Displays for Statistics 5401/8401

Lecture 23

October 31, 2005

Christopher Bingham, Instructor

612-625-1024, kb@umn.edu
372 Ford Hall

Class Web Page

http://www.stat.umn.edu/~kb/classes/5401





# Model for Means in Multi-Group Profile Analysis

The setting is g independent samples of repeated measures data with group mean vectors $\boldsymbol{\mu}_j = [\mu_{\ell j}]$, $j = 1, \ldots g$

In a g by p factorial experiment, you usually represent treatment means as

$$\mu_{\ell j} = \mu + \alpha_j + \beta_\ell + (\alpha\beta)_{\ell j},$$
$$j = 1, 2, \ldots, g,\ \ell = 1,\ldots,p.$$

with "side conditions",

$$\sum_{1\le j\le g}\alpha_j = 0 \qquad \sum_{1\le \ell\le p}\beta_\ell = 0$$
$$\sum_{1\le j\le g}(\alpha\beta)_{j\ell} = 0,\ \ell = 1,\ldots,p$$
$$\sum_{1\le \ell\le p}(\alpha\beta)_{j\ell} = 0,\ j = 1,\ldots,g$$

The group mean vectors are

$$\boldsymbol{\mu}_j = \mu\mathbf{1}_p + \alpha_j\mathbf{1}_p + \boldsymbol{\beta} + (\boldsymbol{\alpha\beta})_j$$
$$\boldsymbol{\beta} = [\beta_1,\ldots,\beta_p]'$$
$$(\boldsymbol{\alpha\beta})_j = [(\alpha\beta)_{1j},\ \ldots,\ (\alpha\beta)_{pj}]'$$

with $\mathbf{1}_p'\boldsymbol{\beta} = \mathbf{1}_p'(\boldsymbol{\alpha\beta})_j = \mathbf{0}$

- The $\alpha_j$ are **between-subjects main effects**, one for each group
- The $\beta_\ell$'s are **within-subjects main effects**, one for each variable
- $(\alpha\beta)_{\ell j}$'s are **interaction effects**. They determine the pattern of interaction between the factors.





## No interaction situation

When $(\alpha\beta)_{\ell j} = 0$ for all $j$ and $\ell$,

- $\mu_{\ell j} = \mu + \alpha_j + \beta_\ell$: effects are <u>additive</u>

$$\boldsymbol{\mu}_j = \mu\mathbf{1}_p + \alpha_j\mathbf{1}_p + \boldsymbol{\beta}$$

- Between <u>group</u> contrasts defined by $\mathbf{c} = \{c_j\}$, $\sum_{1\le j\le g} c_j = 0$, don't depend on $\ell$:

$$\sum_{1\le j\le g} c_j\mu_{\ell j} = \sum_{1\le j\le g} c_j\alpha_j \text{ for all } \ell$$

- Between <u>variable</u> contrasts defined by $\{d_\ell\}$, $\sum_{1\le \ell\le p} d_\ell = 0$, don't depend on group:

$$\mathbf{d}'\boldsymbol{\mu}_j = \sum_{1\le \ell\le p} d_\ell\mu_{\ell j} = \sum_{1\le \ell\le p} d_\ell\beta_\ell = \mathbf{d}'\boldsymbol{\beta},\ 1\le j\le g$$

In particular, pairwise differences are uniquely defined

$$\mu_{\ell j} - \mu_{\ell k} = \alpha_j - \alpha_k,\ \ell = 1, 2, \ldots, p$$

and

$$\mu_{\ell j} - \mu_{mj} = \beta_\ell - \beta_m,\ j = 1, \ldots, g$$

So, when there is no interaction, $\{\alpha_j\}$ and $\{\beta_\ell\}$ completely describe the effects of the factors and, with $\mu$, define $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_g$.

## Interaction present

When some $(\alpha\beta)_{\ell j} \neq 0$, there is interaction between the factors. This implies that

- at least one between group difference
  $$\mu_{\ell j} - \mu_{\ell k} = \alpha_j - \alpha_k + (\alpha\beta)_{\ell j} - (\alpha\beta)_{\ell k}$$
  depends on the level $\ell$ of the within subject factor
- at least one within subject difference
  $$\mu_{\ell j} - \mu_{mj} = \beta_\ell - \beta_m + (\alpha\beta)_{\ell j} - (\alpha\beta)_{mj}$$
  depends on the level j of the between subject factor.

In particular, when there is interaction, it means

- *both* factors have effects
- the effects of a factor are not unique but depend on the level of the other factor.



Statistics 5401 Lecture 23 October 31, 2005

## No interaction ⟺ parallel profiles

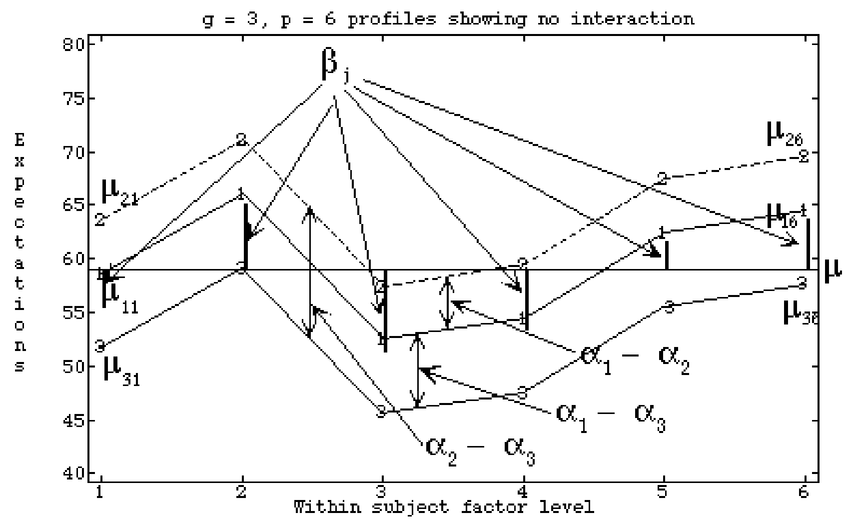
No interaction means additivity:

$$\mu_{\ell j} = \mu + \alpha_j + \beta_\ell$$

Using vectors, this means

$$\boldsymbol{\mu}_j = \mu \mathbf{1}_p + \alpha_j \mathbf{1}_p + \boldsymbol{\beta},\ \boldsymbol{\beta} = [\beta_1,\ldots,\beta_p]'$$
$$j = 1, \ldots, g$$

Geometrically, the g graphs or *profiles* of $\mu_{j\ell}$ vs $\ell$ are *parallel*. with shape set by $\boldsymbol{\beta}$ and height determined by $\mu + \alpha_j$.




p = 6 and g = 3 in this example.

## Test of zero interaction

When at least one $(\alpha\beta)_{\ell j} \neq 0$, that is $H_0$: $(\alpha\beta)_{\ell j} \equiv 0$ is false, then least two profiles are not parallel.

If $\mathbf{C}$ is a full rank p-1 by p matrix defining p-1 within-subject *contrasts* ($\mathbf{C1}_p = \mathbf{0}$), then

$$\mathbf{C}\boldsymbol{\mu}_j \neq \mathbf{C}\boldsymbol{\mu}_k, \text{ some } j \neq k$$

**Conclusion**:

$$H_0: (\alpha\beta)_{\ell j} = 0, \text{ all } j, \ell$$

means the same as

$$H_0: \boldsymbol{\nu}_1 = \boldsymbol{\nu}_2 = \ldots = \boldsymbol{\nu}_g, \boldsymbol{\nu}_j \equiv \mathbf{C}\boldsymbol{\mu}_j, j = 1, \ldots, g$$

Here $\mathbf{C}$ is a matrix of g-1 contrasts like

$$\begin{bmatrix} -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{bmatrix} \text{ or } \begin{bmatrix} -1 & 1 & 0 & \ldots & 0 \\ -1 & 0 & 1 & \ldots & 0 \\ . & . & . & . & . \\ -1 & 0 & 0 & \ldots & 1 \end{bmatrix}$$





But you know how to test the hypothesis that multivariate means are equal.

Define new p - 1 dimensional random vectors

$$\mathbf{w}_{ij} = \mathbf{C}\mathbf{y}_{ij}, i = 1,\ldots,n_j, j = 1,\ldots g$$

with means

$$E[\mathbf{w}_{ij}] = \boldsymbol{\mu}_{wj} = \mathbf{C}\boldsymbol{\mu}_j = \boldsymbol{\nu}_j, j = 1,\ldots g$$

The $\mathbf{W} = \mathbf{YC}'$ is a N by p-1 data matrix whose rows are $\mathbf{w}_{ij}'$, $i = 1,\ldots,n_j$, $j = 1,\ldots g$.

You can test equality of mean vectors by MANOVA with $\mathbf{W}$ as data.

Note that the number of dimensions is now $q = p - 1$ rather than p.

Provided $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \ldots = \boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}_y$ is constant, $\boldsymbol{\Sigma}_w = \mathbf{C}\boldsymbol{\Sigma}_y\mathbf{C}'$ (q by q matrix) is constant.

You don't actually need to compute $\mathbf{W} = \mathbf{YC}'$ since you can compute everything from the results of MANOVA on $\mathbf{Y}$:

- $\mathbf{H}_w = \mathbf{CH}_y\mathbf{C}'$
- $\mathbf{E}_w = \mathbf{CE}_y\mathbf{C}'$

When g = 2, you can use Hotelling's $T^2$:

$$T^2 = (\overline{\mathbf{w}}_{.1} - \overline{\mathbf{w}}_{.2})'\hat{V}[\overline{\mathbf{w}}_{.1} - \overline{\mathbf{w}}_{.2}]^{-1}(\overline{\mathbf{w}}_{.1} - \overline{\mathbf{w}}_{.2})$$
$$= (\mathbf{C}(\overline{\mathbf{y}}_{.1} - \overline{\mathbf{y}}_{.2}))'\{\mathbf{C}\hat{V}[\overline{\mathbf{y}}_{.1} - \overline{\mathbf{y}}_{.2}]\mathbf{C}'\}^{-1}(\mathbf{C}(\overline{\mathbf{y}}_{.1} - \overline{\mathbf{y}}_{.2}))$$

where

$$\hat{V}[\overline{\mathbf{y}}_{.1} - \overline{\mathbf{y}}_{.2}] = (1/n_1 + 1/n_2)\mathbf{S}_{pooled}$$

The null distribution is

$$(qf_e)F_{q,f_e-q+1}/(f_e - q + 1),\ q = p - 1$$

Substituting p-1 for q,

$$(p-1)f_eF_{p-1,f_e-p+2}/(f_e-p+2)$$





When you reject the parallelism hypothesis, you can conclude that $\mathbf{C}\boldsymbol{\mu}_j \neq \mathbf{0}$ for at least one j, so the within subject treatment has some effect.

**Example**:

Data from an experiment comparing the effects of g = 3 doses of vitamin E on the growth of rats over 7 weeks (p = 6 because week 2 was skipped).

```
Cmd> data <- read("","acrtab6.8")
acrtab6.8       15      7 format
) Data from Table 6.8 of Methods of Multivariate Analysis,
) Alvin C. Rencher, Wiley 1995, p. 223
) Data from an experiment comparing 3 vitamin E supplements for
) their effect on the growth of guinea pigs.  Weight recorded at
) the end of weeks 1, 3, 4, 5, 6 and 7
) Col. 1: Group (1, 2, 3 = zero, low, high vitamin E)
) Col. 2: Weight after week 1
) Col. 3: Weight after week 3
) Col. 4: Weight after week 4
) Col. 5: Weight after week 5
) Col. 6: Weight after week 6
) Col. 7: Weight after week 7
Read from file "TP1:Stat5401:Data:guinea.dat"

Cmd> group <- factor(data[,1])

Cmd> y <- data[,-1]
```

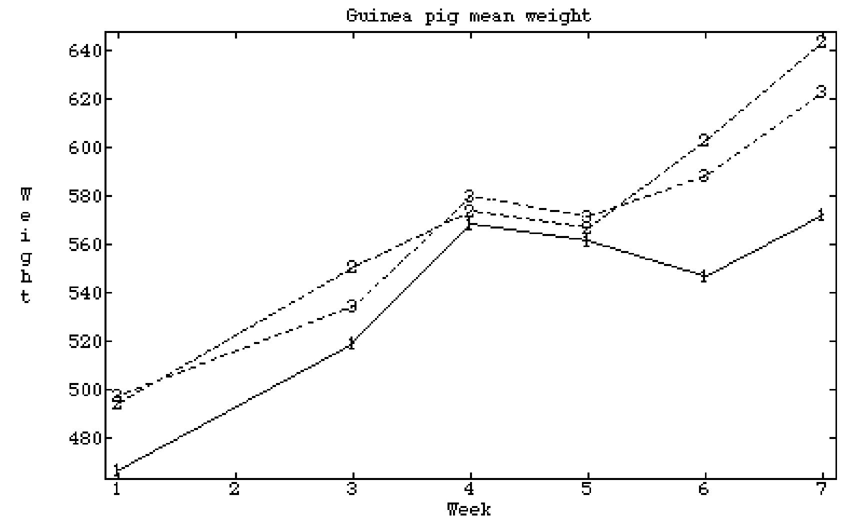
The within-subject factor is time (week). The between-subject factor is vitamin E.

```
Cmd> weeks <- vector(1,3,4,5,6,7) #omits week 2

Cmd> weeklabs <- vector("Week 1","Week 3","Week 4","Week 5",\
        "Week 6","Week 7")

Cmd> setlabels(y,structure("@", weeklabs))

Cmd> list(group,y)
group           REAL   15    1     FACTOR with 3 levels
y               REAL   15    6     (labels)

Cmd> p <- ncols(y); p
(1)          6

Cmd> addmacrofile("") # make sure new Mulvar.mac is available

Cmd> manova("y=group")
Model used is y=group
WARNING: summaries are sequential
NOTE: SS/SP matrices suppressed because of size; use
'manova(,sssp:T)'
                          SS and SP Matrices
                DF
CONSTANT         1
                    Type 'SS[1,,]' to see SS/SP matrix
group            2
                    Type 'SS[2,,]' to see SS/SP matrix
ERROR1          12
                    Type 'SS[3,,]' to see SS/SP matrix

Cmd> h <- matrix(SS[2,,]); fh <- DF[2]

Cmd> e <- matrix(SS[3,,]); fe <- DF[3]

Cmd> print(h, e, format:"7.1f")
h:
        Week 1  Week 3  Week 4  Week 5  Week 6  Week 7
Week 1  2969.2  2177.2   859.4   813.0  4725.2  5921.6
Week 3  2177.2  2497.6   410.0   411.6  4428.8  5657.6
Week 4   859.4   410.0   302.5   280.4  1132.1  1392.5 = H
Week 5   813.0   411.6   280.4   260.4  1096.4  1352.0
Week 6  4725.2  4428.8  1132.1  1096.4  8550.9 10830.9
Week 7  5921.6  5657.6  1392.5  1352.0 10830.9 13730.1
e:
        Week 1  Week 3  Week 4  Week 5  Week 6  Week 7
Week 1  8481.2  8538.8  4819.8  8513.6  8710.0  8468.2
Week 3  8538.8 17170.4 13293.0 19476.4 17034.2 20035.4 = E
Week 4  4819.8 13293.0 12992.4 17077.4 17287.8 17697.2
Week 5  8513.6 19476.4 17077.4 28906.0 26226.4 28625.2
Week 6  8710.0 17034.2 17287.8 26226.4 36898.0 31505.8
Week 7  8468.2 20035.4 17697.2 28625.2 31505.8 33538.8
```





```
Cmd> means <- tabs(y,group,mean:T)' # Note the transpose

Cmd> list(means)
means           REAL   6     3

Cmd> lineplot(weeks,means,symbols:run(3),\
    title:"Guinea pig mean weight",xlab:"Week",ylab:"Weight")
```




Are these significantly non-parallel?

Are these significantly non-linear?

```
Cmd> c <- matrix(vector(-1,1,0,0,0,0, -1,0,1,0,0,0, \
    -1,0,0,1,0,0, -1,0,0,0,1,0, -1,0,0,0,0,1),6)'

Cmd> print(c,format:"4.0f") #contrasts, comparisons with wk 1
c:
(1,1)   -1    1    0    0    0    0   q = p-1 = 5 contrasts
(2,1)   -1    0    1    0    0    0   with week 1
(3,1)   -1    0    0    1    0    0
(4,1)   -1    0    0    0    1    0
(5,1)   -1    0    0    0    0    1

Cmd> chc <- c %*% h %*% c'            C H C'

Cmd> cec <- c %*% e %*% c'            C E C'

Cmd> q <- ncols(y) - 1 # or nrows(chc) = 5

Cmd> list(chc,cec) # chc and cec are q by q
cec             REAL    5     5     (labels)
chc             REAL    5     5     (labels)

Cmd> releigenvals(chc,cec)# s = 2 non-zero relative eigenvalues
(1)      2.6682     0.52252  2.2917e-16 -2.0267e-16 -2.3788e-15

Cmd> cumwilks(det(cec)/det(chc + cec),fh,fe,q)
(1)    0.079316          Exact P-value for Wilks since s = 2

Cmd> cumtrace(trace(solve(cec,chc)),fh,fe,q,upper:T)
(1)    0.092277          P-value for Hotelling trace

Cmd> cumtrace(trace(solve(cec+chc,chc)),fh,fe,q,\
        pillai:T,upper:T)
(1)     0.06563          P-value for Pillai trace
```

**Roy test**: Simulation gave $P \cong .101$ for $\hat{\theta}_{max} = \hat{\lambda}_1/(1+\hat{\lambda}_1) = 2.6682/(1+2.6682) = 0.7274$. The exact P-value is 0.0965.

**Conclusion**: There is no convincing evidence any $(\alpha\beta)_{ij} \neq 0$ and hence the profiles are apparently parallel.





## Within-subject main effect test

When there is no interaction, main effects are well defined. Since $\mathbf{C1}_p = \mathbf{0}$

$$\mathbf{C}\boldsymbol{\mu}_j = (\mu+\alpha_j)\times\mathbf{C1}_p + \mathbf{C}\boldsymbol{\beta} = \mathbf{C}\boldsymbol{\beta},\ j = 1, \ldots, g$$

That is, all $N = n_1 + n_2 + \ldots + n_g$ vectors

$$\mathbf{w}_{11},\ldots,\ \mathbf{w}_{n_1 1},\ \ldots,\mathbf{w}_{1g},\ldots,\ \mathbf{w}_{n_g g}$$

have the same mean.

Now $H_0: \beta_1 = \beta_2 = \ldots = \beta_p$ (no within subject main effect) is equivalent to

$$H_0: E[\mathbf{w}_{ij}] = \mathbf{C}\boldsymbol{\beta} = 0.$$

You can test $H_0$ by Hotelling's $T^2$, treating all N $\mathbf{w}_{ij}$'s as a sample with sample mean $\overline{\overline{\mathbf{w}}} = \mathbf{C}\overline{\overline{\mathbf{y}}}$, where $\overline{\overline{\mathbf{y}}} = \sum_{1\leq j\leq g} n_j \overline{\mathbf{y}}_j / N$.

The estimated variance matrix of $\overline{\overline{\mathbf{y}}}$ is

$$\hat{V}[\overline{\overline{\mathbf{y}}}] = (1/N)\mathbf{S},\ \mathbf{S} = (1/f_e)\mathbf{E},\ f_e = N - g$$

so the estimated variance matrix of $\overline{\overline{\mathbf{w}}}$ is

$$\hat{V}[\overline{\overline{\mathbf{w}}}] = (1/N)\mathbf{CSC}' = (1/N)\mathbf{C}(f_e^{-1}\mathbf{E})\mathbf{C}'.$$

Then

$$T^2 = \overline{\overline{\mathbf{w}}}'\{\hat{V}(\overline{\overline{\mathbf{w}}})\}^{-1}\overline{\overline{\mathbf{w}}} = (\mathbf{C}\overline{\overline{\mathbf{y}}})'\{(1/N)\mathbf{CSC}'\}^{-1}\mathbf{C}\overline{\overline{\mathbf{y}}},$$

with small sample null distribution

$$\{qf_e/(f_e - q + 1)\}F_{q,f_e-q+1},\ q = p-1, f_e = N - g.$$

This provides a test of $H_0$: within-subject factor "main effects" are zero, *assuming no interaction*.

```
Cmd> ybar <- describe(y,mean:T) # grand mean

Cmd> N <- nrows(y); N
(1)          21

Cmd> s <- e/fe # = S_pooled

Cmd> vhat_ybar <- (1/N)*s # estimated V[ybar]

Cmd> vhat_wbar <- c %*% vhat_ybar %*% c' # estimated V[wbar]

Cmd> tsq <- (c %*% ybar)' %*% solve(vhat_wbar) %*% (c %*% ybar)

Cmd> tsq # Hotelling's T^2
             (1)
(1)      297.13

Cmd> fstat <- (fe-q-1)*tsq/(q*fe); vector(fstat, q, fe-q+1)
(1)      29.713           5          8
                    Num df    Denom df

Cmd> cumF(fstat, q, fe-q+1,upper:T) # P-value
(1,1)   5.7779e-05 => significant at any reasonable level
```

You can reject $H_0$: $\boldsymbol{\beta} = 0$.

There is significant difference between the 6 time points.





## Testing between-group main effect

When the profiles are <u>parallel</u>, how do you test the hypothesis that the between-subjects factor has no effect

$$H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \ldots = \boldsymbol{\mu}_g?$$

Since parallelism means

$$\boldsymbol{\mu}_j = \mu\mathbf{1}_p + \alpha_j\mathbf{1}_p + \boldsymbol{\beta}$$

$H_0$ is equivalent to a univariate

$$H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_g = 0$$

for which $\boldsymbol{\mu}_j = \mu\mathbf{1}_p + \boldsymbol{\beta}$, j = 1, 2, ..., g.

MANOVA on $\mathbf{y}$ tests the same hypothesis ($H_0$: $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \ldots = \boldsymbol{\mu}_g$) <u>without</u> assuming an additive structure.

With parallelism, the problem becomes <u>univariate</u>.

The averages <u>across variables</u>

$$x_{ij} \equiv \overline{y}_{ij} = \sum_{1\leq \ell \leq p} y_{\ell ij}/p = (1/p)\mathbf{1}_p'\mathbf{y}_{ij}$$

have means

$$\mu_{x_j} = \mathbf{1}_p'\boldsymbol{\mu}_j/p = \mu + (1/p)\mathbf{1}_p'\boldsymbol{\beta} + \alpha_j$$
$$= \mu + \alpha_j\text{, since } \mathbf{1}_p'\boldsymbol{\beta} = 0$$

When $\alpha_1 = \ldots = \alpha_g$, these are all the same, that is

$$\mu_{x_1} = \mu_{x_2} = \ldots = \mu_{x_g}$$

You can test this by a univariate ANOVA using the averages $x_{ij}$ across variables as data.





```
Cmd> x <- describe(y',mean:T); x # subject means
Cmd> hconcat(group,x) # group and mean for each subject
 (1,1)           1      471.83
 (2,1)           1      561.17
 (3,1)           1      558.83
 (4,1)           1      571.17
 (5,1)           1      532.67
 (6,1)           2         552
 (7,1)           2      512.67
 (8,1)           2      578.67
 (9,1)           2      621.67
(10,1)           2      596.33
(11,1)           3      600.33
(12,1)           3      560.17
(13,1)           3      548.67
(14,1)           3       595.5
(15,1)           3      524.83
```

You now use **x** as the vector of responses in a univariate ANOVA.

```
Cmd> anova("x=group",fstat:T)
Model used is x=group
             DF          SS          MS            F     P-value
CONSTANT      1  4.6889e+06  4.6889e+06   3201.99930     < 1e-08
group         2      3091.3      1545.7      1.05552     0.37821
ERROR1       12       17572      1464.4
```

**Conclusion**: Since P = .37821 > .10 there is no significant effect of vitamin E, that is we cannot reject $H_0$: $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3$

**Reminder** This interpretation assumes no interaction.

You can also get the SS directly from **H** and **E**.

Because $x_{ij} = \mathbf{d}'\mathbf{y}_{ij}$ with $\mathbf{d} = (1/p)\mathbf{1}_p$,

$$SS_h = \mathbf{d}'\mathbf{H}\mathbf{d} = (\sum_\ell \sum_m h_{\ell m})/p^2,$$
$$SS_e = \mathbf{d}'\mathbf{E}\mathbf{d} = (\sum_\ell \sum_m e_{\ell m})/p^2$$

```
Cmd> p <- nrows(h)

Cmd> vector(sum(vector(h))/p^2, sum(vector(e))/p^2)
(1)      3091.3       17572

Cmd> d <- rep(1,p)/p

Cmd> vector(d' %*% h %*% d, d' %*% e %*% d)
          (1)          (2)
       3091.3        17572
```





When g = 2, you can test equality of means assuming parallelism by a univariate two sample **t-statistic**

$$t = (\overline{x}_1 - \overline{x}_2)/\sqrt{\{(1/n_1+1/n_2)(\sum_\ell \sum_m s_{\ell m}/p^2)\}}$$
$$= (\overline{x}_1 - \overline{x}_2)/\sqrt{\{(1/n_1+1/n_2)(\mathbf{1}_p'\mathbf{S}\mathbf{1}_p/p^2)\}}$$

where

- $\overline{x}_j = \sum_{1\leq i\leq n_j} \sum_{1\leq \ell\leq b} y_{\ell ij}/(pn_j)$, j = 1, 2
- $s_{\ell m}$ are the elements of $\mathbf{S} = \mathbf{S}_{pooled} = \mathbf{E}/f_e$.

When $H_0$ is true, t is distributed as Student's $t_{f_e} = t_{n_1+n_2-2}$.

The plot of means against week are close to linear, as well as being parallel.

How can you test the null hypothesis that the means change linearly with time?

A contrast that is linear in weeks is $\mathbf{c}_{lin}$ = [-10, -4, -1, 2, 5, 8]'. If you can find 4 more linearly independent contrasts $\mathbf{c}_2$, $\mathbf{c}_3$, $\mathbf{c}_4$, $\mathbf{c}_5$ that are orthogonal to $\mathbf{c}_{lin}$, then, still assuming parallelism, the hypothesis of linearity is equivalent to

$$H_0: \mathbf{c}_2'\boldsymbol{\beta} = \mathbf{c}_3'\boldsymbol{\beta} = \mathbf{c}_4'\boldsymbol{\beta} = \mathbf{c}_5'\boldsymbol{\beta} = 0$$

or

$$H_0: \mathbf{C}_{nonlin}\boldsymbol{\beta} = \mathbf{0},\ \mathbf{C}_{nonlin} = [\mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4, \mathbf{c}_5]'$$

You can test this by a Hotelling's $T^2$.





I found a suitable $\mathbf{C}_{nonlin}$

```
Cmd> print(c_lin,c_nonlin,format:"5.0f")
c_lin:
(1)   -10    -4    -1     2     5     8
c_nonlin:
(1,1)    -1     1     2    -2     0     0
(2,1)    -1     1     0     0     2    -2
(3,1)     0    -2     2     1     0    -1
(4,1)     4    -4    -1    -5     5     1

Cmd> c_nonlin %*% c_lin # they are orthogonal to c_lin
(1,1)           0
(2,1)           0
(3,1)           0
(4,1)           0

Cmd> tsq_nonlin <- (c_nonlin %*% ybar)' %*% \
        solve(c_nonlin %*% vhat_ybar %*% c_nonlin') %*%\
        (c_nonlin %*% ybar)

Cmd> tsq_nonlin # Hotelling's T^2
            (1)
(1)       41.79
```

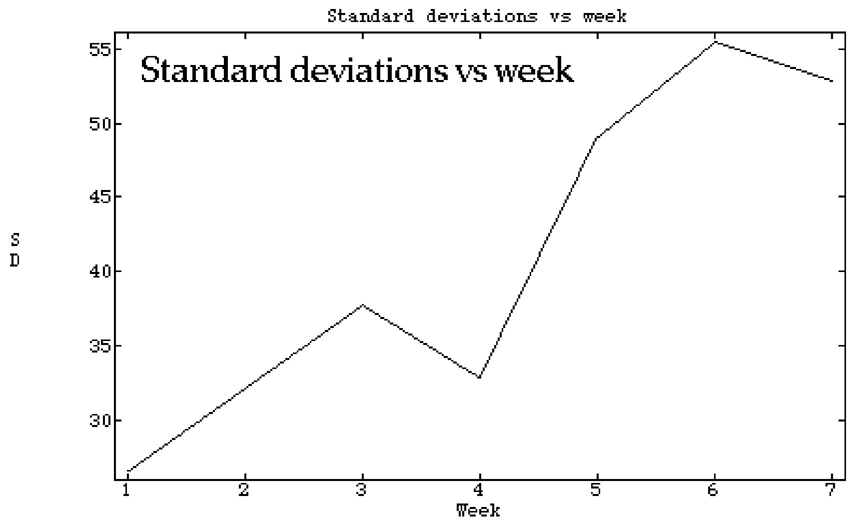
The dimension is now q = 6 - 2 = 4.

```
Cmd> q <- p - 2

Cmd> f_nonlin <- (fe - q + 1)*tsq_nonlin/(fe*q); f_nonlin
(1,1)      7.8357

Cmd> cumF(f_nonlin,q,fe - q + 1,upper:T) # P-values
(1,1)     0.00526
```

You can reject the null hypothesis that the growth is linear at the 1% level of significance.

Could you properly do a <u>split plot</u> ANOVA? Is **Σ** of the proper form?

```
Cmd> sd <- sqrt(diag(s)) # standard deviations

Cmd> lineplot(weeks,sd,\
 title:"Standard deviations vs week", ylab:"SD",xlab:"Week")
```




There seems to be an increase in $\sqrt{s_{ii}}$ as time goes on, contrary to the assumption that $\sigma_{11} = \sigma_{22} = \ldots = \sigma_{pp}$. This is not a formal test.





When variables are determined at fairly widely spaced times, as here, you often expect that the correlations will decrease as the time between determinations increases.

```
Cmd> r <- s/(sd*sd') # or cor(RESIDUALS)

Cmd> print(r,format:"8.4f")
r:
         Week 1   Week 3   Week 4   Week 5   Week 6   Week 7
Week 1   1.0000   0.7076   0.4592   0.5437   0.4924   0.5021
Week 3   0.7076   1.0000   0.8900   0.8742   0.6768   0.8349
Week 4   0.4592   0.8900   1.0000   0.8812   0.7896   0.8478
Week 5   0.5437   0.8742   0.8812   1.0000   0.8031   0.9193
Week 6   0.4924   0.6768   0.7896   0.8031   1.0000   0.8956
Week 7   0.5021   0.8349   0.8478   0.9193   0.8956   1.0000
```

At least the correlations are all positive.

It's conceivable the true correlations are the same but it looks like the longer the time between observations, the lower the correlation.

Here's some somewhat tricky MacAnova output exploring this. It finds the average correlation for each lag, 1, 2, 3, 4, 5 or 6.

```
Cmd> lags <- abs(weeks - weeks') # lags between observations

Cmd> print(lags,format:"4.0f")
MATRIX:
(1,1)    0    2    3    4    5    6      Absolute "lag"
(2,1)    2    0    1    2    3    4      between variables
(3,1)    3    1    0    1    2    3
(4,1)    4    2    1    0    1    2
(5,1)    5    3    2    1    0    1
(6,1)    6    4    3    2    1    0

Cmd> lags <- triupper(lags),pack:T);print(lags,format:4.0f")
lags:      Upper half of preceding, including diagonal
 (1)    0    2    0    3    1    0    4    2    1    0    5    3
(13)    2    1    0    6    4    3    2    1    0

Cmd> rupper <- triupper(r,pack:T); print(rupper,format:"8.4f")
rupper:    Upper half of r, including diagonal
 (1)   1.0000   0.7076   1.0000   0.4592   0.8900   1.0000
 (7)   0.5437   0.8742   0.8812   1.0000   0.4924   0.6768
(13)   0.7896   0.8031   1.0000   0.5021   0.8349   0.8478
(19)   0.9193   0.8956   1.0000

Cmd> r1 <- rupper[lags == 1]  # 4 lag 1 correlations

Cmd> r2 <- rupper[lags == 2]  # 4 lag 2 correlations

Cmd> r3 <- rupper[lags == 3]  # 3 lag 3 correlations

Cmd> r4 <- rupper[lags == 4]  # 2 lag 4 correlations

Cmd> r5 <- rupper[lags == 5]  # 1 lag 5 correlation
```

**r1** contains lag 1 week correlations, **r2** contains lag 2 weeks correlations, ... .

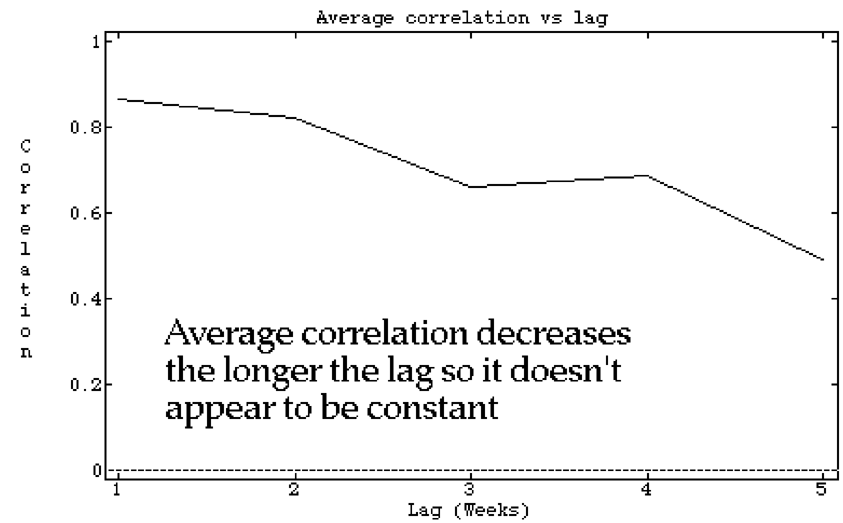




```
Cmd> list(r1,r2,r3,r4,r5)
r1              REAL   4
r2              REAL   4
r3              REAL   3
r4              REAL   2
r5              REAL   1
```

Find the mean lag 1 correlations, the mean lag 2 correlations, ... .

```
Cmd> meanlaggedr <- vector(sum(r1)/4, sum(r2)/4, sum(r3)/3,\
               sum(r4)/2, r5)

Cmd> meanlaggedr # average correlations
(1)     0.86747     0.82268     0.66123     0.68932     0.49237

Cmd> lineplot(1,meanlaggedr,ymin:0,ymax:1,xlab:"Lag (Weeks)",\
     ylab:"Correlation", title:"Average correlation vs lag",\
     xticks:run(5))
```




Of course, this doesn't constitute a formal statistical test.

Here's what the ANOVA would look like.

First create length Np = 90 response vector by stringing the columns of **y** together. Then build factors for the between group and within subject factors

```
Cmd> Y <- vector(y) #
Cmd> GROUP <- factor(rep(group,p))
Cmd> WEEKS <- factor(rep(run(p),rep(N,p)))
```

You also need a factor for subjects (whole plots)

```
Cmd> SUBJECT <- factor(rep(run(N),3)) # 1 to N three times
Cmd> list(Y,GROUP, WEEKS,SUBJECT)
GROUP           REAL   90    FACTOR with 3 levels
SUBJECT         REAL   90    FACTOR with 15 levels
WEEKS           REAL   90    FACTOR with 6 levels
Y               REAL   90
Cmd> anova("Y=GROUP + E(GROUP.SUBJECT) + WEEKS + WEEKS.GROUP",\
     fstat:T) # Split plot anova
Model used is Y=GROUP + E(GROUP.SUBJECT) + WEEKS + WEEKS.GROUP
WARNING: summaries are sequential
               DF          SS          MS           F      P-value
CONSTANT        1  2.8133e+07  2.8133e+07  3201.99930     < 1e-08
GROUP           2       18548        9274     1.05552     0.37821
ERROR1         12  1.0543e+05      8786.2    16.19444     < 1e-08
WEEKS           5  1.4255e+05       28511    52.55046     < 1e-08
GROUP.WEEKS    10      9762.7      976.27     1.79944    0.080144
ERROR2         60       32553      542.54
```

**Note**: F for **GROUP** is same as F for groups before.





- The **ERROR1** MS is the denominator for the between groups F
- The **ERROR2** MS is the denominator for the between variables and interaction F.

The conclusions are the same as before

- No apparent interaction
- Strong week effect
- No apparent vitamin E effect.