

## Topic: Designed experiments with multiple factors

- $2 \times 2$  experiments
- $a \times b$  experiments
- three-factor ANOVA
- nested vs. crossed designs (not described in packet)

### An example of a $2 \times 2$ study

Cholesterol measurements for random samples of  $n_j \equiv 7$  people from four populations given in the table below. Groups (cohorts) defined as follows:

- I The population of women younger than 50
- II The population of men younger than 50
- III The population of women 50 years or older
- IV The population of men 50 years or older

Sources of variability? \_\_\_\_\_, \_\_\_\_\_, \_\_\_\_\_

Group	Cholesterol level							avg	std. dev.
I	221	213	202	183	185	197	162	$\bar{y}_I = 194.7$	$s = 20$
II	271	192	189	209	227	236	142	$\bar{y}_{II} = 209.4$	$s = 41$
III	262	193	224	201	161	178	265	$\bar{y}_{III} = 212.0$	$s = 40$
IV	192	253	248	278	232	267	289	$\bar{y}_{IV} = 251.3$	$s = 32$

## One-way ANOVA Model:

$$\begin{aligned}
 Y_{ij} &= \mu_i + E_{ij} && \text{each estimable} \\
 &= \mu + \tau_i + E_{ij} && \text{each nonestimable}
 \end{aligned}$$

$i = 1, 2, 3, 4 \quad j = 1, 2, \dots, 7$  and  $E_{ij}$  i.i.d.  $N(0, \sigma^2)$

Parameters:  $\mu, \tau_1, \tau_2, \tau_3, \tau_4, \sigma^2$

One-way ANOVA table:

### The GLM Procedure

	Class cohort	Levels 4	Values				F Value	Pr > F
			I	II	III	IV		
Source	DF	Sum of Squares	Mean Square					
Model	3	12280.85714	4093.61905				3.46	0.0323
Error	24	28434.57143	1184.77381					
Corrected Total	27	40715.42857						

R-Square	Coeff Var	Root MSE	y Mean
0.301627	15.87245	34.42054	216.8571

Source	DF	Type I SS	Mean Square	F Value	Pr > F
cohort	3	12280.85714	4093.61905	3.46	0.0323

## Some terminology

Definition: A \_\_\_\_\_ in an experiment or study is a variable whose effect on the response is of primary interest. The values that a factor takes in the experiment are called factor \_\_\_\_\_ or treatments.

Definition: In \_\_\_\_\_, experimental units are randomly assigned to factor levels, or treatment groups.

Note: The cholesterol study is NOT a completely randomized design, as randomization of subjects to different levels of AGE and GENDER isn't possible.

Definition: When the same number of units are used for each treatment, the design is \_\_\_\_\_.

In one-way ANOVA of cholesterol data, COHORT is the factor, but it can be broken down into two factors in two-way ANOVA: AGE (factor A) and GENDER (factor B).

Definition: If there are observations at all combinations of all factors, the design is \_\_\_\_\_, otherwise it is \_\_\_\_\_.

- ① Estimate the mean difference in cholesterol between young men and young women.
- ② Estimate the mean difference between old men and old women.
- ③ Estimate the mean difference between men and women.
- ④ Estimate the mean difference between older and younger folks.
- ⑤ Estimate the mean difference between the differences estimated in 1. and 2.
- ⑥ Provide standard errors for all of these estimated contrasts
- ⑦ Specify the vectors defining these contrasts. For example, the first contrast of cohort means can be written

$$\theta_1 = (-1, 1, 0, 0)' \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} = \mu_2 - \mu_1$$

Consider the following contrasts of the cohort cholesterol means in the population:

$$\theta_3 = (-1, 1, -1, 1)' \mu$$

$$\theta_4 = (-1, -1, 1, 1)' \mu$$

$$\theta_5 = (-1, 1, 1, -1)' \mu$$

Q: Are these contrasts orthogonal?

Q: True/False:  $SS(\hat{\theta}_3) + SS(\hat{\theta}_4) + SS(\hat{\theta}_5) = SS[Trt]$

Another exercise:

- ① Compute the sums of squares for the estimated contrasts in 3., 4. and 5. using the exercise just completed and the fact that if  $\hat{\theta} = \sum c_i \bar{y}_{i+}$  then

$$SS[\hat{\theta}] = \frac{\hat{\theta}^2}{\sum \frac{c_i^2}{n_i}}.$$

- ② For each  $i = 3, 4, 5$ , obtain the  $F$ -ratio to test  $H_0 : \theta_i = 0$ .
- ③ Critical value for each test? \_\_\_\_\_. Draw conclusions.
- (3) an age effect
  - (4) a gender effect
  - (5) an age  $\times$  gender interaction

## Types of effects

Two-way ANOVA model for the cholesterol measurements:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

$i = 1, 2 = a$  and  $j = 1, 2 = b$  and  $k = 1, 2, \dots, 7 = n$ .

$E_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$ . Factors A, B: each with two levels ( $a = b = 2$ ).

Three types of effects: \_\_\_\_\_, \_\_\_\_\_, \_\_\_\_\_.

Exercise: Classify the contrasts from slide 4 as simple, interaction or main effects:

- ① Estimate the mean difference in cholesterol between young men and young women.
- ② Estimate the mean difference between old men and old women.
- ③ Estimate the mean difference between men and women.
- ④ Estimate the mean difference between older and younger folks.
- ⑤ Estimate the mean difference between the differences estimated in 1. and 2.

		GENDER	
		female ( $j = 1$ )	male ( $j = 2$ )
AGE	younger ( $i = 1$ )	194.7	209.4
	older ( $i = 2$ )	212.0	251.3

Partitioning the treatment SS into  $t - 1$  orthogonal components

$$12281 = SS[Trt] = SS[\hat{\theta}_3] + SS[\hat{\theta}_4] + SS[\hat{\theta}_5] = 5103 + 6121 + 1056$$

SS and F test for interaction effect

$H_0 : (\alpha\beta)_{ij} \equiv 0$  (i.e. no interaction) versus  $H_1 : (\alpha\beta)_{ij} \neq 0$  for some  $i, j$ .

$$\text{Use } \theta_5 = \mu(AB) \quad \text{and} \quad F = \frac{SS(\hat{\theta}_{AB})/((a-1)(b-1))}{MS[E]}$$

on  $df = 1, 24$ . For cholesterol data the estimated interaction effect is

$$\hat{\theta}_{AB} = \hat{\theta}_5 = \hat{\mu}(AB) = (251.3 - 209.4) - (212 - 194.7) = 41.9 - 17.3 = 24.6$$

the associated sum of squares is

$$SS(\hat{\theta}_5) = \frac{(24.6)^2}{\frac{1}{7} + \frac{(-1)^2}{7} + \frac{(-1)^2}{7} + \frac{1}{7}} = \frac{(24.6)^2}{\frac{4}{7}} = 1056$$

and  $F = 1056/1185 = 0.9$  which (is/isn't) significant at  $\alpha = 0.05$  on 1,24  $df$ .

Conclusion: men's age effect (41.9) not significantly greater than women's (17.3)



### F test for main effects

To test for main effect of A: AGE  $H_0 : \alpha_1 = \alpha_2 = 0$  vs.  $H_1 : \alpha_i$  not both 0

$$\text{use } \theta_4 = \mu(A) \quad \text{and} \quad F = \frac{SS(\hat{\theta}_4)}{MS[E]}$$

on 1, 24 *df*. The estimated main effect of AGE is

$$\hat{\mu}(A) = \frac{(251.3 - 209.4)}{2} + \frac{(212 - 194.7)}{2} = \frac{59.2}{2} = 29.6$$

the associated sum of squares is

$$SS(\hat{\theta}_4) = \frac{(29.6)^2}{\frac{(\frac{1}{2})^2}{7} + \frac{(\frac{1}{2})^2}{7} + \frac{(\frac{1}{2})^2}{7} + \frac{(\frac{1}{2})^2}{7}} = \frac{(29.6)^2}{\frac{1}{7}} = 6121$$

and

$$F = 6121/1185 = 5.2$$

Similarly for the main effect of B: gender

$$\hat{\mu}(B) = \frac{209.4 - 194.7}{2} + \frac{251.3 - 212}{2} = 27, SS(\hat{\theta}_3) = 5103$$

Leading to  $F_A = 6121/1185 = 5.2$ ,  $F_B = 5103/1185 = 4.3$  since  $F(0.05, \quad, \quad) = 4.26$  both GENDER and AGE effects significant at  $\alpha = 0.05$ .

Reminder: If  $SS[\text{Model}]$  on  $df = t - 1$  can be partitioned into  $t - 1$  orthogonal contrast sums of squares, then

$$SS[Trt] = \sum_1^{t-1} SS(\hat{\theta}_i)$$

. For cholesterol,

$$\begin{aligned} SS[Trt] &= SS( \quad ) + SS( \quad ) + SS( \quad ) \\ &= \quad + \quad + \quad \end{aligned}$$

## Confidence intervals for effects

If  $\theta = c'\mu$ ,  $100(1 - \alpha)\%$  confidence interval given by

$$\hat{\theta} \pm t(\alpha/2, N - t) \sqrt{MS[E] \sum \frac{c_i^2}{n_i}}$$

For the cholesterol data, with  $t(0.025, 24) = 2.06$  we have a 95% confidence interval for the AGE  $\times$  GENDER interaction effect:

$$24.6 \pm 2.06 \sqrt{\frac{4}{7} 1185} \quad \text{or} \quad 24.6 \pm 2.06(26.0) \quad \text{or} \quad (-29, 78)$$

a 95% confidence interval for the AGE effect:

$$29.6 \pm 2.06 \sqrt{\frac{1}{7} 1185} \quad \text{or} \quad 29.6 \pm 2.06(13.0) \quad \text{or} \quad 29.6 \pm 26.8 \quad \text{or} \quad (2.7, 56.4)$$

and a 95% confidence interval for the GENDER effect: \_\_\_\_\_

The term under the  $\sqrt{\phantom{x}}$  is the estimated standard error of the estimated contrast:

$$\widehat{SE}(\sum c_i \bar{y}_{i.}) = \sqrt{MS[E] \sum \frac{c_i^2}{n_i}}$$

## SAS code for cholesterol problem

```
data one;
  input cohort $ @;
  do subj=1 to 7;
    input y @;
    if cohort="I" then do; gender="W"; age="y"; end;
    else if cohort="II" then do; gender="M"; age="y";end;
    else if cohort="III" then do; gender="W" ; age="o";end;
    else if cohort="IV" then do; gender="M" ; age="o";end;
    output;
  end;
cards;
I      221    213    202    183    185    197    162
II     271    192    189    209    227    236    142
III    262    193    224    201    161    178    265
IV     192    253    248    278    232    267    289
;
run;

proc glm;
  class cohort;
  model y=cohort/clparm;
  contrast "main effect of age" cohort -1 -1 1 1;
  contrast "main effect of gender" cohort -1 1 -1 1;
  contrast "interaction effect" cohort -1 1 1 -1;
  estimate "main effect of age" cohort -1 -1 1 1/divisor=2;
  estimate "main effect of gender" cohort -1 1 -1 1/divisor=2;
  estimate "interaction effect" cohort -1 1 1 -1;
run;

proc glm;
  class gender age;
  model y=age|gender;
run;
```

(SAS will overlook misspelling of contrast.)

## SAS output (abbreviated) for cholesterol problem

The SAS System  
The GLM Procedure

1

### Class Level Information

Class	Levels	Values
cohort	4	I II III IV

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	12280.85714	4093.61905	3.46	0.0323
Error	24	28434.57143	1184.77381		
Corrected Total	27	40715.42857			

R-Square	Coeff Var	Root MSE	y Mean
0.301627	15.87245	34.42054	216.8571

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
main effect of age	1	6121.285714	6121.285714	5.17	0.0323
main effect of gender	1	5103.000000	5103.000000	4.31	0.0488
interaction effect	1	1056.571429	1056.571429	0.89	0.3544

Parameter	Estimate	Standard Error	t Value	Pr >  t
main effect of age	29.5714286	13.0097426	2.27	0.0323
main effect of gender	27.0000000	13.0097426	2.08	0.0488
interaction effect	-24.5714286	26.0194851	-0.94	0.3544

Parameter	95% Confidence Limits	
main effect of age	2.7206396	56.4222175
main effect of gender	0.1492111	53.8507889
interaction effect	-78.2730065	29.1301493

The GLM Procedure

Class	Levels	Values
gender	2	M W
age	2	o y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	12280.85714	4093.61905	3.46	0.0323
Error	24	28434.57143	1184.77381		
Corrected Total	27	40715.42857			

R-Square	Coeff Var	Root MSE	y Mean
0.301627	15.87245	34.42054	216.8571

Source	DF	Type I SS	Mean Square	F Value	Pr > F
age	1	6121.285714	6121.285714	5.17	0.0323
gender	1	5103.000000	5103.000000	4.31	0.0488
gender*age	1	1056.571429	1056.571429	0.89	0.3544

## Exercise: (space on next page)

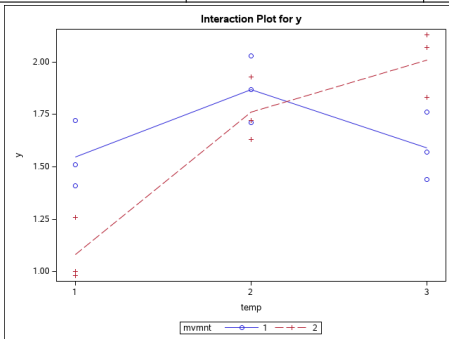
- Express the effects below in terms of model parameters  $\alpha_i, \beta_j, (\alpha\beta)_{ij}$ :
  - $\mu(AB_1)$  - \_\_\_\_\_ effect of \_\_\_\_\_ for \_\_\_\_\_ fixed at \_\_\_\_\_
  - $\mu(AB_2)$  - simple effect of \_\_\_\_\_ for \_\_\_\_\_ fixed at \_\_\_\_\_
  - $\mu(A_1B)$  - simple effect of \_\_\_\_\_ for \_\_\_\_\_ fixed at \_\_\_\_\_
  - $\mu(A)$  - \_\_\_\_\_ effect of \_\_\_\_\_
  - $\mu(B)$
  - $\mu(AB)$  - difference of \_\_\_\_\_ and \_\_\_\_\_ (sometimes divided by two)
- Estimate these effects

		GENDER	
		female ( $j = 1$ )	male ( $j = 2$ )
AGE	younger ( $i = 1$ )	194.7	209.4
	older ( $i = 2$ )	212.0	251.3

### $a \times b$ designs

Weight gain by  $N = 18$  tanks of fish randomized to  $a \times b = 3 \times 2$  combinations of temperature and movement: (moodle: "fishwtgain.dat").

Temp	Movement	Tanks			mean
1	1	1.41	1.72	1.51	1.55
1	2	1.00	1.26	0.98	1.08
2	1	1.87	1.71	2.03	1.87
2	2	1.93	1.72	1.63	1.76
3	1	1.76	1.44	1.57	1.59
3	2	2.07	2.13	1.83	2.01





```
proc glm data=two;
  class temp mvmnt;
  model y=temp|mvmnt;
run;
```

The SAS System  
The GLM Procedure

Class	Levels	Values
temp	3	1 2 3
mvmnt	2	1 2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1.58689444	0.31737889	12.71	0.0002
Error	12	0.29966667	0.02497222		
Corrected Total	17	1.88656111			

R-Square	Coeff Var	Root MSE	y Mean
0.841157	9.619440	0.158026	1.642778

Source	DF	Type I SS	Mean Square	F Value	Pr > F
temp	2	0.97747778	0.48873889	19.57	0.0002
mvmnt	1	0.01227222	0.01227222	0.49	0.4967
temp*mvmnt	2	0.59714444	0.29857222	11.96	0.0014

Source	DF	Type III SS	Mean Square	F Value	Pr > F
temp	2	0.97747778	0.48873889	19.57	0.0002
mvmnt	1	0.01227222	0.01227222	0.49	0.4967
temp*mvmnt	2	0.59714444	0.29857222	11.96	0.0014

An aside: only treatment totals and ANOVA table available in textbook, so data were synthesized:

```
data one;
  array ttotals{3,2} (4.65,3.24,5.61,5.28,4.77,6.03);
  do temp=1 to 3;
    do mvmnt=1 to 2;
      ymean=ttotals{temp,mvmnt}/3;
      do rep=1 to 3;
        error=rannor(123);
        output;
      end;
    end;
  end;
end;
run;
proc standard data=one out=two mean=0 std=.1581139;
  by temp mvmnt ;
  var error;
run;
data two;
  set two;
  y=ymean+error;
  y=round(y,0.01);
run;
```

### 3 × 2 fish weight gain example continued

The effect of temperature (1=cold,2=lukewarm,3=warm) depends on movement:

- when water still (mvmnt=1), temp effect is \_\_\_\_\_
- when water moving (mvmnt=2), temp effect is \_\_\_\_\_

The GLM Procedure

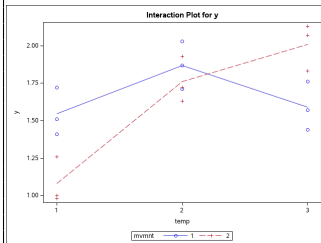
Level of temp	N	Mean	Std Dev
1	6	1.31333333	0.29173047
2	6	1.81500000	0.15280707
3	6	1.80000000	0.27085051

Level of mvmnt	N	Mean	Std Dev
1	9	1.66888889	0.20551426
2	9	1.61666667	0.43823510

Level of temp	Level of mvmnt	N	Mean	Std Dev
1	1	3	1.54666667	0.15821926
1	2	3	1.08000000	0.15620499
2	1	3	1.87000000	0.16000000
2	2	3	1.76000000	0.15394804
3	1	3	1.59000000	0.16093477
3	2	3	2.01000000	0.15874508



## Partitioning $SS[Total]$ in $a \times b$ design (Two-way ANOVA)

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + E_{ijk}$$

( $i = 1, 2, \dots, a$  and  $j = 1, 2, \dots, b$  and  $k = 1, 2, \dots, n$ )

Deviations:

total :  $y_{ijk} - \bar{y}...$

due to level  $i$  of factor A: \_\_\_\_\_

due to level  $j$  of factor B: \_\_\_\_\_

due to levels  $i$  of factor A and  $j$  of factor B after subtracting main effects:

$$\bar{y}_{ij.} - \bar{y}... -$$

$$SS[Total] = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}...) ^2 = \sum_i \sum_j \sum_k (\bar{y}_{ij.} - \bar{y}...) ^2 + \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.}) ^2$$

$$SS[A] = \sum_i \sum_j \sum_k (\bar{y}_{i..} - \bar{y}...) ^2, \quad SS[B] = \sum_i \sum_j \sum_k (\bar{y}_{.j.} - \bar{y}...) ^2$$

$$SS[AB] = \sum_i \sum_j \sum_k (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}...) ^2, \quad SS[E] = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.}) ^2$$

## ANOVA for two-factor crossed design

$$y_{ijk} = \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) + y_{ijk} - \bar{y}_{ij.}$$

$$y_{ijk} - \bar{y}_{...} = (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) + y_{ijk} - \bar{y}_{ij.}$$

Square both sides, sum over  $i, j, k$ , and the  $\times$ -products vanish.

$$SS(Tot) = SS(Trt) + SS( \quad )$$

$$SS(Trt) = SS(A) + \quad +$$

Analysis of replicated two (or more) factor designs often proceed according to the following steps:

- ➊ Check for interaction
  - ➊ If no interaction, analyze main effects
  - ➋ If interaction, analyze simple effects

### $a \times b$ example continued

Test for interaction effect in  $2 \times 2$  generalizes to  $a \times b$ :

$$H_0 : (\alpha\beta)_{ij} \equiv 0 \text{ vs. } H_1 : (\alpha\beta)_{ij} \neq 0 \text{ for some } i, j$$

$$F = \frac{MS[AB]}{MS[E]}$$

on  $(a - 1)(b - 1)$  and  $N - ab$  numerator, denominator  $df$ .

$$SS[AB] = n \sum_{i=1}^3 \sum_{j=1}^3 (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot\cdot})^2 = 0.597$$

$$F = \frac{.597/2}{0.025} = 11.96$$

which is highly significant ( $p = 0.0014$ ) on 2,12  $df$ .

We could proceed to test for main effects, but we won't.

Q: Why not?

A: Because effect of one factor depends on the level of the other factor, it might not make sense to talk about main effects.

If one insists on main effects, the appropriate  $F$ -ratios are

$$F_A = \frac{SS[A]/(a-1)}{MS[E]} \text{ on } a-1, N-ab \text{ df}$$

$$F_B = \frac{SS[B]/(b-1)}{MS[E]} \text{ on } b-1, N-ab \text{ df}$$

but the significance of the interaction effect suggests that the effect of one factor, say  $A$ , differs across levels of the other factor. A test for the main effect of  $A$  is based on the effect of  $A$  after *averaging over levels of  $B$* . (Draw a picture.)



### $a \times b$ designs

Yields on 36 tomato crops from balanced, complete, crossed design with  $a = 3$  varieties ( $A$ ) at  $b = 4$  planting densities ( $B$ ) :

Variety	Density $k$ /hectare	Sample		
1	10	7.9	9.2	10.5
2	10	8.1	8.6	10.1
3	10	15.3	16.1	17.5
1	20	11.2	12.8	13.3
2	20	11.5	12.7	13.7
3	20	16.6	18.5	19.2
1	30	12.1	12.6	14.0
2	30	13.7	14.4	15.4
3	30	18.0	20.8	21.0
1	40	9.1	10.8	12.5
2	40	11.3	12.5	14.5
3	40	17.2	18.4	18.9

Statistical model?

$$Y_{ijk} =$$

## ANOVA table

The SAS System  
The GLM Procedure

1

Class	Levels	Values
a	3	1 2 3
b	4	10 20 30 40

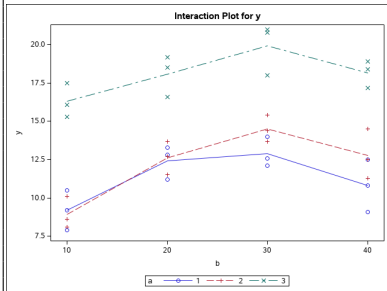
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	422.3155556	38.3923232	24.22	<.0001
Error	24	38.0400000	1.5850000		
Corrected Total	35	460.3555556			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
a	2	327.5972222	163.7986111	103.34	<.0001
b	3	86.6866667	28.8955556	18.23	<.0001
a*b	6	8.0316667	1.3386111	0.84	0.5484

Level of a	N	Mean	Std Dev
1	12	11.3333333	1.88309867
2	12	12.2083333	2.34887142
3	12	18.1250000	1.73369023

Level of b	N	Mean	Std Dev
10	9	11.4777778	3.75458978
20	9	14.3888889	2.96835158
30	9	15.7777778	3.36480972
40	9	13.9111111	3.53250777

Level of a	Level of b	N	Mean	Std Dev
1	10	3	9.20000000	1.30000000
1	20	3	12.4333333	1.09696551
1	30	3	12.9000000	0.98488578
1	40	3	10.8000000	1.70000000
2	10	3	8.93333333	1.04083300
2	20	3	12.6333333	1.10151411
2	30	3	14.5000000	0.85440037
2	40	3	12.7666667	1.61658075
3	10	3	16.3000000	1.11355287
3	20	3	18.1000000	1.34536240
3	30	3	19.9333333	1.67729942
3	40	3	18.1666667	0.87368949



A conventional look at main effects is just to make pairwise comparisons among marginal means, after averaging over other factors. Pairwise comparisons of density means using Tukey's procedure with  $\alpha = 0.05$  are given below. (Use means b/tukey;. to obtain the output.)

The GLM Procedure

Tukey's Studentized Range (HSD) Test for y

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	24
Error Mean Square	1.585
Critical Value of Studentized Range	3.90126
Minimum Significant Difference	1.6372

Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	b
A	15.7778	9	30
A			
B A	14.3889	9	20
B			
B	13.9111	9	40
C	11.4778	9	10

### A three-factor example

In a balanced, complete, crossed design,  $N = 36$  <sup>tanks</sup> ~~shrimp~~ were randomized to  $abc = 12$  treatment combinations from the factors below:

A1: Temperature at 25° C

A2: Temperature at 35° C

B1: Density of shrimp population at 80 shrimp/40l

B2: Density of shrimp population at 160 shrimp/40l

C1: Salinity at 10 units

C2: Salinity at 25 units

C3: Salinity at 40 units

The response variable of interest is weight gain  $Y_{ijkl}$  after four weeks.

#### Three-way ANOVA Model:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\beta\gamma)_{jk} + (\alpha\gamma)_{ik} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkl}$$

$i = 1, 2 \quad j = 1, 2 \quad k = 1, 2, 3 \quad l = 1, 2, 3$

$\epsilon_{ijkl} \stackrel{iid}{\sim} N(0, \sigma^2)$

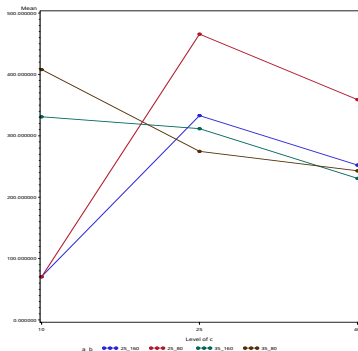
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	467636.3333	42512.3939	14.64	<.0001
Error	24	69690.6667	2903.7778		
Corrected Total	35	537327.0000			

$\sqrt{2MS(E)/3} \approx 44$

Source	DF	Type I SS	Mean Square	F Value	Pr > F
a	1	15376.0000	15376.0000	5.30	0.0304
b	1	21218.7778	21218.7778	7.31	0.0124
a*b	1	8711.1111	8711.1111	3.00	0.0961
c	2	96762.5000	48381.2500	16.66	<.0001
a*c	2	300855.1667	150427.5833	51.80	<.0001
b*c	2	674.3889	337.1944	0.12	0.8909
a*b*c	2	24038.3889	12019.1944	4.14	0.0285

by

sig



$$H_0: \mu_{ijk} \equiv \mu$$

$$F = \frac{MS_{a,b}}{MS_{a,b,c}}$$

Level of a	Level of b	N	Mean	Std Dev
25	80	9	298.333333	185.106051
25	160	9	218.666667	128.739077
35	80	9	308.555556	85.475305
35	160	9	291.111111	57.953525

means a/b/c;

Level of a	Level of c	N	Mean	Std Dev
25	10	6	70.500000	15.109600
25	25	6	399.333333	114.206246
25	40	6	305.666667	69.987618
35	10	6	369.500000	56.450864
35	25	6	293.166667	45.375838
35	40	6	236.833333	38.096807

Level of b	Level of c	N	Mean	Std Dev
80	10	6	239.166667	188.065326
80	25	6	370.166667	122.218520
80	40	6	301.000000	77.415761
160	10	6	200.833333	144.240655
160	25	6	322.333333	74.529636
160	40	6	241.500000	32.788718

Level of a	Level of b	Level of c	N	Mean	Std Dev
25	80	10	3	70.333333	17.156146
25	80	25	3	465.666667	87.648921
25	80	40	3	359.000000	59.858166
25	160	10	3	70.666667	16.623277
25	160	25	3	333.000000	108.282039
25	160	40	3	252.333333	11.372481
35	80	10	3	408.000000	51.117512
35	80	25	3	274.666667	47.961790
35	80	40	3	243.000000	36.166283
35	160	10	3	331.000000	30.116441
35	160	25	3	311.666667	42.665365
35	160	40	3	230.666667	46.971623

Interpretation of third order interaction  
Interpretation of second order interaction

1<sup>st</sup> order interaction is between two factors

2<sup>nd</sup> order interaction is between three factors

Upon inspection of the interaction plot, what do you see?

What is the primary two-factor/first-order interaction? \_\_\_\_\_

Consider the means for low temperature (red and blue). Do you see evidence of BC interaction for temperature is low? Characterize it.

Do you see evidence of BC interaction for temperature is high?

If there is a BC interaction at one level of A but not the other, this is a second-order interaction.

Characterization of a three-factor interaction may not be unique. Here we first fixed A, but another analyst might first fix some other factor and characterize factorial effects in a different order.



```

%let d=divisor;    *an example of a macro variable;
data one;
  drop i;          /* a=temp, b=density, c=salinity */
  input a b c @;   * @ hold the line. prevent DATA step from loading ;
  do i=1 to 3;     * new record when next INPUT encountered;
    input y @;     * @ hold the line;    *love isn't always on time! (Toto) ;
    y0=sqrt(y);
    output;
  end;
  cards;
25 80 10 86 52 73
25 80 25 544 371 482
25 80 40 390 290 397
25 160 10 53 73 86
25 160 25 393 398 208
25 160 40 249 265 243
35 80 10 439 436 349
35 80 25 249 245 330
35 80 40 247 277 205
35 160 10 324 305 364
35 160 25 352 267 316
35 160 40 188 223 281
;
run;

proc glimmix data=one;
  class a b c;
  model y=a|b|c;
  lsmeans a*b*c/slicediff=a*c;
run;

```

*will consider every diff among levels of B fixed for each combo of A & C*

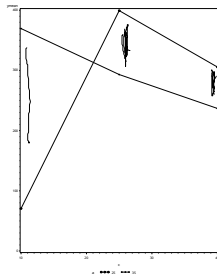
```

proc glimmix data=one;
  class a b c;
  model y=a|b|c;
  estimate "temp effect at c=1" a -1 1 a*c -1 0 0 1 0 0;
  estimate "temp effect at c=2" a -1 1 a*c 0 -1 0 0 1 0;
  estimate "temp effect at c=3" a -1 1 a*c 0 0 -1 0 0 1;
  estimate "avg of temp effects at c=2,3" a -2 2 a*c 0 -1 -1 0 1 1/&d=2;
  estimate "mu[AC1]-.5(mu[AC2]+mu[AC3])" a*c -2 1 1 2 -1 -1/divisor=2;
run;

```

Handwritten notes above the code:

- $-d_1 + d_2$
- $-(28)_{11} + (48)_{21} = 22 - d_1 + (28)_{21} - (48)_{11}$
- low temp
- Hi temp

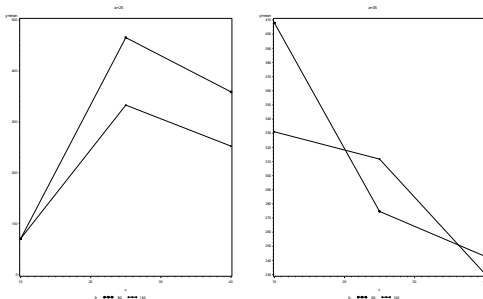


$$\frac{1}{2}(\mu[AC_2] + \mu[AC_3]) = 22 - d_1 + \frac{1}{2}[(28)_{22} - (48)_{12} + (28)_{23} - (48)_{13}]$$

Estimates					
Label	Estimate	Standard Error	DF	t Value	Pr >  t
temp effect at c=1	299.00	31.1115	24	9.61	<.0001
temp effect at c=2	-106.17	31.1115	24	-3.41	0.0023
temp effect at c=3	-68.8333	31.1115	24	-2.21	0.0367
avg of temp effects at c=2,3	-87.5000	21.9992	24	-3.98	0.0006
mu[AC1]-.5(mu[AC2]+mu[AC3])	386.50	38.1037	24	10.14	<.0001

We've characterized the  $A \times C$  interaction. Note  $SS(AC)$ .

What else? The  $B \times C$  interaction for fixed  $A$ .



$$\mu[A_1BC_1] = \mu_{121} - \mu_{111}$$

$$\mu[A_1BC_2] = \mu_{122} - \mu_{112}$$

$$\mu[A_1BC_3] = \mu_{123} - \mu_{113}$$

$$\mu[A_2BC_1] = \mu_{221} - \mu_{211}$$

$$\mu[A_2BC_2] = \mu_{222} - \mu_{212}$$

$$\mu[A_2BC_3] = \mu_{223} - \mu_{213}$$

We can see now why it might not make sense to test for the “main effect” of  $B$ . The effect of  $B$  appears to be quite different across levels of  $A$  and  $C$ .

We can look at the effects of  $B$  after fixing  $A$  and  $C$ . For example,

$$\mu[A_1BC_1] = \beta_2 - \beta_1 + (\alpha\beta)_{12} - (\alpha\beta)_{11} + (\beta\gamma)_{21} - (\beta\gamma)_{11} + (\alpha\beta\gamma)_{121} - (\alpha\beta\gamma)_{111}$$

```
*class a b c;  
estimate "density effect at c=1,a=1"  
      b -1 1 a*b -1 1 b*c -1 0 0      1 0 0 a*b*c -1 0 0      1 0 0 ;
```

Similarly, we can look at lots of contrasts and compute contrast sums of squares:

```
proc glm data=one;
  class a b c;
  model y=a|b|c;
  estimate "density effect at c=1,a=1"
    b -1 1 a*b -1 1 b*c -1 0 0 1 0 0 a*b*c -1 0 0 1 0 0 ;
  estimate "density effect at c=2,a=1" b -1 1 b*c 0 -1 0 0 1 0
    b -1 1 a*b -1 1 b*c 0 -1 0 0 1 0 a*b*c 0 -1 0 0 1 0 ;
  estimate "density effect at c=3,a=1" b -1 1 b*c 0 0 -1 0 0 1
    b -1 1 a*b -1 1 b*c 0 0 -1 0 0 1 a*b*c 0 0 -1 0 0 1 ;
  estimate "mu[A1BC1]-.5(mu[A1BC2]+mu[A1BC3])" b*c -2 1 1 2 -1 -1
    a*b*c -2 1 1 2 -1 -1/divisor=2;
  estimate "mu[A2BC1]-.5(mu[A2BC2]+mu[A2BC3])" b*c -2 1 1 2 -1 -1
    a*b*c 0 0 0 0 0 0 -2 1 1 2 -1 -1/divisor=2;
  estimate "mu[A1BC1.....] - mu[A2BC1.....]"
    a*b*c -2 1 1 2 -1 -1 2 -1 -1 -2 1 1/divisor=2;

  contrast "density effect at c=1,a=1"
    b -1 1 a*b -1 1 b*c -1 0 0 1 0 0 a*b*c -1 0 0 1 0 0 ;
  contrast "density effect at c=2,a=1" b -1 1 b*c 0 -1 0 0 1 0
    b -1 1 a*b -1 1 b*c 0 -1 0 0 1 0 a*b*c 0 -1 0 0 1 0 ;
  contrast "density effect at c=3,a=1" b -1 1 b*c 0 0 -1 0 0 1
    b -1 1 a*b -1 1 b*c 0 0 -1 0 0 1 a*b*c 0 0 -1 0 0 1 ;
  contrast "mu[A1BC1]-.5(mu[A1BC2]+mu[A1BC3])" b*c -2 1 1 2 -1 -1
    a*b*c -2 1 1 2 -1 -1;
  contrast "mu[A2BC1]-.5(mu[A2BC2]+mu[A2BC3])" b*c -2 1 1 2 -1 -1
    a*b*c 0 0 0 0 0 0 -2 1 1 2 -1 -1;
  contrast "diff between last two"
    a*b*c -2 1 1 2 -1 -1 2 -1 -1 -2 1 1;

run;
```

Parameter	Estimate	Standard Error	t Value	Pr >  t
density effect at c=1,a=1	0.333333	43.9983165	0.01	0.9940
density effect at c=2,a=1	-132.666667	43.9983165	-3.02	0.0060
density effect at c=3,a=1	-106.666667	43.9983165	-2.42	0.0232
mu[A1BC1]-.5(mu[A1BC2]+mu[A1BC3])	120.000000	53.8867124	2.23	0.0356
mu[A2BC1]-.5(mu[A2BC2]+mu[A2BC3])	-89.333333	53.8867124	-1.66	0.1104
mu[A1BC1.....] - mu[A2BC1.....]	209.333333	76.2073196	2.75	0.0112

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
mu[AC1]-.5(mu[AC2]+mu[AC3])	1	298764.5000	298764.5000	102.89	<.0001
density effect at c=1,a=1	1	0.1667	0.1667	0.00	0.9940
density effect at c=2,a=1	1	26400.6667	26400.6667	9.09	0.0060
density effect at c=3,a=1	1	17066.6667	17066.6667	5.88	0.0232
mu[A1BC1]-.5(mu[A1BC2]+mu[A1BC3])	1	14400.0000	14400.0000	4.96	0.0356
mu[A2BC1]-.5(mu[A2BC2]+mu[A2BC3])	1	7980.4444	7980.4444	2.75	0.1104
diff between last two	1	21910.2222	21910.2222	7.55	0.0112

a\*b\*c Effect Sliced by a\*c for y

a	c	DF	Sum of Squares	Mean Square	F Value	Pr > F
25	10	1	0.166667	0.166667	0.00	0.9940
25	25	1	26401	26401	9.09	0.0060
25	40	1	17067	17067	5.88	0.0232
35	10	1	8893.500000	8893.500000	3.06	0.0929
35	25	1	2053.500000	2053.500000	0.71	0.4087
35	40	1	228.166667	228.166667	0.08	0.7816

Recall,  $SS(AC) \sim 301000$  and  $SS(ABC) \sim 24000$ . We've explained most of this variation with two single *df* contrasts.

## Activity w/ ESTIMATE statement

Exercise: identify the estimable contrasts in each of the ESTIMATE statements in the correspondence below, which pertains to a  $3 \times 2$  study with factors and levels

Factor	Levels
A : additive	acetic, nothing, sorbate
B : uv	0,1.

To: osborne@stat.ncsu.edu  
Subject: non estimatable estimate statements

I am still having trouble with the estimate statements, the only ones that work for the additive\*uv interaction are where we contrast the same additive over the uv, can anything be done about this??

proc glm;

*order important*

```

class additive uv;
model ycount=additive uv uv*additive;
estimate 'acetic uv=0 vs acetic uv=1' uv 1 -1 uv*additive 1 -1 0 0 0 0; ✓
estimate 'acetic uv=0 vs nothing uv=0' uv 1 -1 uv*additive 1 0 -1 0 0 0; X
estimate 'acetic uv=0 vs nothing uv=1' uv 1 -1 uv*additive 1 0 0 -1 0 0;
estimate 'acetic uv=0 vs sorbate uv=0' uv 1 -1 uv*additive 1 0 0 0 -1 0;
estimate 'acetic uv=0 vs sorbate uv=1' uv 1 -1 uv*additive 1 0 0 0 0 -1;
estimate 'acetic uv=1 vs nothing uv=0' uv 1 -1 uv*additive 0 1 -1 0 0 0;
estimate 'acetic uv=1 vs nothing uv=1' uv 1 -1 uv*additive 0 1 0 -1 0 0;
estimate 'acetic uv=1 vs sorbate uv=0' uv 1 -1 uv*additive 0 1 0 0 -1 0;
estimate 'acetic uv=1 vs sorbate uv=1' uv 1 -1 uv*additive 0 1 0 0 0 -1;
estimate 'nothing uv=0 vs nothing uv=1' uv 1 -1 uv*additive 0 0 1 -1 0 0;
estimate 'nothing uv=0 vs sorbate uv=0' uv 1 -1 uv*additive 0 0 1 0 -1 0;
estimate 'nothing uv=0 vs sorbate uv=1' uv 1 -1 uv*additive 0 0 1 0 0 -1;
estimate 'nothing uv=1 vs sorbate uv=0' uv 1 -1 uv*additive 0 0 0 1 -1 0;
estimate 'nothing uv=1 vs sorbate uv=1' uv 1 -1 uv*additive 0 0 0 1 0 -1;
estimate 'sorbate uv=0 vs sorbate uv=1' uv 1 -1 uv*additive 0 0 0 0 1 -1;
estimate 'uv=0 vs uv=1' uv 1 -1;
estimate 'acetic vs nothing' additive 1 -1;
estimate 'acetic vs sorbate' additive 1 0 -1;
estimate 'nothing vs sorbate' additive 0 1 -1;

```

*correct  
any  
unspecified*

## Unbalanced design

Recall the  $2 \times 2$  cholesterol study. Suppose the study is unbalanced:

$$E(\bar{Y}_{2..} - \bar{Y}_{1..}) = \alpha_2 - \alpha_1 + \text{bias}$$

Age	Gender		Marginal mean
	Male	Female	
young	271, 192, 189, 209, 227, 236	162	$\bar{y}_{1++} = 212.3$
old	289	262, 193, 224, 201, 161, 178, 265	$\bar{y}_{2++} = 221.6$
	$\bar{y}_{+1+} = 230.4$	$\bar{y}_{+2+} = 205.8$	

$$\bar{y}_{11} = 220.7, \bar{y}_{12} = 162, \bar{y}_{21} = 289, \bar{y}_{22} = 212.$$

Consider an additive two-factor ANOVA model:  $Y_{ijk} = \mu + \alpha_i + \beta_j + E_{ijk}$

Exercise: finish parametric expressions for expected values below:

$$\begin{aligned} \frac{1}{7} \sum_{j=1}^2 \sum_{k=1}^7 E(Y_{ijk}) &= E(\bar{Y}_{1++}) = \mu + \alpha_1 + \frac{1}{7}(6\beta_1 + \beta_2) \quad \neq \mu + \alpha_1 + \frac{1}{2}(\beta_1 + \beta_2) \\ E(\bar{Y}_{2++}) &= \mu + \alpha_2 + \frac{1}{8}(\beta_1 + 7\beta_2) \quad \neq \mu + \alpha_2 + \frac{1}{2}(\beta_1 + \beta_2) \\ E(\bar{Y}_{+1+}) &= \mu + \frac{1}{7}(6\alpha_1 + \alpha_2) + \beta_1 \\ E(\bar{Y}_{+2+}) &= \mu + \frac{1}{8}(\alpha_1 + 7\alpha_2) + \beta_2 \end{aligned}$$

Marginal sample means are not real useful in this unbalanced study.

*Least squares means:* (what would be estimated by marginal means if design were balanced).



Parametric expressions for population means given below for additive model:

Population group	effect of interest	estimate	SE
Young folks	$\mu + \alpha_1 + \frac{1}{2}(\beta_1 + \beta_2)$	188.03	
Older folks	$\mu + \alpha_2 + \frac{1}{2}(\beta_1 + \beta_2)$	247.1	
Men	$\mu + \frac{1}{2}(\alpha_1 + \alpha_2) + \beta_1$	251.5	
Women	$\mu + \frac{1}{2}(\alpha_1 + \alpha_2) + \beta_2$	183.6	
Young men	$\mu + \alpha_1 + \beta_1$	222	
Older men	$\mu + \alpha_2 + \beta_1$		
Young women	$\mu + \alpha_1 + \beta_2$	154	
Older women	$\mu + \alpha_2 + \beta_2$		

`lsmeans age gender;` will report least squares estimates for first four means above.

		Standard		
gender	y LSMEAN	Error	Pr	>  t
m	251.525773	16.233482	<.0001	
w	183.597938	15.842256	<.0001	
		Standard		
age	y LSMEAN	Error	Pr	>  t
jr	188.025773 ✓	16.233482	<.0001	
sr	247.097938	15.842256	<.0001	

These quantities estimated using linear combinations of the treatment means of the form:

$$\hat{\theta} = c_{11}\bar{y}_{11+} + c_{12}\bar{y}_{12+} + c_{21}\bar{y}_{21+} + c_{22}\bar{y}_{22+}.$$

The coefficients are chosen so that  $E(\hat{\theta}) = \theta$  and  $\sum c_{ij}^2/n_{ij}$  is minimized.

Example: What are the coefficients for the contrast which estimates the population mean for young folks,  $\mu + \alpha_1 + \frac{1}{2}(\beta_1 + \beta_2)$  with minimum variance?

$$c_{11} + c_{12} = 1(\text{coeff for } \alpha_1)$$

$$c_{21} + c_{22} = 0(\text{coeff for } \alpha_2)$$

$$c_{11} + c_{21} = \frac{1}{2}(\text{coeff for } \beta_1)$$

$$c_{12} + c_{22} = \frac{1}{2}(\text{coeff for } \beta_2)$$

Variance is then proportional to

$$\frac{c_{11}^2}{n_{11}} + \frac{c_{12}^2}{n_{12}} + \frac{c_{21}^2}{n_{21}} + \frac{c_{22}^2}{n_{22}} = \frac{c_{11}^2}{n_{11}} + \frac{(1 - c_{11})^2}{n_{12}} + \frac{(\frac{1}{2} - c_{11})^2}{n_{21}} + \frac{(c_{11} - \frac{1}{2})^2}{n_{22}}$$

minimized at  $c_{11} = \frac{66}{97}$  by setting derivative to 0, solving. LS mean then

$$\hat{\theta} = \frac{66}{97}\bar{y}_{11+} + (1 - \frac{66}{97})\bar{y}_{12+} + (\frac{1}{2} - \frac{66}{97})\bar{y}_{21+} + (\frac{66}{97} - \frac{1}{2})\bar{y}_{22+} = 188.03.$$

Similarly for old folks, the contrast with minimum variance has

$$c_{11} = 18/97 = -c_{12}$$

and

$$c_{21} = \frac{1}{2} - 18/97, c_{22} = \frac{1}{2} + 18/97$$

so that the estimate for the old folks mean is

$$\frac{18}{97}\bar{y}_{11+} - \frac{18}{97}\bar{y}_{12+} + (\frac{1}{2} - \frac{18}{97})\bar{y}_{21+} + (\frac{1}{2} + \frac{18}{97})\bar{y}_{22+} = 247.1.$$

Exercise: obtain least squares estimators and estimates of marginal means for men and women as well as for each age $\times$ gender combination.

Q: Is there an age effect? Should we base our conclusion on

$$\sum_i \sum_j \sum_k (\bar{y}_{i++} - \bar{y}_{+++})^2 = 325.6?$$

A: Might not be a good idea if factor B has an effect.

(This is the unadjusted sum of squares for age.)

Alternatively, consider the contrast

$$\theta_{age} = \alpha_1 - \alpha_2.$$

We can obtain the coefficients of the LS estimate of this contrast and then use them to get sum of squares for age effect adjusted for gender, or type II SS

$$\hat{\theta}_{age} = \hat{\alpha}_1 - \hat{\alpha}_2 = c_{11}\bar{y}_{11+} + c_{12}\bar{y}_{12+} + c_{21}\bar{y}_{21+} + c_{22}\bar{y}_{22+} \quad \text{where}$$

$$c_{11} + c_{12} = 1(\text{coeff for } \alpha_1)$$

$$c_{21} + c_{22} = -1(\text{coeff for } \alpha_2)$$

$$c_{11} + c_{21} = 0(\text{coeff for } \beta_1)$$

$$c_{12} + c_{22} = 0(\text{coeff for } \beta_2)$$

$\text{Var}(\hat{\theta}_{age})$  is minimized when  $c_{11} = \frac{48}{97}$  which leads to

$$\hat{\theta}_{age} = -59.07$$

with

$$SS[\hat{\theta}_{age}] = \frac{(-59.07)^2}{\frac{(\frac{48}{97})^2}{6} + \frac{(-1 - \frac{48}{97})^2}{1} + \frac{(\frac{-48}{97})^2}{1} + \frac{(-1 + \frac{48}{97})^2}{7}} = 6044.$$

```

/*
I   221  213  202  183  185  197  162
II  271  192  189  209  227  236  142
III 262  193  224  201  161  178  265
IV  192  253  248  278  232  267  289
*/
options ls=75;
data one;
    input gender $ age $2. @;
    do i=1 to 7;
        input y @@;
        output;
    end;
cards;
w jr . . . . . 162
m jr 271 192 189 209 227 236 .
w sr 262 193 224 201 161 178 265
m sr . . . . . 289
;
run;

```

```

proc glm;
    class age gender;
    model y=age gender/solution;
    lsmeans gender age/stderr;
    estimate "lsmean for young folks" intercept 2 age 2 0 gender 1 1/divisor=2;
    estimate "lsmean for older folks" intercept 2 age 0 2 gender 1 1/divisor=2;
    estimate "lsmean for men" intercept 2 age 1 1 gender 2 0/divisor=2;
    estimate "lsmean for women" intercept 2 age 1 1 gender 0 2/divisor=2;
    estimate "lsmean for young men" intercept 1 age 1 0 gender 1 0;
    estimate "lsmean for young women" intercept 1 age 1 0 gender 0 1;
    estimate "lsmean for old men" intercept 1 age 0 1 gender 1 0;
    estimate "lsmean for old women" intercept 1 age 0 1 gender 0 1;
    contrast "age effect" age 1 -1;
    contrast "age effect" age 1 -1;
    contrast "gender effect" gender 1 -1;
    means gender age;
run;

```

$$+ \alpha_i$$

$$\mu + \frac{1}{2}(\beta_1 + \beta_2)$$

$$\mu + \alpha_1 + \beta_1$$

$$\mu + \alpha_1 + \beta_2$$

$$\mu + \alpha_2 + \beta_1$$

$$\mu + \alpha_2 + \beta_2$$

The SAS System  
The GLM Procedure  
Class Level Information

1

Class	Levels	Values
age	2	jr sr
gender	2	m w

Number of observations 28

NOTE: Due to missing values, only 15 observations can be used in this analysis.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	8318.06735	4159.03368	3.42	0.0669
Error	12	14606.86598	1217.23883		
Corrected Total	14	22924.93333			

R-Square	Coeff Var	Root MSE	y Mean
0.362839	16.05812	34.88895	217.2667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
age	1	325.629762	325.629762	0.27	0.6144
gender	1	7992.437592	7992.437592	6.57	0.0249

Source	DF	Type III SS	Mean Square	F Value	Pr > F
age	1	6044.348306	6044.348306	4.97	0.0457
gender	1	7992.437592	7992.437592	6.57	0.0249

Parameter	Estimate	Standard Error	t Value	Pr >  t
Intercept	213.1340206 B	12.77243521	16.69	<.0001
age jr	-59.0721649 B	26.50916484	-2.23	0.0457
age sr	0.0000000 B	.	.	.
gender m	67.9278351 B	26.50916484	2.56	0.0249
gender w	0.0000000 B	.	.	.

## Least Squares Means

gender	y LSMEAN	Standard Error	Pr >  t
m	251.525773	16.233482	<.0001
w	183.597938	15.842256	<.0001

age	y LSMEAN	Standard Error	Pr >  t
jr	188.025773	16.233482	<.0001
sr	247.097938	15.842256	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
age effect	1	6044.348306	6044.348306	4.97	0.0457
gender effect	1	7992.437592	7992.437592	6.57	0.0249

Parameter	Estimate	Standard Error	t Value	Pr >  t
lsmean for young folks	188.025773	16.2334818	11.58	<.0001
lsmean for older folks	247.097938	15.8422561	15.60	<.0001
lsmean for men	251.525773	16.2334818	15.49	<.0001
lsmean for women	183.597938	15.8422561	11.59	<.0001
lsmean for young men	221.989691	13.7197962	16.18	<.0001
lsmean for young women	154.061856	26.2714097	5.86	<.0001
lsmean for old men	281.061856	26.2714097	10.70	<.0001
lsmean for old women	213.134021	12.7724352	16.69	<.0001
age effect	-59.072165	26.5091648	-2.23	0.0457

$\mu + \alpha_i + \beta_j$   
 $\mu + \alpha_1 + \beta_2$