0	-300.50	-74.63	151.25
CSD2 - C0	-309.50	-83.63	142.25
CKD2 - C0	-310.75	-84.88	141.00
CMD1 - C0	-328.50	-102.63	123.25
CSD1 - C0	-360.00	-134.13	91.75
CKD1 - C0	-369.00	-143.13	82.75

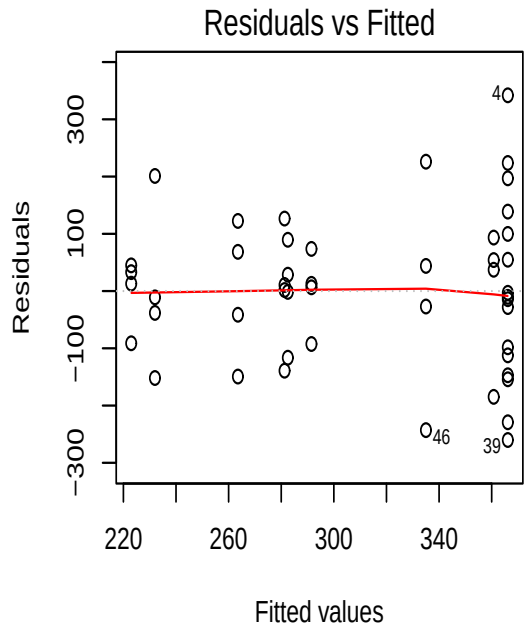

LS-Means for DOSE*FUM
With 95% Confidence Limits
EELWORM COUNT LS-Mean
500
400
300
200
100
1 CK
1 CM
1 CN
1 CS
2 CK
2 CM
2 CN
2 CS
DOSE*FUM

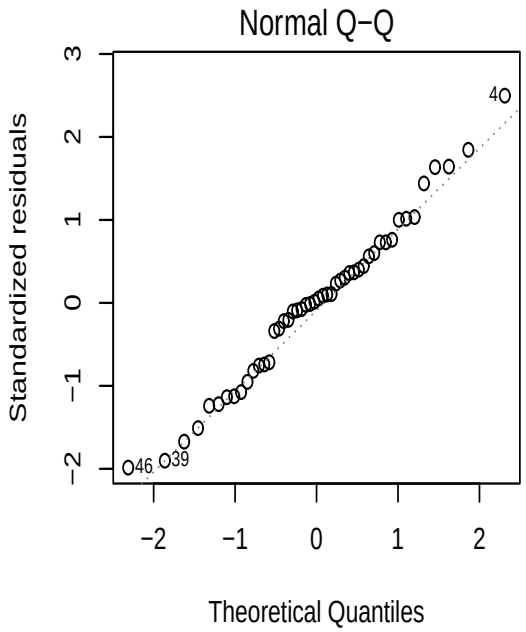
R Code for analyzing the Eelworm data - augmentedFac.R in eCampus

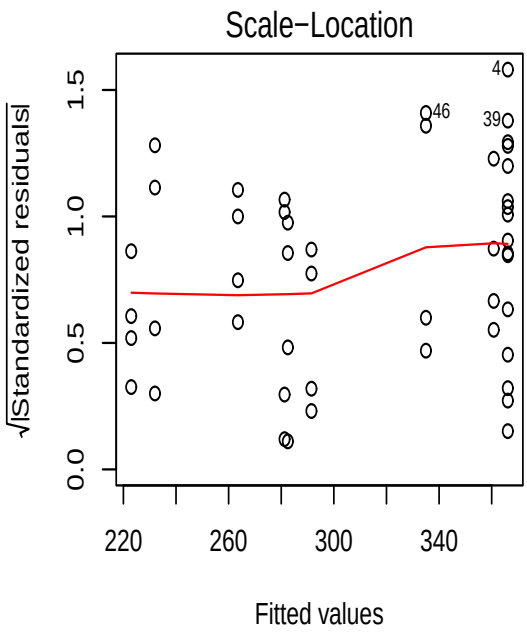
```
#augmentedFac.R

library("multcomp")
library("lsmeans")
options(contrasts = c("contr.sum", "contr.poly"))

ne = c(466,219,421,708,256,283,398,304,386,379,194,372,
       590,137,356,212,236,142,176,199,332,308,221,166,
       505,363,563,338,268,408,415,365,222,561,433,311,
       352,254,106,268,132,292,454,298,114, 92, 80,281)
dose = rep(c("0","0","0","0","D1","D2","D1","D2","D1","D2","D1","D2"),4)
fung = rep(c("C","C","C","C","CK","CK","CN","CN","CM","CM","CS","CS"),4)
dose  = factor(dose)
fung  = factor(fung)
datafac = cbind(ne,factor(dose),factor(fung))
datafac = data.frame(datafac)

#Model 1: Just Dose in Model

model1  = lm(ne~dose,data=datafac)
anova(model1)

#Output from R:
#Analysis of Variance Table
#Response: ne
#          Df Sum Sq Mean Sq F value Pr(>F)
#dose       2  78651   39325  2.1198 0.1319
#Residuals 45 834828   18552

#Model 2: Factorial model without control data

newo = c(256,283,398,304,386,379,194,372,
         236,142,176,199,332,308,221,166,
         268,408,415,365,222,561,433,311,
         132,292,454,298,114, 92, 80,281)
dosewo = rep(c("D1","D2","D1","D2","D1","D2","D1","D2"),4)
fungwo =  rep(c("CK","CK","CN","CN","CM","CM","CS","CS"),4)
datafacwo = data.frame(cbind(newo,factor(dosewo),factor(fungwo)))

model2  = lm(newo~dosewo  + fungwo  + dosewo:fungwo,data=datafacwo)
anova(model2)
```

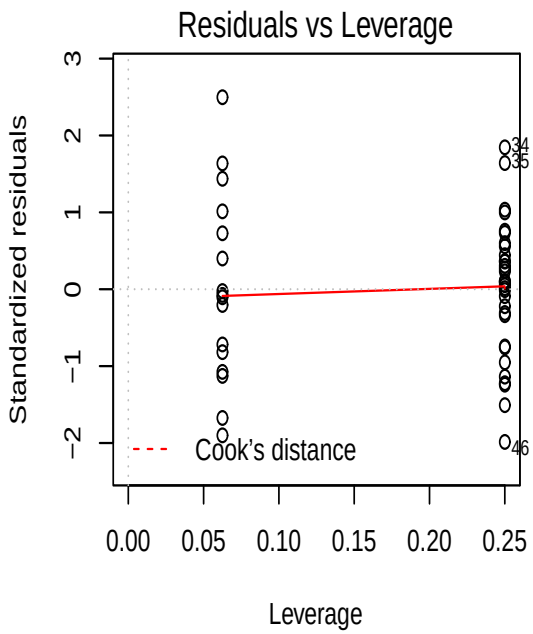
```
#Output from R:

#Analysis of Variance Table
#Response: newo
#               Df Sum Sq Mean Sq F value Pr(>F)
#dosewo          1   6161  6160.5  0.4195 0.5234
#fungwo          3  29906  9968.7  0.6788 0.5736
#dosewo:fungwo   3  25542  8513.9  0.5797 0.6340
#Residuals      24 352482 14686.8

#Model 3: Cell Means model

trt = rep(c("C0","C0","C0","C0","CKD1","CKD2","CND1","CND2","CMD1","CMD2","CSD1","CSD2"),4)
datatrt = data.frame(cbind(ne,factor(trt)))
trt = factor(trt)
model3  = lm(ne~trt,data=datatrt)
anova(model3)
aovtrt = aov(model3)


#Output from R:

#Analysis of Variance Table

#Response: ne
#           Df Sum Sq Mean Sq F value Pr(>F)
#trtfac      8 134098   16762  0.8388 0.5746
#Residuals 39 779380   19984


#Pairwise Comparisons of Treatment means:

lsmeans(aovtrt,~trt)
plot(lsmeans(aovtrt,~trt))

#Output from R:
#trtfac  lsmean       SE df  lower.CL upper.CL
# C0     366.125 35.34129 39 294.64049 437.6095
# CKD1   223.000 70.68258 39  80.03099 365.9690
# CKD2   281.250 70.68258 39 138.28099 424.2190
# CMD1   263.500 70.68258 39 120.53099 406.4690
# CMD2   335.000 70.68258 39 192.03099 477.9690
# CND1   360.750 70.68258 39 217.78099 503.7190
# CND2   291.500 70.68258 39 148.53099 434.4690
# CSD1   232.000 70.68258 39  89.03099 374.9690
# CSD2   282.500 70.68258 39 139.53099 425.4690
```

```
#Dunnett's comparison of treatments to Control
trt.mc = glht(aovtrt,linfct = mcp(trt="Dunnett"),alternative = "less")
summary(trt.mc)

#Output from R:


# Simultaneous Tests for General Linear Hypotheses

#Multiple Comparisons of Means: Dunnett Contrasts


#Fit: aov(formula = ne ~ trtfac)

#Linear Hypotheses:
#               Estimate Std. Error t value Pr(<t)
#CKD1 - C0 >= 0 -143.125     79.026  -1.811  0.232
#CKD2 - C0 >= 0  -84.875     79.026  -1.074  0.604
#CMD1 - C0 >= 0 -102.625     79.026  -1.299  0.480
#CMD2 - C0 >= 0  -31.125     79.026  -0.394  0.898
#CND1 - C0 >= 0   -5.375     79.026  -0.068  0.962
#CND2 - C0 >= 0  -74.625     79.026  -0.944  0.674
#CSD1 - C0 >= 0 -134.125     79.026  -1.697  0.279
#CSD2 - C0 >= 0  -83.625     79.026  -1.058  0.613
#pl(Adjusted p values reported -- single-step method)


#Residual Analysis

par(mfrow=c(2,2))
plot(model3)
```









STOP Wednesday 3/30/22 (week 10, before 26)

START Friday 4/1/22 (week 10, lecture 24)

Analysis Using Contrasts

An alternative analysis is done by examining various contrasts in the nine treatments. The analysis will partition the 8 degrees of freedom for treatment into a 'Control versus average of Fumigants' sum of squares with one degree of freedom, and the usual main effect and interaction sums of squares for the 2x4 factorial stucture. We can accomplish this partition by pooling the sums of squares for three orthogonal contrasts for the fumigant main effect, using a contrast to compare Dose=D1 to Dose =D2, and pooling the sums of squares for three orthogonal contrasts corresponding to the Dose*Fumigant interaction for the 2x4 factorial stucture.

Contrast	C	CKD1	CKD2	CND1	CND2	CMD1	CMD2	CSD1	CSD2	Interpretation
Main-Fumigant	0	-1	-1	-1	-1	1	1	1	1	CK,CN vs CM,CS
	0	-1	-1	1	1	0	0	0	0	CK vs CN
	0	0	0	0	0	-1	-1	1	1	CM vs CS
Main-Dose	0	-1	1	-1	1	-1	1	-1	1	D1 vs D3
Fum*Dose	0	1	-1	1	-1	-1	1	-1	1	- Row 1 * Row 4
	0	1	-1	-1	1	0	0	0	0	- Row 2 * Row 4
	0	0	0	0	0	1	-1	-1	1	- Row 3 * Row 4.

In addition we can use contrasts to compare the fumigants within each dose by pooling the sums of squares for orthogonal contrasts within each dose.

Contrast	C	CKD1	CKD2	CND1	CND2	CMD1	CMD2	CSD1	CSD2	Interpretation
Fum @ Dose 1	0	1	0	1	0	-1	0	-1	0	
	0	1	0	-1	0	0	0	0	0	
	0	0	0	0	0	1	0	-1	0	
Fum @ Dose 2	0	0	1	0	1	0	-1	0	-1	
	0	0	1	0	-1	0	0	0	0	
	0	0	0	0	0	0	1	0	-1	

Finally, we can compare the main effect for the three doses of application of the fumigants by pooling the sums of squares for orthogonal contrasts on the marginal means for the three doses:

Contrast	C	CKD1	CKD2	CND1	CND2	CMD1	CMD2	CSD1	CSD2	Interpretation
Control vs Trt	-8	1	1	1	1	1	1	1	1	
D1 vs D2	0	1	-1	1	-1	1	-1	1	-1	

```
*augFac_Contrasts.sas;

OPTION LS=75 PS=55 NOCENTER NODATE;
TITLE 'QUALITATIVE FACTOR WITH A QUANITATIVE FACTOR WITH A 0 LEVEL';
DATA AUGFAC;
INPUT  DOSE $ FUM $ TRT $ Y @@;
LABEL Y='EELWORM COUNT';
CARDS;
0 C  C0    466 0 C C0    219 0 C C0   421 0 C C0   708
1 CK CKD1  256 2 CK CKD2  283 1 CN CND1 398 2 CN CND2 304
1 CM CMD1  386 2 CM CMD2 379 1 CS CSD1 194 2 CS CSD2 372
0 C C0   590 0 C C0    137 0 C C0   356 0 C C0   212
1 CK CKD1  236 2 CK CKD2  142 1 CN CND1 176 2 CN CND2 199
1 CM CMD1  332 2 CM CMD2 308 1 CS CSD1 221 2 CS CSD2 166
0 C C0   505 0 C C0    363 0 C C0   563 0 C C0   338
1 CK CKD1  268 2 CK CKD2  408 1 CN CND1 415 2 CN CND2 365
1 CM CMD1  222 2 CM CMD2 561 1 CS CSD1 433 2 CS CSD2 311
0 C C0   352 0 C C0    254 0 C C0   106 0 C C0   268
1 CK CKD1  132 2 CK CKD2  292 1 CN CND1 454 2 CN CND2 298
1 CM CMD1  114 2 CM CMD2  92 1 CS CSD1  80 2 CS CSD2 281
RUN;


* ANALYSIS AS A CR WITH 9 TREATMENTS;

PROC GLM DATA=AUGFAC;
  CLASS TRT;
  MODEL   Y  = TRT;
  MEANS TRT;
  MEANS TRT/DUNNETT( 'C0');

 *                              TRT C CK1 CK2 CM1 CM2 CN1 CN2 CS1 CS2;
CONTRAST 'FUMIGANT MAIN EFFECT' TRT 0  -1  -1  -1  -1   1   1   1   1,
                                TRT 0  -1  -1   1   1   0   0   0   0,
                                TRT 0   0   0   0   0  -1  -1   1   1;

CONTRAST 'BETWEEN DOSE 1 & 2'   TRT 0  -1   1  -1   1  -1   1  -1   1;

CONTRAST 'FUMIGANT*DOSE'        TRT 0   1  -1   1  -1  -1   1  -1   1,
                                TRT 0   1  -1  -1   1   0   0   0   0,
                                TRT 0   0   0   0   0   1  -1  -1   1;

CONTRAST 'FUMIGANT @ DOSE 1'    TRT 0   1   0   1   0  -1   0  -1   0,
                                TRT 0   1   0  -1   0   0   0   0   0,
                                TRT 0   0   0   0   0   1   0  -1   0;

CONTRAST 'FUMIGANT @ DOSE 2'    TRT 0   0   1   0   1   0  -1   0  -1,
                                TRT 0   0   1   0  -1   0   0   0   0,
                                TRT 0   0   0   0   0   0   1   0  -1;

CONTRAST 'BETWEEN DOSES'        TRT -8  1   1   1   1   1   1   1   1,
                                TRT  0  1  -1   1  -1   1  -1   1  -1;
RUN;
```

```
Class    Levels    Values
TRT           9    C0 CKD1 CKD2 CMD1 CMD2 CND1 CND2 CSD1 CSD2
Number of observations in data set = 48

Dependent Variable: Y   EELWORM COUNT
                                   Sum of          Mean
Source                   DF       Squares        Square  F Value    Pr > F
Model                     8     134098.417     16762.302     0.84    0.5746
Error                    39     779380.250     19984.109
Corrected Total          47     913478.667

Source                   DF    Type III SS   Mean Square  F Value    Pr > F
TRT                       8     134098.417     16762.302     0.84    0.5746


Contrast                 DF    Contrast SS   Mean Square  F Value    Pr > F
FUMIGANT MAIN EFFECT      3     29906.1250     9968.7083     0.50    0.6853
BETWEEN DOSE 1 & 2        1      6160.5000     6160.5000     0.31    0.5819
FUMIGANT*DOSE             3     25541.7500     8513.9167     0.43    0.7354
FUMIGANT @ DOSE 1         3     47722.6875    15907.5625     0.80    0.5036
FUMIGANT @ DOSE 2         3      7725.1875     2575.0625     0.13    0.9424
BETWEEN DOSES             2     78650.5417    39325.2708     1.97    0.1534
```

Part II: Experiments with Unobserved Treatments

In a factorial experiment, if one of more of the treatments are not observed, $n_{ij} = 0$, we cannot use standard methods of analysis. One of the major problems is that the unobserved treatments will have means μ_{ij} which cannot be estimated with the given data. This will result in the main effects containing these treatment means to be non-estimable. Consider the following definition:

DEFINITION For the linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, consider estimating a linear combination or subset of the coefficients $\boldsymbol{\beta}$: $\mathbf{L}\boldsymbol{\beta}$. We state $\mathbf{L}\boldsymbol{\beta}$ is **Estimable** if and only if there exist a matrix $\mathbf{A}$ such that $E[\mathbf{AY}] = \mathbf{L}\boldsymbol{\beta}$.

That is, a linear combination of the population parameters, $\mathbf{L}\boldsymbol{\beta}$, is estimable if there exist a linear combination of the observed data which is an unbiased estimator of $\mathbf{L}\boldsymbol{\beta}$.

This is where the problem occurs with unobserved treatments, factor level combinations. Suppose we have two factors A and B with 3 and 2 levels, respectively. Suppose that one of the $t = 6$ treatments is not observed, for example the (A,B) = (1,1) treatment, so that we have no data associated with the treatment mean $\mu_{1,1}$. This will cause problems when we attempt to estimate main effects associated with both factor A and factor B. For example, the factor A main effect $\mu_{1.} - \mu_{2.}$ cannot be estimated because

$$\mu_{1.} - \mu_{2.} = (\mu_{1,1} + \mu_{1,2}) - (\mu_{2,1} + \mu_{2,2})$$

and we have no data from which we can estimate $\mu_{1,1}$.

Similarly, the factor B main effect $\mu_{.1} - \mu_{.2}$ cannot be estimated because

$$\mu_{.1} - \mu_{.2} = (\mu_{1,1} + \mu_{2,1} + \mu_{3,1}) - (\mu_{1,2} + \mu_{2,2} + \mu_{3,2})$$

and once again we have no data from which we can estimate $\mu_{1,1}$. However, we could estimate the factor A main effect $\mu_{2.} - \mu_{3.}$ because

$$\mu_{2.} - \mu_{3.} = (\mu_{2,1} + \mu_{2,2}) - (\mu_{3,1} + \mu_{3,2})$$

and we have data on all four of the treatment means represented in this contrast.

Also, we will be able to obtain only some of the least squares estimates of the marginal means:

$\mu_{1.}$ and $\mu_{.1}$ will not be estimable but $\mu_{2.}, \mu_{3.}, \mu_{.2}$ will be estimable.

The following example will illustrate the types of problems encountered by blindly running a SAS analysis of the data from an experiment having no observations on some of the factor combinations. We will also discuss alternative forms of analysis.

Example of CRD 3×3 Factorial Experiment - With Unobserved Treatments (From: *Analysis of Messy Data*, by Johnson and Millikin)

A bakery scientist wanted to study the effects of combining three different fats(F_1) with each of three surfactants(F_2) on the specific volume of bread loaves baked from doughs mixed from each of the nine treatment combinations. Four loaves were made from each of the nine treatment combinations. Unfortunately, one container of yeast turned out to be ineffective, and the data from the 15 loaves made with that yeast had to be removed from the analysis. The data is given below:

Specific Volumes from Baking Experiment

			LOAVE					
TRT	Fat	Surfacant	1	2	3	4	n_{ij}	$\bar{Y}_{ij.} = \hat{\mu}_{ij}$
1	1	1	6.7	4.3	5.7	*	3	5.57
2	1	2	7.1	*	5.9	5.6	3	6.20
3	1	3	*	*	*	*	0	*
4	2	1	*	5.9	7.4	7.1	3	6.80
5	2	2	*	*	*	*	0	*
6	2	3	6.4	5.1	6.2	6.3	4	6.00
7	3	1	7.1	5.9	*	*	2	6.50
8	3	2	7.3	6.6	8.1	6.8	4	7.00
9	3	3	*	7.5	9.1	*	2	8.30
TOTAL							21	6.58

The above experiment is a CRD with a 3×3 factorial treatment structure and 4 replications. However, a number of the replications are not observed. This results in several treatments having no observations in the experiment. Often experimental data such as the above is analyzed using computer software. The following analysis from SAS demonstrates the problems which result from such an analysis.

The effects model is used in the following analysis:

$$y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + e_{ijk} \quad \text{with} \quad i = 1, 2, 3; \quad j = 1, 2, 3$$

This was designed as an equally replicated experiment with $r = 4$, however, because of problems that arose during the experiment the number of actual observations per treatment are given below:

$$n_{11} = 3; \quad n_{12} = 3; \quad n_{13} = 0; \quad n_{21} = 3; \quad n_{22} = 0; \quad n_{23} = 4; \quad n_{31} = 2; \quad n_{32} = 4; \quad n_{33} = 2$$

Having some of the treatments unobserved would not be a problem in the Cell Means Model because we would just have 7 treatments with

$$n_1 = 3; \quad n_2 = 3; \quad n_3 = 3; \quad n_4 = 4; \quad n_5 = 2; \quad n_6 = 4; \quad n_7 = 2$$

When there are missing treatments in an experiment with a factorial treatment structure, there are a number of complications in doing an analysis using the Effects Model. These problems in conducting the analysis are not present in the Cell Means Model.

```
*On Dostat as  twofact_missing,effects.sas;
options ls=80 ps=58 nocenter nodate;
* This is Example 15.1 on page 192 in the
"Analysis of Messy Data, Vol I", by G. Milliken and D. Johnson;
*This experiment is a CR 3x3 factorial with 4 reps but there
is missing reps for some treatments and 0 reps for 2 treatments;
options pagesize=55 linesize=72;
data raw;
input trt fat surf fl1-fl4;
drop fl1-fl4;
sv=fl1;output;
sv=fl2;output;
sv=fl3;output;
sv=fl4;output;
cards;
1 1  1  6.7 4.3 5.7  .
2 1  2  7.1  .  5.9 5.6
3 1  3   .   .   .  .
4 2  1   .  5.9 7.4 7.1
5 2  2   .   .   .   .
6 2  3  6.4 5.1 6.2 6.3
7 3  1  7.1 5.9  .   .
8 3  2  7.3 6.6 8.1 6.8
9 3  3   .  7.5 9.1  .
;
title 'Analysis as a CR 3x3 factorial';
proc glm;
class fat surf;
model sv = fat surf fat*surf/ss1 ss2 ss3 ss4;
means fat surf fat*surf;
lsmeans fat surf fat*surf/stderr pdiff ;
RUN;
```

Analysis as a CR 3x3 factorial

Class	Levels	Values
fat	3	1 2 3
surf	3	1 2 3
Number of observations	36	

NOTE: Due to missing values, only 21 observations can be used in this analysis.

Dependent Variable: sv

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	12.47142857	2.07857143	2.95	0.0447
Error	14	9.86666667	0.70476190		
Corrected Total	20	22.33809524			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
fat	2	7.45261905	3.72630952	5.29	0.0195
surf	2	0.29722997	0.14861498	0.21	0.8124
fat*surf	2	4.72157956	2.36078978	3.35	0.0647

Source	DF	Type II SS	Mean Square	F Value	Pr > F
fat	2	6.47812282	3.23906141	4.60	0.0292
surf	2	0.29722997	0.14861498	0.21	0.8124
fat*surf	2	4.72157956	2.36078978	3.35	0.0647

Source	DF	Type III SS	Mean Square	F Value	Pr > F
fat	2	6.00174091	3.00087046	4.26	0.0359
surf	2	0.99963357	0.49981678	0.71	0.5089
fat*surf	2	4.72157956	2.36078978	3.35	0.0647

Source	DF	Type IV SS	Mean Square	F Value	Pr > F
fat	2*	3.87252033	1.93626016	2.75	0.0985
surf	2*	1.67022222	0.83511111	1.18	0.3346
fat*surf	2	4.72157956	2.36078978	3.35	0.0647

* NOTE: Other Type IV Testable Hypotheses exist which may yield different SS.

fat	N	Mean	Std Dev
1	6	5.88333333	0.97655858
2	7	6.34285714	0.75907212
3	8	7.30000000	0.97541200

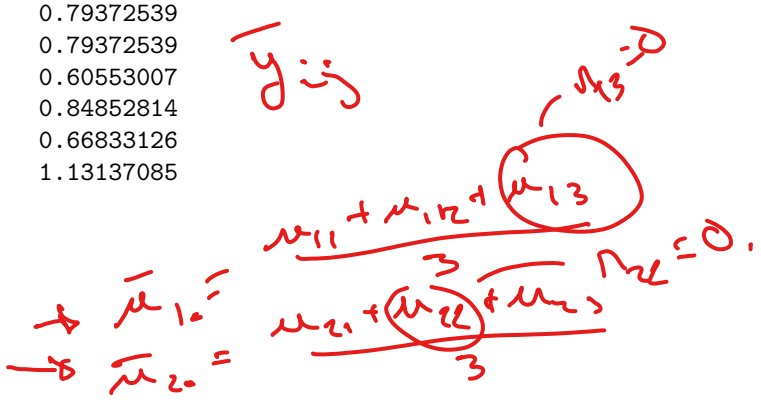
surf	N	Mean	Std Dev
1	8	6.26250000	1.02251441
2	7	6.77142857	0.84796676
3	6	6.76666667	1.37355985

fat	surf	N	Mean	Std Dev
1	1	3	5.56666667	1.20554275
1	2	3	6.20000000	0.79372539
2	1	3	6.80000000	0.79372539
2	3	4	6.00000000	0.60553007
3	1	2	6.50000000	0.84852814
3	2	4	7.20000000	0.66833126
3	3	2	8.30000000	1.13137085

Least Squares Means

fat	sv LSMEAN	Standard Error	Pr > |t|
1	Non-est	.	.
2	Non-est	.	.
3	7.33333333	0.31286355	<.0001

surf	sv LSMEAN	Standard Error	Pr > |t|
1	6.28888889	0.30225490	<.0001
2	Non-est	.	.
3	Non-est	.	.



fat	surf	sv LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
1	1	5.56666667	0.48468612	<.0001	1
1	2	6.20000000	0.48468612	<.0001	2
2	1	6.80000000	0.48468612	<.0001	3
2	3	6.00000000	0.41975049	<.0001	4
3	1	6.50000000	0.59361684	<.0001	5
3	2	7.20000000	0.41975049	<.0001	6
3	3	8.30000000	0.59361684	<.0001	7

Least Squares Means for effect fat*surf
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: sv

i/j	1	2	3	4	5	6	7
1		0.3712	0.0936	0.5102	0.2434	0.0232	0.0031
2	0.3712		0.3962	0.7597	0.7013	0.1412	0.0159
3	0.0936	0.3962		0.2326	0.7013	0.5428	0.0706
4	0.5102	0.7597	0.2326		0.5029	0.0628	0.0069
5	0.2434	0.7013	0.7013	0.5029		0.3520	0.0501
6	0.0232	0.1412	0.5428	0.0628	0.3520		0.1525
7	0.0031	0.0159	0.0706	0.0069	0.0501	0.1525	

STOP Monday 3/28/22 (week 10, Lecture 25)

START Wed 3/30/22 (Week 10, Lecture 26)

Four Sum of Squares from SAS

The SAS output has four types of Sum of Squares. When fitting a model with multiple effects, for example,

$y_{ij} = \mu + \alpha_{1i} + \alpha_{2i} + \alpha_{3i} + \alpha_{4i} + e_{ij}$, SAS uses the following notation

$R(\alpha_1|\alpha_2, \alpha_3, \alpha_4)$ = reduction in SSE due to adding the term α_1 to a model containing the terms $\alpha_2, \alpha_3, \alpha_4$

That is, fit two models

Model 1 : $y_{ij} = \mu + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + e_{ij}$ compute SSE_1 - Full Model

Model 2 : $y_{ij} = \mu + \alpha_2 + \alpha_3 + \alpha_4 + e_{ij}$ compute SSE_2 - Reduced Model

$R(\alpha_1|\alpha_2, \alpha_3, \alpha_4) = SSE_2 - SSE_1 \geq 0$

The four types of sum of squares will be defined below using the model for a three factor experiment:

$$y_{ijk} = \mu + \tau_i + \gamma_j + \delta_k + (\tau\gamma)_{ij} + (\tau\delta)_{ik} + (\gamma\delta)_{jk} + (\tau\gamma\delta)_{ijk} + e_{ijkl}$$

The above model is appropriate for a CRD with a Completely Crossed Factorial Treatment Structure allowing for all possible main effects and interactions. Thus,

1. Main Effects $F_1: \quad \tau_i, \qquad F_2: \quad \gamma_j, \qquad F_3: \quad \delta_j,$
2. Two-way Interaction Effects $F_1 * F_2: \quad (\tau\gamma)_{ij}, \qquad F_1 * F_3: \quad (\tau\delta)_{ik}, \qquad F_2 * F_3: \quad (\gamma\delta)_{jk},$
3. Three-way Interaction Effects $F_1 * F_2 * F_3: \quad (\tau\gamma\delta)_{ijk}$

Type I Sum of Squares: Sequential

Type I sum of squares for each of the terms in the model are computed in a sequential manner using the order in which the terms appear in the model statement.

Factor	Interpretation of Type I SS		
F_1	SS_{F_1}	$=$	$R(\tau \mid \mu)$
F_2	SS_{F_2}	$=$	$R(\gamma \mid \mu, \tau)$
F_3	SS_{F_3}	$=$	$R(\delta \mid \mu, \gamma, \tau)$
$F_1 * F_2$	$SS_{F_1*F_2}$	$=$	$R((\tau * \gamma) \mid \mu, \delta, \gamma, \tau)$
$F_1 * F_3$	$SS_{F_1*F_3}$	$=$	$R((\tau * \delta) \mid \mu, \delta, \gamma, \tau, (\tau * \gamma))$
$F_2 * F_3$	$SS_{F_2*F_3}$	$=$	$R((\gamma * \delta) \mid \mu, \delta, \gamma, \tau, (\tau * \delta), (\tau * \gamma))$
$F_1 * F_2 * F_3$	$SS_{F_1*F_2*F_3}$	$=$	$R((\tau * \gamma * \delta) \mid \mu, \delta, \gamma, \tau, (\gamma * \delta), (\tau * \delta), (\tau * \gamma))$

Type I Sum of Squares are generally useful only for regression models where a polynomial model is being built from a low degree of complexity to a higher degree of complexity. Type I Sum of Squares are also used in analyzing Hierarchial Nested Treatment Factors which will be discussed later.

A CRD experiment with a 3×3 factorial treatment structure and missing Treatment Combinations having sample sizes given below will be used to illustrate the hypotheses being tested by Type I SS:,

		F_2			
		1	2	3	$n_{i.}$
F_1	1	$n_{11} = 2$	$n_{12} = 0$	$n_{13} = 2$	4
	2	$n_{21} = 1$	$n_{22} = 1$	$n_{23} = 2$	4
	3	$n_{31} = 1$	$n_{32} = 1$	$n_{33} = 0$	2
	$n_{.j}$	$n_{.1} = 4$	$n_{.2} = 2$	$n_{.3} = 4$	10

The Type I SS test the following hypotheses:

Source of Variation	Hypotheses
F_1	$H_o : \mu_{11} + \mu_{13} - \mu_{31} - \mu_{32} = 0$ and $\mu_{21} + \mu_{22} + 2\mu_{23} - 2\mu_{31} - 2\mu_{32} = 0$
F_2	$H_o : 5\mu_{11} - 5\mu_{13} + 3\mu_{21} + \mu_{22} - 4\mu_{23} + \mu_{31} - \mu_{32} = 0$ and $\mu_{11} - \mu_{13} + 2\mu_{22} - 2\mu_{23} - \mu_{31} + \mu_{32} = 0$
$F_1 * F_2$	$H_o : \mu_{11} - \mu_{13} - \mu_{22} + \mu_{23} - \mu_{31} + \mu_{32} = 0$ and $\mu_{21} - \mu_{22} - \mu_{31} + \mu_{32} = 0$

Type II Sum of Squares: Adjusted for Same or Lower Order Factors

The Type II sum of squares corresponding to a term in the model is adjusted for every other effect in the model that is at the same or lower level.

Factor	Interpretation of Type II SS		
F_1	SS_{F_1}	$=$	$R(\tau \mid \ \mu, \gamma, \delta)$
F_2	SS_{F_2}	$=$	$R(\gamma \mid \ \mu, \tau, \delta)$
F_3	SS_{F_3}	$=$	$R(\delta \mid \ \mu, \tau, \gamma)$
$F_1 * F_2$	$SS_{F_1*F_2}$	$=$	$R((\tau * \gamma) \mid \ \mu, \delta, \gamma, \tau, (\tau * \delta), (\gamma * \delta))$
$F_1 * F_3$	$SS_{F_1*F_3}$	$=$	$R((\tau * \delta) \mid \ \mu, \delta, \gamma, \tau, (\tau * \gamma), (\gamma * \delta))$
$F_2 * F_3$	$SS_{F_2*F_3}$	$=$	$R((\gamma * \delta) \mid \ \mu, \delta, \gamma, \tau, (\tau * \delta), (\tau * \gamma))$
$F_1 * F_2 * F_3$	$SS_{F_1*F_2*F_3}$	$=$	$R((\tau * \gamma * \delta) \mid \ \mu, \delta, \gamma, \tau, (\gamma * \delta), (\tau * \delta), (\tau * \gamma))$

Type II Sum of Squares are used in the analysis of sample survey data in which the hypotheses being tested are weighted means where weights are estimates of the population weights.

For the CRD experiment introduced previously, the Type II SS test the following hypotheses:

Source of Variation	Hypotheses
FAT	$H_o: 2\mu_{11} + \mu_{13} + \mu_{22} - \mu_{23} - 2\mu_{31} - \mu_{32} = 0$ and $2\mu_{11} - 2\mu_{23} + 3\mu_{21} + 4\mu_{22} + 2\mu_{23} - 5\mu_{31} - 4\mu_{32} = 0$
$SURF$	$H_o: 5\mu_{11} - 5\mu_{13} + 3\mu_{21} + \mu_{22} - 4\mu_{23} + \mu_{31} - \mu_{32} = 0$ and $\mu_{11} - \mu_{13} + 2\mu_{22} - 2\mu_{23} - \mu_{31} + \mu_{32} = 0$
$FAT * SURF$	$H_o: \mu_{11} - \mu_{13} - \mu_{22} + \mu_{23} - \mu_{31} - \mu_{32} = 0$ and $\mu_{21} - \mu_{22} - \mu_{31} + \mu_{32} = 0$

Type III & IV Sum of Squares: Adjusted for All Other Factors

The Type III & IV sum of squares are identical for experiments in which $n_{ij} > 0$ for all treatments. The Sum of Squares corresponding to a term in the model is adjusted for ALL other effects in the model.

Factor	Interpretation of Type III & Type IV SS		
F_1	SS_{F_1}	$=$	$R(\tau \mid \ \mu, \gamma, \delta, (\tau * \gamma), (\tau * \delta), (\gamma * \delta), (\tau * \gamma * \delta))$
F_2	SS_{F_2}	$=$	$R(\gamma \mid \ \mu, \tau, \delta, (\tau * \gamma), (\tau * \delta), (\gamma * \delta), (\tau * \gamma * \delta))$
F_3	SS_{F_3}	$=$	$R(\delta \mid \ \mu, \tau, \gamma, (\tau * \gamma), (\tau * \delta), (\gamma * \delta), (\tau * \gamma * \delta))$
$F_1 * F_2$	$SS_{F_1 * F_2}$	$=$	$R((\tau * \gamma) \mid \ \mu, \tau, \gamma, \delta, (\tau * \delta), (\gamma * \delta), (\tau * \gamma * \delta))$
$F_1 * F_3$	$SS_{F_1 * F_3}$	$=$	$R((\tau * \delta) \mid \ \mu, \tau, \gamma, \delta, (\tau * \gamma), (\gamma * \delta), (\tau * \gamma * \delta))$
$F_2 * F_3$	$SS_{F_2 * F_3}$	$=$	$R((\gamma * \delta) \mid \ \mu, \tau, \gamma, \delta, (\tau * \gamma), (\tau * \delta), (\tau * \gamma * \delta))$
$F_1 * F_2 * F_3$	$SS_{F_1 * F_2 * F_3}$	$=$	$R((\tau * \gamma * \delta) \mid \ \mu, \tau, \gamma, \delta, (\tau * \gamma), (\tau * \delta), (\gamma * \delta))$

Type III & IV Sum of Squares are the mostly widely used in the analysis of experiments. They test the type of hypotheses of most interest to experimenters. When some of the treatments are not observed in the experiment, that is, $n_{ij} = 0$ for some treatments, Type IV Sum of Squares adjusts factor effects by averaging over one or more common levels of the other factor effects. In most cases, when some treatments are not observed, Type IV Sum of Squares are testing hypotheses which are most likely to have reasonable interpretations. However, as is true, for all four types of sum of squares, it is difficult to determine the actual hypotheses being tested.

For the CRD experiment described previously in which $n_{12} = 0$ and $n_{33} = 0$, the **Type III SS** test the following hypotheses:

Source of Variation	Hypotheses
F_1	$H_o : 2\mu_{11} + \mu_{13} + \mu_{22} - \mu_{23} - 2\mu_{31} - \mu_{32} = 0$ and $2\mu_{11} - 2\mu_{23} + 3\mu_{21} + 4\mu_{22} + 2\mu_{23} - 5\mu_{31} - 4\mu_{32} = 0$
F_2	$H_o : 7\mu_{11} - 7\mu_{13} + 6\mu_{21} + 2\mu_{22} - 8\mu_{23} + 2\mu_{31} - 2\mu_{32} = 0$ and $\mu_{11} - \mu_{13} + 2\mu_{22} - 2\mu_{23} - \mu_{31} + \mu_{32} = 0$
$F_1 * F_2$	$H_o : \mu_{11} - \mu_{13} - \mu_{22} + \mu_{23} - \mu_{31} - \mu_{32} = 0$ and $\mu_{21} - \mu_{22} - \mu_{31} + \mu_{32} = 0$

For the CRD experiment described previously in which $n_{12} = 0$ and $n_{33} = 0$, the **Type IV SS** test the following hypotheses:

Source of Variation	Hypotheses
F_1	$H_o : \frac{1}{2}(\mu_{11} + \mu_{13}) = \frac{1}{2}(\mu_{21} + \mu_{23})$ and $\mu_{11} = \mu_{31}$
F_2	$H_o : \frac{1}{2}(\mu_{11} + \mu_{21}) = \frac{1}{2}(\mu_{13} + \mu_{23})$ and $\mu_{22} = \mu_{23}$
$F_1 * F_2$	$H_o : \mu_{11} - \mu_{13} - \mu_{22} + \mu_{23} - \mu_{31} - \mu_{32} = 0$ and $\mu_{21} - \mu_{22} - \mu_{31} + \mu_{32} = 0$

There are many other possible Type IV hypotheses that can be generated. PROC GLM in SAS automatically generates a set of Type IV hypotheses. Thus, it is impossible to interpret the significance of the effects using the $p - value$ for the main and interaction effects because which set of hypotheses being tested are not displayed. The interpretation problem is shown in the SAS output with the displaying of the statement, **"OTHER TYPE IV TESTABLE HYPOTHESES EXIST WHICH MAY YIELD DIFFERENT SS"**.

APPROACH II:
CELL MEANS MODEL USING CONTRASTS TO TEST HYPOTHESES

The more appropriate methodology is to use a Cell Means Model and construct contrasts which are testing hypotheses that are directly of interest to the researcher.

Let Y_{ijk} =specific volume of the kth loave using ith level of Fat and jth level of Surfacant.

Model: $Y_{ijk} = \mu_{ij} + e_{ijk}$; for $i, j = 1, 2, 3$; $k = 1, \cdots, r_{ij}$

$SS_{TOT} = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{r_{ij}}[Y_{ijk} - \bar{Y}_{...}]^2 = 22.338095 \quad df_{TOT} = 21 - 1 = 20$

$SS_E = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{r_{ij}}[Y_{ijk} - \hat{\mu}_{ij}]^2 = 9.8666667 \quad df_E = N - t = 21 - 7 = 14$

$SS_{MODEL} = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{r_{ij}}[\hat{\mu}_{ij} - \bar{Y}_{...}]^2 = 12.471429 \quad df_M = t - 1 = 7 - 1 = 6$

We want to decompose SS_{MODEL} into terms which represent differences in the $t = 7$ treatments: "Fat Main Effect", "Surfacant Main Effect", and "FxS Interaction."

I. First test for overall difference in the 7 Treatments using SS_{Model}:

Test $H_o : \mu_{11} = \mu_{12} = \mu_{21} = \mu_{23} = \mu_{31} = \mu_{32} = \mu_{33}$ versus H_1 : Not all μ_{ij} are equal

$F = \frac{MS_{MODEL}}{MS_E} = \frac{12.471429/6}{9.866667/14} = 2.95$ with $df = 6, 14 \quad \Rightarrow \quad p - value = 1 - G(2.95) = .0447$

Therefore, there appears to be some evidence of a difference in the 7 treatment means.

II. Construct Contrasts which represent comparisons between treatment means which are main effects and two-way interactions:

The following table would be 8 mutually orthogonal contrasts which would represent the 9-1=8 df for decomposing SS_{MODEL} into components for Main Effects and Interaction provided all 9 treatments were observed:

Coefficients for Mutually Orthogonal Contrasts in 9 Treatment Means

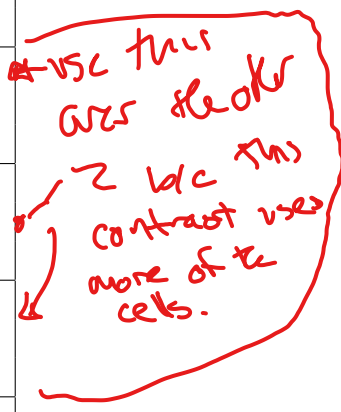
		TREATMENT MEANS								
CONTRAST	EFFECT	μ_{11}	μ_{12}	μ^*_{13}	μ_{21}	μ^*_{22}	μ_{23}	μ_{31}	μ_{32}	μ_{33}
Main Fat	C_1	1	1	1	-1	-1	-1	0	0	0
	C_2	1	1	1	1	1	1	-2	-2	-2
Main Surf.	C_3	1	-1	0	1	-1	0	1	-1	0
	C_4	1	1	-2	1	1	-2	1	1	-2
Interaction	C_5	1	-1	0	-1	1	0	0	0	0
	C_6	1	1	-2	-1	-1	2	0	0	0
	C_7	1	-1	0	1	-1	0	-2	2	0
	C_8	1	1	-2	1	1	-2	-2	-2	4

Note: * indicates that treatment was not observed

Because not all factor combinations were observed, the contrasts which represent Main Effects and Interactions are modified to the following contrasts:

Coefficients for Contrasts in Observed 7 Treatment Means

CONTRAST	EFFECT	TREATMENT MEANS μ_{11}	μ_{12}	μ_{21}	μ_{23}	μ_{31}	μ_{32}	μ_{33}
MAIN,FAT								
	C_1	1	1	0	0	-1	-1	0
	C_2	0	0	1	1	-1	0	-1
MAIN,SURF,1								
	C_3	1	-1	0	0	1	-1	0
	C_4	0	0	1	-1	1	0	-1
MAIN,SURF,2								
	C_3	0	0	0	0	1	0	-1
	C_4	0	0	0	0	1	-2	1
MAIN,SURF,3								
	C_3	0	0	0	0	0	1	-1
	C_4	0	0	1	-1	1	0	-1
INTERACTION								
	C_5	1	-1	0	0	-1	1	0
	C_6	0	0	1	-1	-1	0	1



The choice for the contrasts are not unique as is illustrated with 3 possible sets of contrasts for evaluating the Main Effect of Surfacant. Furthermore, the set of 6 contrasts is not a set of orthogonal contrasts.

The determination of whether there is significant evidence of a "Main Effect" for Fat or Surfacant and whether there is a significant evidence of an "Interaction" between Fat and Surfacant relies on testing the significance of the above contrasts.

III. Test if "Interaction-Type" Contrasts are Significant:

$\mu^{\mathbf{t}} = (\mu_{11},\ \mu_{12},\ \mu_{21},\ \mu_{23},\ \mu_{31},\ \mu_{32},\ \mu_{33})$

$\hat{\mu}^{\mathbf{t}} = (5.567,\ 6.2,\ 6.8,\ 6.0,\ 6.5,\ 7.2,\ 8.3)$

With $\mathbf{C_1} = (1,-1,0,0,-1,1,0)$; $H_o : \mathbf{C_1}\mu = 0 \ \Rightarrow \ H_o : (\mu_{11} - \mu_{12}) = (\mu_{31} - \mu_{32})$

With $\mathbf{C_2} = (0,0,1,-1,-1,0,1)$; $H_o : \mathbf{C_2}\mu = 0 \ \Rightarrow \ H_o : (\mu_{21} - \mu_{23}) = (\mu_{31} - \mu_{33})$

$$\mathbf{C_{F*S}} = \begin{pmatrix} \mathbf{C_1} \\ \mathbf{C_2} \end{pmatrix}$$

$$\mathbf{C_{F*S}}\hat{\mu} = \begin{pmatrix} .066667 \\ 2.6 \end{pmatrix}$$

$$\mathbf{D} = Diag\left(\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{4}, \tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{2}\right)$$

$$\left(\mathbf{C_{F*S}}\mathbf{D}\mathbf{C}^{\mathbf{t}}_{\mathbf{F*S}}\right)^{-1} = \tfrac{144}{287}\begin{pmatrix} \frac{19}{12} & \frac{-1}{2} \\ \frac{-1}{2} & \frac{17}{12} \end{pmatrix}$$

$SS_C = \left(\mathbf{C_{F*S}}\hat{\mu}\right)^{\mathbf{t}}\left(\mathbf{C_{F*S}}\mathbf{D}\mathbf{C}^{\mathbf{t}}_{\mathbf{F*S}}\right)^{-1}\left(\mathbf{C_{F*S}}\hat{\mu}\right) = \mathbf{4.7215796}$ with df_{F*S}= Row Rank($\mathbf{C_{F*S}}$) = 2

Test $H_o : \mathbf{C_{F*S}}\mu = 0$ vs $H_o : \mathbf{C_{F*S}}\mu \neq 0$

$F = \frac{MS_C}{MSE} = \frac{4.7216/2}{9.86667/14} = 3.35 \ \Rightarrow p - value = Pr[F_{2,14} \geq 3.35] = 0.065$

There is not significant evidence of an "Interaction between Fat and Surfacant".

R Code to compute the above:

```
library(MASS)

mu = c(5.56666667,6.2,6.8,6.0,6.5,7.2,8.3)
cfs = matrix(c(1,0,-1,0,0,1,0,-1,-1,-1,1,0,0,1),nrow=2)
cfs%*%mu
D=diag(c(1/3,1/3,1/3,1/4,1/2,1/4,1/2))
SSC = t(cfs%*%mu)%*%ginv(cfs%*%D%*%t(cfs))%*%(cfs%*%mu)
SSC
4.72158
```

IV. Test if Fat "Main Effect-Type" Contrasts are Significant:

$\mathbf{C_1} = (1,1,0,0,-1,-1,0) \Rightarrow \quad H_o : \frac{1}{2}(\mu_{11}+\mu_{12}) = \frac{1}{2}(\mu_{31}+\mu_{32})$

$\mathbf{C_2} = (0,0,1,-1,-1,0,1) \Rightarrow \quad H_o : \frac{1}{2}(\mu_{21}+\mu_{23}) = \frac{1}{2}(\mu_{31}+\mu_{33})$

$$\mathbf{C_F} = \begin{pmatrix} \mathbf{C_1} \\ \mathbf{C_2} \end{pmatrix}$$

$$\mathbf{C_F}\hat{\mu} = \begin{pmatrix} -1.9333 \\ -2.0 \end{pmatrix}$$

$$\left(\mathbf{C_F D C_F^t}\right)^{-1} = \frac{144}{287}\begin{pmatrix} \frac{19}{12} & \frac{-1}{2} \\ \frac{-1}{2} & \frac{17}{12} \end{pmatrix}$$

$SS_C = \left(\mathbf{C_F}\hat{\mu}\right)^{\mathbf{t}}\left(\mathbf{C_F D C_F^t}\right)^{-\mathbf{1}}\left(\mathbf{C_F}\hat{\mu}\right) = \mathbf{3.873}$ with df_F=Row Rank($\mathbf{C_F}$) = 2

Test $H_o : \mathbf{C_F}\mu = 0 \quad$ vs $\quad H_o : \mathbf{C_F}\mu \neq 0$

$F = \frac{MS_C}{MSE} = \frac{3.873/2}{9.86667/14} = 2.75 \quad \Rightarrow p - value = Pr[F_{2,14} \geq 2.75] = 0.099$

There is not significant evidence of a "Fat Main Effect".

V. Test if Surfacant "Main Effect-Type" Contrasts are Significant:

$\mathbf{C_1} = (0,0,0,0,0,1,-1) \Rightarrow \quad H_o : \mu_{32} = \mu_{33}$

$\mathbf{C_2} = (0,0,1,-1,1,0,-1) \Rightarrow \quad H_o : \frac{1}{2}(\mu_{21} + \mu_{31}) = \frac{1}{2}(\mu_{23} + \mu_{33})$

$\mathbf{C_S} = \begin{pmatrix} \mathbf{C_1} \\ \mathbf{C_2} \end{pmatrix}$

$\mathbf{C_S}\hat{\mu} = \begin{pmatrix} -1.1 \\ -1.0 \end{pmatrix}$

$\left(\mathbf{C_S D C_S^t}\right)^{-1} = \frac{144}{135} \begin{pmatrix} \frac{19}{12} & \frac{-1}{2} \\ \frac{-1}{2} & \frac{9}{12} \end{pmatrix}$

$SS_C = \left(\mathbf{C_S}\hat{\mu}\right)^{\mathbf{t}} \left(\mathbf{C_S D C_S^t}\right)^{-1} \left(\mathbf{C_S}\hat{\mu}\right) = \mathbf{1.6702}$ with df_S=Row Rank($\mathbf{C_F}$) = 2

Test $H_o : \mathbf{C_S}\mu = 0$ vs $H_o : \mathbf{C_S}\mu \neq 0$

$F = \frac{MS_C}{MSE} = \frac{1.6702/2}{9.86667/14} = 1.18 \quad \Rightarrow p - value = Pr[F_{2,14} \geq 1.18] = 0.3346$

There is not significant evidence of a "Surfacant Main Effect".

VI. Multiple Comparison of 7 Treatment Means and SAS Computations for Contrasts:

The following SAS code computes the tests of specified contrasts and implements Tukey's HSD procedure on the 7 treatments:

```
*On Dostat as twofact_missing,cellmeans.sas;
options ls=72 ps=58 nocenter nodate;
* This is Example 15.1 on page 192 in the
"Analysis of Messy Data, Vol I", by G. Milliken and D. Johnson;
*This experiment is a CR 3x3 factorial with 4 reps but there
is missing reps for some treatments and 0 reps for 2 treatments;
options pagesize=55 linesize=120;
data raw;
input trt fat surf fl1-fl4;
*drop fl1-fl4;
sv=fl1;output;
sv=fl2;output;
sv=fl3;output;
sv=fl4;output;
cards;
1 1  1  6.7 4.3 5.7  .
2 1  2  7.1  .  5.9 5.6
3 1  3   .   .   .  .
4 2  1   .  5.9 7.4 7.1
5 2  2   .   .   .  .
6 2  3  6.4 5.1 6.2 6.3
7 3  1  7.1 5.9  .   .
8 3  2  7.3 6.6 8.1 6.8
9 3  3   .  7.5 9.1  .
;
title 'Analysis as a CR 3x3 factorial';
proc glm;
class trt;
model sv = trt / ss1 ss2 ss3 ss4;
estimate 'Main1 of Fat'       trt  1   1  0  0 -1 -1  0;
estimate 'Main2 of Fat'       trt  0   0  1  1 -1  0 -1;
estimate 'Main1 of Surf'      trt  0   0  0  0  0  1 -1;
estimate 'Main2 of Surf'      trt  0   0  1 -1  1  0 -1;
estimate 'Inter1 of Fat&Surf' trt  1  -1  0  0 -1  1  0;
estimate 'Inter2 of Fat&Surf' trt  0   0  1 -1 -1  0  1;

contrast 'Main of Fat'        trt  1   1  0  0 -1 -1  0,
                              trt  0   0  1  1 -1  0 -1;
contrast 'Main of Surf'       trt  0   0  0  0  0  1 -1,
                              trt  0   0  1 -1  1  0 -1;
contrast 'Inter of Fat&Surf' trt  1  -1  0  0 -1  1  0,
                              trt  0   0  1 -1 -1  0  1;

means trt/tukey;
lsmeans trt / stderr pdiff adjust=tukey ;
run;
```

using cellmeans model.

Analysis as a CR with 7 treatments

Class	Levels	Values
trt	9	1 2 3 4 5 6 7 8 9

Number of observations 36

NOTE: Due to missing values, only 21 observations can be used in this analysis.

Dependent Variable: sv

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	12.47142857	2.07857143	2.95	0.0447
Error	14	9.86666667	0.70476190		
Corrected Total	20	22.33809524			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
trt	6	12.47142857	2.07857143	2.95	0.0447

Source	DF	Type II SS	Mean Square	F Value	Pr > F
trt	6	12.47142857	2.07857143	2.95	0.0447

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	6	12.47142857	2.07857143	2.95	0.0447

Source	DF	Type IV SS	Mean Square	F Value	Pr > F
trt	6	12.47142857	2.07857143	2.95	0.0447

- ANOVA

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Main of Fat	2	3.87252033	1.93626016	2.75	0.0985
Main of Surf	2	1.67022222	0.83511111	1.18	0.3346
Interaction	2	4.72157956	2.36078978	3.35	0.0647

$\alpha_{pc} = \frac{0.05}{6}$

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Main1 of Fat	-1.93333333	0.99920603	-1.93	0.0735
Main2 of Fat	-2.00000000	1.05634891	-1.89	0.0792
Main1 of Surf	-1.10000000	0.72702918	-1.51	0.1525
Main2 of Surf	-1.00000000	1.05634891	-0.95	0.3599
Interaction1	0.06666667	0.99920603	0.07	0.9477
Interaction2	2.60000000	1.05634891	2.46	0.0274

Level of trt	N	sv Mean	sv Std Dev
1	3	5.56666667	1.20554275
2	3	6.20000000	0.79372539
4	3	6.80000000	0.79372539
6	4	6.00000000	0.60553007
7	2	6.50000000	0.84852814
8	4	7.20000000	0.66833126
9	2	8.30000000	1.13137085

$\bar{y}_{i\cdot}$

Tukey's Studentized Range (HSD) Test for sv

NOTE: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	14
Error Mean Square	0.704762
Critical Value of Studentized Range	4.82895

Comparisons significant at the 0.05 level are indicated by ***.

trt Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
9 - 8	1.1000	-1.3825	3.5825	
9 - 4	1.5000	-1.1168	4.1168	
9 - 7	1.8000	-1.0665	4.6665	
9 - 2	2.1000	-0.5168	4.7168	
9 - 6	2.3000	-0.1825	4.7825	
9 - 1	2.7333	0.1165	5.3501	***
8 - 4	0.4000	-1.7894	2.5894	
8 - 7	0.7000	-1.7825	3.1825	
8 - 2	1.0000	-1.1894	3.1894	
8 - 6	1.2000	-0.8270	3.2270	
8 - 1	1.6333	-0.5560	3.8227	
4 - 7	0.3000	-2.3168	2.9168	
4 - 2	0.6000	-1.7405	2.9405	
4 - 6	0.8000	-1.3894	2.9894	
4 - 1	1.2333	-1.1072	3.5739	
7 - 2	0.3000	-2.3168	2.9168	
7 - 6	0.5000	-1.9825	2.9825	
7 - 1	0.9333	-1.6835	3.5501	
2 - 6	0.2000	-1.9894	2.3894	
2 - 1	0.6333	-1.7072	2.9739	
6 - 1	0.4333	-1.7560	2.6227	

only 1 significant difference

Least Squares Means

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

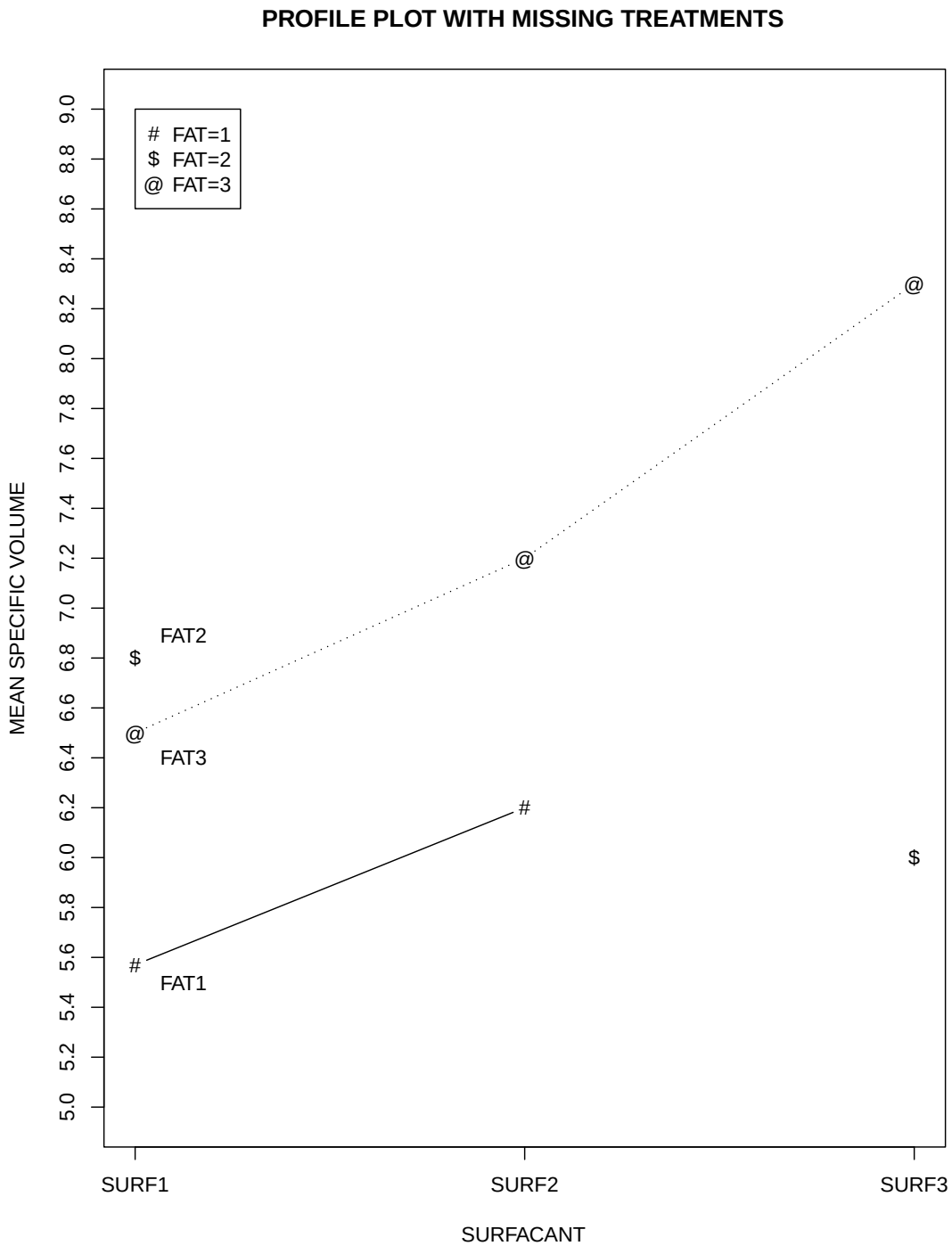
trt	sv LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
1	5.56666667	0.48468612	<.0001	1
2	6.20000000	0.48468612	<.0001	2
4	6.80000000	0.48468612	<.0001	3
6	6.00000000	0.41975049	<.0001	4
7	6.50000000	0.59361684	<.0001	5
8	7.20000000	0.41975049	<.0001	6
9	8.30000000	0.59361684	<.0001	7

Least Squares Means for effect trt
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: sv

i/j	1	2	3	4	5	6	7
1		0.9622	0.5687	0.9920	0.8760	0.2145	0.0381
2	0.9622		0.9707	0.9999	0.9996	0.7077	0.1584
3	0.5687	0.9707		0.8639	0.9996	0.9948	0.4787
4	0.9920	0.9999	0.8639		0.9912	0.4437	0.0778
5	0.8760	0.9996	0.9996	0.9912		0.9543	0.3805
6	0.2145	0.7077	0.9948	0.4437	0.9543		0.7336
7	0.0381	0.1584	0.4787	0.0778	0.3805	0.7336	

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used.



Return to Beginning of 11.0. (Recall we started on pg 23) at ≈ 29 min mark Wed 3/30/22 (week 10, lecture 26)