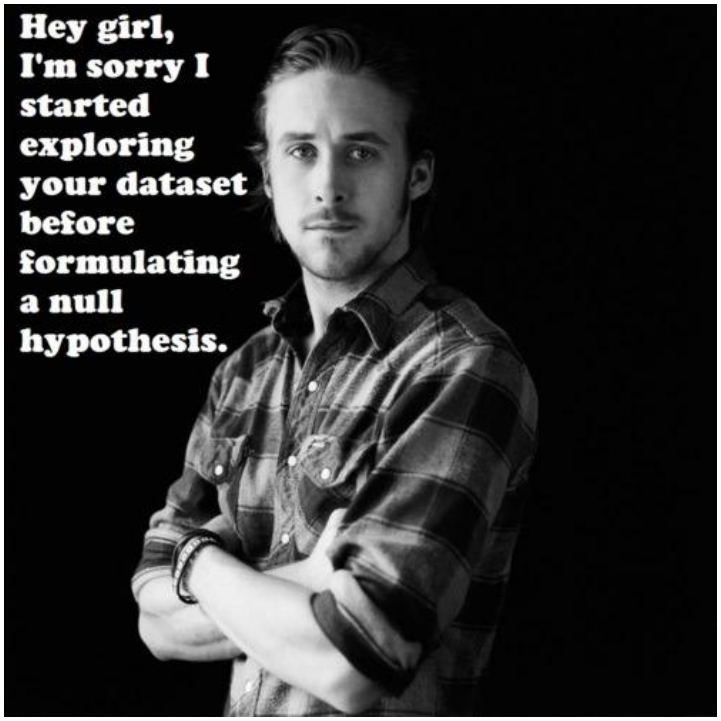


BIOS 6612
Lecture 12

General Linear Models III
Hypothesis Tests

Reading: Section 5 'General linear models' longitudinal course notes

Review (Lecture 11)/ Current (Lecture 12)/ Preview (Lecture 13)

- Lecture 11: General Linear Models II
  - Estimation
    - One-way effects model
    - Two-way effects model
  - Time as class versus continuous
  - Estimators: forms and properties
  - Standard errors and confidence intervals

- Lecture 12: General Linear Models III
  - Hypothesis tests and estimation in the general linear model framework
    - t-tests
    - F-tests
    - Main effect tests
  - Using SAS and R for custom tests and estimates

- Lecture 13: Introduction to Linear Mixed Models (LMM)
  - 2 measurements per subject
    - Paired t-test
    - Linear regression
  - Methods for repeated measurements
    - RMANOVA
    - Random effects
    - Covariance structures

Recall: Tests of General Linear Hypotheses

- BIOS 6611 Lecture 23-24 Categorical Variable and General Linear Hypothesis
  - General linear hypothesis:

$$H_0: \mathbf{L}\boldsymbol{\beta} = \mathbf{h}$$
$$H_1: \mathbf{L}\boldsymbol{\beta} \neq \mathbf{h}$$

- Notation:
  - H_0: $\mathbf{C}\boldsymbol{\beta} = \mathbf{h}$
    - Elements of **C** are constrained to sum to 0
    - Contrast: $\sum_{i=1}^{k} c_i = 0$

$$E[\mathrm{Y}] = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

    - Example:

$$H_0 : \begin{pmatrix} 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = 0 \Rightarrow H_0 : \beta_1 - \beta_2 = 0 \ \ or \ \ H_0 : \beta_1 = \beta_2$$

  - H_0: $\mathbf{L}\boldsymbol{\beta} = \mathbf{h}$
    - Elements of **L** are NOT constrained to sum to 0
    - Not a contrast: $\sum_{i=1}^{k} l_i \neq 0$

$$E[\mathrm{Y}] = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

    - For example:

$$H_0 : \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow H_0 : \begin{matrix} \beta_1 = 0 \\ \beta_2 = 0 \end{matrix}$$

Test of Linear Hypotheses: t-Tests

- Using methodology and notation presented in BIOS 6611 & 6612, we can construct a *t*-test for H_0: $\mathbf{C\beta} = 0$ using contrasts

$$t = C\hat{\boldsymbol{\beta}} / SE(C\hat{\boldsymbol{\beta}})$$

  - Which has a *t*-distribution with $n-k$ degrees of freedom under the null hypothesis
  - Where k= 1 (intercept) + #covariates in the model

- Note that $\mathbf{C\beta}$ is a scalar in this case

- Tests can be carried out in SAS using the ESTIMATE statement
  - Outputs
    - The *t*-test
    - The estimate $C\hat{\boldsymbol{\beta}}$
    - Its standard error
- These are mostly useful for comparing means between two specific groups
  - Main effect and interaction tests are more easily carried out using ANOVA

Example: Myostatin Data

- Myostatin protein is an inhibitor of skeletal muscle mass
- 2×3 factorial treatment structure in completely randomized design
  - 2 levels of treatment: myostatin Y or N; called 'group' variable
  - 3 levels of time: 24, 48 and 72 hours.
- Total of 24 muscle cell samples (4 replicates for each treatment)
- Outcome variable: measure of protein in the sample for the given condition (time and treatment)
  - Hypothesized that myostatin samples would have greater protein degradation than controls
- Table for population mean leucine protein levels for group*time combinations.

		Time			
		24h	48h	72h	
Group	C	μ_{11}	μ_{12}	μ_{13}	$\bar{\mu}_{1\bullet}$
	M	μ_{21}	μ_{22}	μ_{23}	$\bar{\mu}_{2\bullet}$
		$\bar{\mu}_{\bullet 1}$	$\bar{\mu}_{\bullet 2}$	$\bar{\mu}_{\bullet 3}$	

SAS program and output summary: Myostatin Example

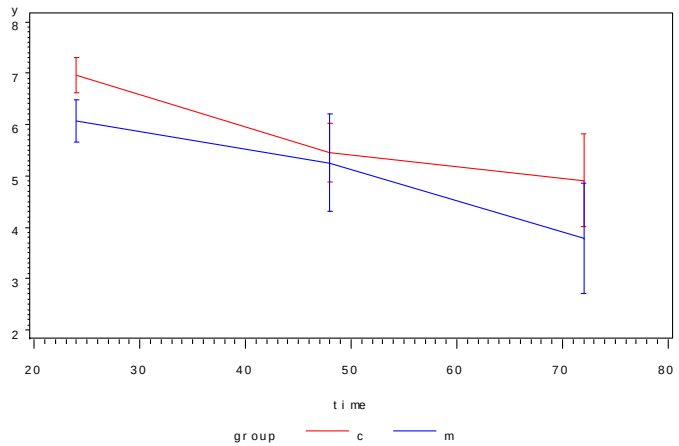
```
data myostatin;
input leucine group $ time @@;
y=leucine/1000; cards;
6568 c 24 6802 c 24 7198 c 24 7280 c 24
4992 c 48 5242 c 48 5285 c 48 6284 c 48
4092 c 72 4331 c 72 5135 c 72 6087 c 72
5516 m 24 6023 m 24 6334 m 24 6400 m 24
4512 m 48 4706 m 48 5175 m 48 6612 m 48
3076 m 72 3209 m 72 3462 m 72 5364 m 72
;
proc means data=myostatin noprint;
by group time; var y;
output out=myo_out mean=my stddev=sy n=ny;
run;
proc print data=myo_out;
var group time my sy ny; run;

proc gplot data=myostatin;
plot y*time=group / vaxis= 2 to 8;
symbol1 i=stdlmtj mode=include c=red;
symbol2 i=stdlmtj mode=include c=blue; run;
```

The output from PROC MEANS:

group	time	my	sy	ny
c	24	6.962	0.335	4
c	48	5.451	0.570	4
c	72	4.911	0.902	4
m	24	6.068	0.403	4
m	48	5.251	0.949	4
m	72	3.778	1.070	4

Graph from PROC GPLOT:



Two-way effects model: $Y_{ijk} = \mu + \alpha_i + \tau_j + \gamma_{ij} + \varepsilon_{ijk}$	group i time j replicate k	MODEL y=group|time;
One-way effects model: $Y_{ij} = \mu + \kappa_i + \varepsilon_{ij}$	group×time i replicate j	MODEL y=group*time;
Means model: $Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$	group i time j replicate k	MODEL y=group*time / noint;

Example: for the Myostatin data in the one-way effects model

$\mathbf{Y} = (Y_{11} \;\; Y_{12} \;\; Y_{13} \;\; Y_{14} \;\; Y_{21} \ldots Y_{61} \;\; Y_{62} \;\; Y_{63} \;\; Y_{64})^t$

$\boldsymbol{\beta} = (\mu \;\; \kappa_1 \;\; \kappa_2 \;\; \kappa_3 \;\; \kappa_4 \;\; \kappa_5 \;\; \kappa_6)^t$

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 1 & 0 & \ldots & 0 \\ 1 & 0 & 1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & 0 & 0 & 0 & \ldots & 1 \\ 1 & 0 & 0 & 0 & \ldots & 1 \\ 1 & 0 & 0 & 0 & \ldots & 1 \\ 1 & 0 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

Note that **X** does not have full rank (e.g., first column is the sum of the next 6). Thus, $(\mathbf{X}^t\mathbf{X})^{-1}$ does not exist. We'll need to use a generalized inverse.

Myostatin data in the one-way effects model

```
*One-way version;
proc glm data=myostatin;
class group time;
model y = group*time / solution;
run;
```

```
The GLM Procedure
Dependent Variable: y

                                 Sum of
Source                     DF       Squares     Mean Square    F Value    Pr > F
Model                       5    23.12640221      4.62528044       8.02    0.0004
Error                      18    10.37454375      0.57636354
Corrected Total            23    33.50094596

      R-Square     Coeff Var      Root MSE        y Mean
      0.690321      14.04979      0.759186      5.403542

                                                Standard
Parameter                  Estimate                 Error    t Value    Pr > |t|
Intercept             3.777750000 B      0.37959305       9.95      <.0001
group*time c 24       3.184250000 B      0.53682564       5.93      <.0001
group*time c 48       1.673000000 B      0.53682564       3.12      0.0060
group*time c 72       1.133500000 B      0.53682564       2.11      0.0490
group*time m 24       2.290500000 B      0.53682564       4.27      0.0005
group*time m 48       1.473500000 B      0.53682564       2.74      0.0133
group*time m 72       0.000000000 B       .                .         .

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal
equations.  Terms whose estimates are followed by the letter 'B' are not uniquely estimable.
```

t-test Example: Myostatin data in the one-way effects model

One-way effects model: $Y_{ij} = \mu + \kappa_i + \varepsilon_{ij}$

$Test\ \ H_0 : \kappa_1 - \kappa_2 = 0$

$E[\mathrm{Y}_i] = \mu + \kappa_1 * \mathrm{G}roupTime1_i + \kappa_2 * \mathrm{G}roupTime2_i + \kappa_3 * \mathrm{G}roupTime3_i$

$+\kappa_4 * \mathrm{G}roupTime4_i + \kappa_5 * \mathrm{G}roupTime5_i + \kappa_6 * \mathrm{G}roupTime6_i$

$$H_0 : \begin{pmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mu \\ \kappa_1 \\ \kappa_2 \\ \kappa_3 \\ \kappa_4 \\ \kappa_5 \\ \kappa_6 \end{pmatrix} = 0 \Rightarrow H_0 : \kappa_1 - \kappa_2 = 0$$

$$t = \frac{\mathbf{c}\hat{\boldsymbol{\beta}}}{\sqrt{\mathbf{c}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}'\hat{\sigma}^2_{Y|X}}}$$

$$c\hat{B} = 0 + (6.568 + 6.802 + 7.198 + 7.280)/4 - (4.992 + 5.242 + 5.285 + 6.284)/4 + 0 + 0 + 0 + 0 = 1.511$$

$$\sqrt{Var(\mathrm{c}\,\hat{B})} = \sqrt{\hat{\sigma}^2_{Y|X} \sum_{i=1}^{k} c_i^2 / n_i} = \sqrt{0.57636354 * [0 + 1^2/4 + (-1)^2/4 + 0 + 0 + 0 + 0]} = 0.537$$

$t = 1.511 / 0.537 = 2.82$; p=0.01

- We would conclude that for the Control group, there is protein degradation between 24 and 48 hours, on average.

```
> # one-way effects model
> mod2 <- lm(leucine/1000 ~ as.factor(time):group,data=myostatin)
> summary(mod2)

Call:
lm(formula = leucine/1000 ~ as.factor(time):group, data = myostatin)

Residuals:
    Min      1Q  Median      3Q     Max
-0.8193 -0.5470 -0.1629  0.2788  1.5862

Coefficients: (1 not defined because of singularities)
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                        3.7778     0.3796   9.952 9.62e-09 ***
as.factor(time)24:groupcontrol     3.1843     0.5368   5.932 1.30e-05 ***
as.factor(time)48:groupcontrol     1.6730     0.5368   3.116 0.005961 **
as.factor(time)72:groupcontrol     1.1335     0.5368   2.111 0.048975 *
as.factor(time)24:groupmyostatin   2.2905     0.5368   4.267 0.000464 ***
as.factor(time)48:groupmyostatin   1.4735     0.5368   2.745 0.013318 *
as.factor(time)72:groupmyostatin       NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7592 on 18 degrees of freedom
Multiple R-squared:  0.6903,    Adjusted R-squared:  0.6043
F-statistic: 8.025 on 5 and 18 DF,  p-value: 0.000396

> # t test example
> cvec <- c(0,1,-1,0,0,0)
> est <- cvec %*% coef(mod2)[1:6]
> # have to drop the last element since the covariance matrix of the estimates does not include row
/column for non-identifiable parameters
> se.est <- sqrt(t(cvec) %*% vcov(mod2) %*% cvec)
> # t stat
> est/se.est
         [,1]
[1,] 2.81516
> # p.value
> 2*pt(-abs(est/se.est),nrow(myostatin)-nrow(vcov(mod2)))
            [,1]
[1,] 0.01145815
```

Test of Linear Hypotheses: F-Tests

- Test statistic:

$$W = \frac{[SSE_{red} - SSE_{full}]/s}{SSE_{full}/(n-k)} = \frac{[\mathbf{Y}^t(\mathbf{P}_{full} - \mathbf{P}_{red})\mathbf{Y})]/s}{[\mathbf{Y}^t(\mathbf{I} - \mathbf{P}_{full})\mathbf{Y}]/(n-k)} \sim F_{s,n-k} \text{ under } H_0$$

- Notes
  - *red*=reduced model and *full*=full model
  - *SSE*=residual sum of squares
  - The denominator of W is the MSE
  - $\mathbf{P}_{full} = \mathbf{P}_{\mathbf{X}}$
  - $k = r(\mathbf{X}_{full})$, $s = r(\mathbf{X}_{full}) - r(\mathbf{X}_{red})$

- Three approaches to carrying out the test:
  (1) Employ PROC GLM (SAS) or LM function (R) directly.
  (2) Fit full and reduced models separately with PROC GLM / LM function and obtain the RSS quantities to calculate W
  (3) Work with projection matrices using PROC IML (or R)

- In SAS, we can conduct generalized likelihood ratio F-tests using the CONTRAST statement

- In R, it can be carried out using the *glh.test* function that is applied to a *glm* object
  - The function is available via the *gmodels* package

Example: Myostatin Data 2-way model (group and time as class variables)

Two-way effects model: $Y_{ijk} = \mu + \alpha_i + \tau_j + \gamma_{ij} + \varepsilon_{ijk}$	group i time j replicate k	MODEL y=group|time;

Write **X**, **Y** and **β** (for the two-way effects model).

The model: $Y_{ijk} = \mu + \alpha_i + \tau_j + \gamma_{ij} + \varepsilon_{ijk}$

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ . & . & . & & & & & & & & & \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{114} \\ y_{121} \\ y_{122} \\ y_{123} \\ y_{124} \\ \dots \\ y_{231} \\ y_{232} \\ y_{233} \\ y_{234} \end{pmatrix} = \begin{pmatrix} 6568 \\ 6802 \\ 7198 \\ 7280 \\ 4992 \\ 5242 \\ 5285 \\ 6284 \\ \dots \\ 3076 \\ 3209 \\ 3462 \\ 5364 \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{23} \end{pmatrix}$$

```
proc glm data=myostatin; class group time; model y = group|time / solution; run;
```

F-test Example: Myostatin Data 2-way model (group and time as class variables)

Question: do we need the interaction term?

- The null hypothesis: $H_0 : \gamma_{ij} = 0 \quad \forall\, i, j$

- The projection matrices:

$$\mathbf{P}_{full} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-}\mathbf{X}^t$$

  - where $\mathbf{X}$ is defined for the Myostatin data in the 2-way effects model including interaction

$$\mathbf{P}_{red} = \mathbf{X}_{red}(\mathbf{X}_{red}^t\mathbf{X}_{red})^{-}\mathbf{X}_{red}^t$$

  - where $\mathbf{X}_{red}$ is same as $\mathbf{X}$ without last 6 columns

- The full model has *group*, *time* and *group*time* as predictors, and the reduced model has just *group* and *time*

- The SSE for the full and reduced models are 10.375 and 11.316, respectively;
  $s = r(\mathbf{X}_{full}) - r(\mathbf{X}_{red}) = 2$ (the number of degrees of freedom for the interaction), and
  $k = r(\mathbf{X}_{full}) = 6$

  - Thus, W = {[11.316 – 10.375] / 2} / {10.375 / (24 – 6)} = 0.82 (p=0.45)
  - This matches the F-statistic generated by the CONTRAST statement
  - Based on the test, you could argue to drop the interaction term

Interaction F test in R

```
> # full model is
> mod1 <- lm(leucine/1000 ~ group*as.factor(time),data=myostatin)
> # reduced model is
> mod0 <- lm(leucine/1000 ~ group*as.factor(time) - group:as.factor(time),data=myostatin)
>
> # SSE is
> sse.red <- mod0$df.residual*summary(mod0)$sigma^2
> sse.full <- mod1$df.residual*summary(mod1)$sigma^2
>
> # degrees of freedom are
> df.red <- mod0$df.residual
> df.full <- mod1$df.residual
>
> # so F statistic is
> Fstat <- (sse.red - sse.full)/(df.red-df.full) /
+   (sse.full/df.full)
> Fstat
[1] 0.8165096
> # we are using RESIDUAL degrees of freedom here:
> # the difference between them is still equal to the difference in number of parameters between th
e reduced and full models
> # p-value is
> pf(Fstat,df.red-df.full,df.full,lower.tail=FALSE)
[1] 0.4576871
> # compare with interaction row of the ANOVA table for the full model
> anova(mod1)
Analysis of Variance Table

Response: leucine/1000
                     Df  Sum Sq Mean Sq F value    Pr(>F)
group                 1  3.3056  3.3056  5.7353   0.02772 *
as.factor(time)       2 18.8796  9.4398 16.3782 8.872e-05 ***
group:as.factor(time)  2  0.9412  0.4706  0.8165   0.45769
Residuals            18 10.3745  0.5764
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Main effect tests, interaction tests
More detail on CONTRAST and ESTIMATE statements

- Discussed theory for tests of the form H_0: $\mathbf{C\beta = h}$
  - What form of **C** and **h** are associated with tests of interest?
    - This will depend on the data at hand, and the specific hypotheses that the researcher is interested in testing.
  - Usually we use **h=0**.
- To illustrate the forms of **C**, consider the main effect tests for group and time, and the test for interaction, using the means model for the Myostatin data
- These tests are outputted directly without having to specify CONTRAST or ESTIMATE statements
  - In SAS, comes from Type III sum of squares table
  - In R, can use the anova() command on the model object
- But you can obtain them with the CONTRAST or ESTIMATE statements
  - Why?
  - To better understand both these particular tests as well as the statements

Contrasts

- In the strict sense
  - A CONTRAST is a linear combination of beta elements, $\mathbf{c}_i^t\boldsymbol{\beta}$, such that $\sum \mathbf{c}_i^t = 0$

- If all rows of **C** have this property
  - Then **Cβ** is a set of contrasts
  - Which are generally estimable

- When we estimate **Lβ** using the ESTIMATE statement
  - Elements of **L** are not constrained to sum to 0

- However, if **Lβ** is not estimable for the particular **L** that you specify
  - Then SAS will tell you that

- The **C** matrix may often be defined to have row contrasts
  - This was true in BIOS 6611
  - But generally in the class notes it will not be forced to have such constraints

Main effect test for group: Myostatin Data Means Model

Means model: remove intercept from one-way effects model $Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$	group i time j replicate k	MODEL y=group*time / noint;

Notation for the means model

		Time	
Trt Group	μ_{11}	μ_{12}	μ_{13}
	μ_{21}	μ_{22}	μ_{23}

For this means model, $\boldsymbol{\beta} = (\mu_{11}, \mu_{12}, \mu_{13}, \mu_{21}, \mu_{22}, \mu_{23})^t$

```
proc glm data=myostatin;
class group time;
model y = group*time   / solution noint;
run;
```

- The main effect test for group tests for differences in marginal means for groups
- For the application above, the test can be written as H_0: $\mathbf{C\beta} = 0$
  - where C = (1/3 1/3 1/3 -1/3 -1/3 -1/3)
- This also reduces to $H_0: \bar{\mu}_{1\bullet} = \bar{\mu}_{2\bullet}$
- We can add the following statements that will yield the same test results:

```
CONTRAST  'group factor' group*time 1  1  1  -1  -1 -1;
ESTIMATE  'group factor' group*time 1  1  1  -1  -1 -1 / divisor=3;
```

Main effect test for group using contrasts: R

```
> # means model is
> mod3 <- lm(leucine/1000 ~ 0+as.factor(time):group,data=myostatin)
> # putting group second in the model formula orders the means by group first, then time
> cvec <- c(rep(1,3),rep(-1,3))#/3
> # we can divide by 3 because this is a balanced design with 3 levels of the second factor (time)
but this factor will cancel out of the test statistic because it appears in numerator and denominat
or
> # estimate
> group.cont <- cvec %*% coef(mod3)
> # standard error
> se.group.cont <- sqrt(t(cvec) %*% vcov(mod3) %*% cvec)
> # t statistic is
> group.cont/se.group.cont
         [,1]
[1,] 2.394846
> # compare
> (group.cont/se.group.cont)^2
         [,1]
[1,] 5.735287
> pf((group.cont/se.group.cont)^2,1,mod3$df.residual,lower.tail=FALSE)
           [,1]
[1,] 0.02771845
> # with F statistic in group row of ANOVA table for full (two-factor effects) model
> anova(mod1)
Analysis of Variance Table

Response: leucine/1000
                      Df  Sum Sq Mean Sq F value    Pr(>F)
group                  1  3.3056  3.3056  5.7353   0.02772 *
as.factor(time)        2 18.8796  9.4398 16.3782 8.872e-05 ***
group:as.factor(time)  2  0.9412  0.4706  0.8165   0.45769
Residuals             18 10.3745  0.5764
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Notes:

- The CONTRAST statement produces an F-test
- The ESTIMATE statement produces a t-test
  - And an estimate corresponding to the coefficients specified
- In many cases these will produce the same results
  - If the data and coefficients are the same
- For either the CONTRAST or ESTIMATE approach, the test will be the same if the coefficients are all scaled by the same amount
  - Because the scalar will cancel out in the test statistic numerator and denominator
  - Rescaling will change the estimate but will not change the either the *t*-test invoked by the ESTIMATE statement or the *F*-test invoked by the CONTRAST statement
- The divisor is often used to simplify the code
  - Becomes important if numbers have unending decimals (e.g., 0.333…)
  - The CONTRAST does not have the divisor option as it is not important
  - It is really only necessarily for the estimate in the ESTIMATE statement, not the test
- Generally, **C** will have a different form and dimensions depending on the model used
  - Means model, one-way effects model, two-way effects model

Main Effect Test for Time

Means model: $Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$	group i time j replicate k	MODEL y=group*time / noint;

For this means model, $\boldsymbol{\beta} = (\mu_{11}, \mu_{12}, \mu_{13}, \mu_{21}, \mu_{22}, \mu_{23})^t$

- The main effect test for time tests for differences in marginal means for time
  - In this case there are 3 times
    - Thus the test is $H_o: \bar{\mu}_{\bullet 1} = \bar{\mu}_{\bullet 2} = \bar{\mu}_{\bullet 3}$
  - In this case we will need 2 rows in the **C** matrix for the test
    - One line for each equation in the hypothesis

- There are different possibilities but one is:

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & 1 & -1 & 0 \\ 1 & 0 & -1 & 1 & 0 & -1 \end{pmatrix}$$

- The first line of $\mathbf{C\beta}=\mathbf{0}$ is
  - $\mu_{11} - \mu_{12} + \mu_{21} - \mu_{22} = 0$, or $\frac{1}{2}\mu_{11} + \frac{1}{2}\mu_{21} = \frac{1}{2}\mu_{12} + \frac{1}{2}\mu_{21}$, or more simply, $\bar{\mu}_{\bullet 1} = \bar{\mu}_{\bullet 2}$
- Similarly, the second line is $\bar{\mu}_{\bullet 1} = \bar{\mu}_{\bullet 3}$
- Note that $\bar{\mu}_{\bullet 2} = \bar{\mu}_{\bullet 3}$ is implied through the other equalities

- We can carry out the test with the following statement in SAS:

```
CONTRAST  'time factor' group*time 1  -1   0  1  -1  0,
                        group*time 1   0  -1  1   0  -1;
```

Main Effect Test for Time

- Can do this in R directly using the C matrix and formula for the F statistic

```
> cmat <- rbind(c(1,-1,0,1,-1,0), # mean at time 1 = mean at time 2
+               c(1,0,-1,1,0,-1)) # mean at time 1 = mean at time 3
>
> # estimates
> time.cont <- cmat %*% coef(mod3)
> # covariance matrix
> v.time.cont <- cmat %*% vcov(mod3) %*% t(cmat)
> # (formula is F = (C Beta-hat - d)' ( C (X'X)^-1 C' ) (C Beta-hat - d) / r / ( SSE / (n-p) ), but
SSE/(n-p) = MSE and the covariance matrix of the estimates is MSE*(X'X)^-1)
> Fstat <- (t(time.cont) %*% solve(v.time.cont) %*% time.cont)/2
> Fstat
          [,1]
[1,] 16.37819
> pf(Fstat,2,mod3$df.residual,lower.tail=FALSE)
              [,1]
[1,] 8.872323e-05
> # compare with time row of ANOVA table for full model
> anova(mod1)
Analysis of Variance Table

Response: leucine/1000
                      Df  Sum Sq Mean Sq F value    Pr(>F)
group                  1  3.3056  3.3056  5.7353   0.02772 *
as.factor(time)        2 18.8796  9.4398 16.3782 8.872e-05 ***
group:as.factor(time)  2  0.9412  0.4706  0.8165   0.45769
Residuals             18 10.3745  0.5764
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- … Or using the gmodels package

```
library(gmodels)

glm2<-glm(y~gt-1,data=myostatin) #Note! No-intercept model
```

```
summary(glm2)

#F-test for Time
    C<-rbind(c(1,0,-1,1,0,-1),c(1,-1,0,1,-1,0))

   mycontrast<-glh.test(glm2,C)

 summary(mycontrast)

#C matrix:
#     gtc24 gtc48 gtc72 gtm24 gtm48 gtm72
#[1,]     1     0    -1     1     0    -1
#[2,]     1    -1     0     1    -1     0

#C %*% Beta-hat:
#[1] 4.34125 2.32825

#F = 16.3782, df1 =  2, df2 = 18, p-value = 8.872e-05
```

Notes

- Since there are 2 d.f.
  - The test can NOT be carried out using the ESTIMATE statement
- Other forms of **C** may yield the same test
  - This can be explained by the *Full Rank Reparameterization Theorem* which states that

$$SS\left(\underset{q\times p}{\mathbf{C}}\ \underset{p\times 1}{\hat{\boldsymbol{\beta}}}\right) = SS\left(\underset{q\times q}{\mathbf{D}}\ \underset{q\times p}{\mathbf{C}}\ \underset{p\times 1}{\hat{\boldsymbol{\beta}}}\right) \text{ for any nonsingular } \underset{q\times q}{\mathbf{D}}$$

- **Key note:**
  - CONTRAST statements may have multiple rows
  - But ESTIMATE statements are restricted to one row

Time*group interaction: Myostatin means model

- Does the difference between group means depend on time?
  - If so, then there is interaction
  - Similarly, you can ask the question whether differences over time are similar between groups

- If the differences between group means is in fact the same at each time, then there is no interaction and this would comprise the null hypothesis:

$$H_0: \mu_{11} - \mu_{21} = \mu_{12} - \mu_{22} = \mu_{13} - \mu_{23}$$

- The **C** matrix associated with this hypothesis is:

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & -1 & 1 & 0 \\ 1 & 0 & -1 & -1 & 0 & 1 \end{pmatrix}$$

- As with the main effect tests, the interaction test will be part of the default output

- The test can also be carried out with the following added statement:

```
CONTRAST 'interact' group*time 1 -1 0 -1 1 0, group*time 1 0 -1 -1 0 1;
```

- Again, d.f.>1 so cannot get the test with an ESTIMATE statement

Interaction contrast: test in R

```
> cmat <- rbind(c(1,-1,0,-1,1,0), # mean at time 1 = mean at time 2
+               c(1,0,-1,-1,0,1)) # mean at time 1 = mean at time 3
>
> # estimates
> interxn.cont <- cmat %*% coef(mod3)
> # covariance matrix
> v.interxn.cont <- cmat %*% vcov(mod3) %*% t(cmat)
> # (formula is F = (C Beta-hat - d)' ( C (X'X)^-1 C' ) (C Beta-hat - d) / r / ( SSE / (n-p) ), but
SSE/(n-p) = MSE and the covariance matrix of the estimates is MSE*(X'X)^-1)
> Fstat <- (t(interxn.cont) %*% solve(v.interxn.cont) %*% interxn.cont)/2
> Fstat
           [,1]
[1,] 0.8165096
> pf(Fstat,2,mod3$df.residual,lower.tail=FALSE)
           [,1]
[1,] 0.4576871
> # compare with interaction row of ANOVA table for full model
> anova(mod1)
Analysis of Variance Table

Response: leucine/1000
                     Df  Sum Sq Mean Sq F value    Pr(>F)
group                 1  3.3056  3.3056  5.7353   0.02772 *
as.factor(time)       2 18.8796  9.4398 16.3782 8.872e-05 ***
group:as.factor(time)  2  0.9412  0.4706  0.8165   0.45769
Residuals            18 10.3745  0.5764
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Orthogonality of Contrasts

- For the **C** matrices, pairs of rows are *orthogonal* if
  - For row vectors **c** and **d**, $\sum_j c_j d_j = 0$, where j denotes the j^{th} element of **c** or **d**
  - As a consequence, Type III sums of squares for the 3 factors will add up nicely to the total sum of squares, for these data since sample sizes are equal across treatments (or cells)
  - Independence of tests also follows from orthogonality of the contrasts
- Example

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}$$

- Not Orthogonal (1*1+-1*0+0*-1+0*0=1 (row 1 and row2))

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & -1 & -1 \end{pmatrix}$$

- Orthogonal