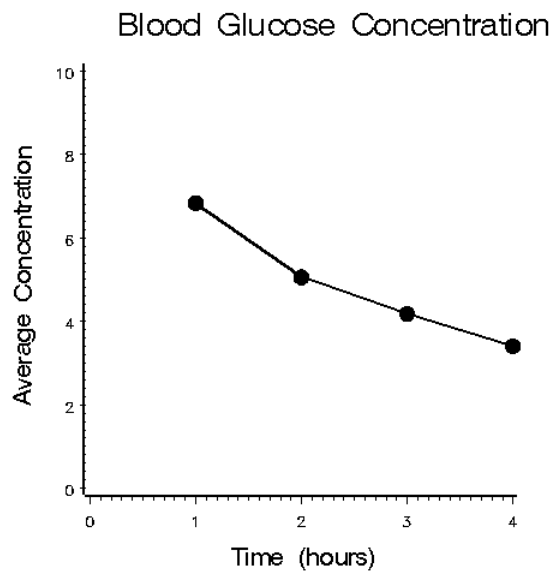


Stat 501 Solutions and Comments on Assignment 3 Spring 2005

1. (a) The following plot suggests a decreasing trend in the average blood glucose concentration across time, but it does not show the level of variation among individual subjects.



95%CI: μ_1 : [3.90,9.78]
 μ_2 : [2.70,7.42]
 μ_3 : [2.13,6.23]
 μ_4 : [2.32,4.50]

$$(b) \quad H_0: \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad T^2 = 8.236, \quad F = 1.83 \quad \text{with d.f.} = (3,4)$$

Since the p-value associated with the F-test is 0.283, the data do not provide conclusive evidence against the null hypothesis that the mean blood glucose concentration did not change across time.

(c)

Contrast	Lower limit	upper limit
$\mu_1 - \mu_2$	-3.40	6.96
$\mu_1 - \mu_3$	-2.56	7.84
$\mu_1 - \mu_4$	-1.43	8.29
$\mu_2 - \mu_3$	-3.03	4.78
$\mu_2 - \mu_4$	-2.18	5.49
$\mu_3 - \mu_4$	-2.19	3.73

$$(d) \quad H_0: \begin{bmatrix} -1 & 2 & -1 & 0 \\ 0 & -1 & 3 & -1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad T^2 = 0.397, \quad F = 0.166 \quad \text{with d.f.} = (2,5)$$

Since the p-value associated with the F-test is 0.85, the null hypothesis that the mean concentrations lie on a straight line cannot be rejected.

$$(e) \quad R = \begin{bmatrix} 1.00 & .2468 & .1563 & .0532 \\ .2468 & 1.00 & .3721 & .1658 \\ .1563 & .3721 & 1.00 & .4260 \\ .0532 & .1658 & .4260 & 1.00 \end{bmatrix} \quad \text{and} \quad X^2 = \left[n - 1 - \frac{2p+5}{6} \right] \log(|R|) = 2.88$$

with 6 d.f. and p-value = .824. The null hypothesis of zero correlations is not rejected.

(f) $X^2 = 4.58$ with 8 d.f. and p-value = .801. The null hypothesis of equal correlations and equal variances is not rejected. It is not surprising that none of the null hypotheses in the first six parts of this problem were rejected. The tests have little power because there are only seven subjects.

(g) Using the error sum of squares and corrected total sum of squares that I gave you, the ANOVA table is:

Source of Variation	d.f.	Sums of squares	Mean Square	F	Conser. d.f.
Subjects	6	55.35	9.225		
Time points	3	45.67	15.22	2.02	1
Error	18	82.05	4.558		6
Cor. total	27	183.07			

Note that the subject sum of squares is obtained by summing all of the elements in the S matrix, multiplying the sum by (n-1)=6, and dividing the result by p=4.

The test in part (f) does not indicate a need to use conservative degrees of freedom.

(h) The variance components are estimated as

$$\hat{\sigma}_{\text{error}}^2 = \text{MS}_{\text{error}} = 4.558 \quad \text{and} \quad \hat{\sigma}_{\text{subjects}}^2 = \frac{\text{MS}_{\text{subjects}} - \text{MS}_{\text{error}}}{4} = 1.167$$

Then, the common within subject correlation is estimated as

$$r = \frac{\hat{\sigma}_{\text{subjects}}^2}{\hat{\sigma}_{\text{error}}^2 + \hat{\sigma}_{\text{subjects}}^2} = \frac{1.167}{4.558 + 1.167} = .204$$

which is consistent with the correlations corresponding to the off-diagonal elements of S.

3. (a)

$$\text{For the females: } n_1 = 7 \quad \bar{X}_1 = \begin{bmatrix} 5 \\ 8 \end{bmatrix} \quad \text{and} \quad S_1 = \begin{bmatrix} 4/6 & 5/6 \\ 5/6 & 14/6 \end{bmatrix}$$

$$\text{For the males: } n_2 = 5 \quad \bar{X}_2 = \begin{bmatrix} 6 \\ 9 \end{bmatrix} \quad \text{and} \quad S_2 = \begin{bmatrix} 1.0 & 1.5 \\ 1.5 & 2.5 \end{bmatrix}$$

Then the pooled estimate of the covariance matrix is

$$S = \frac{(n_1 - 1) S_1 + (n_2 - 1) S_2}{(n_1 - 1) + (n_2 - 1)} = \begin{bmatrix} 0.8 & 1.1 \\ 1.1 & 2.4 \end{bmatrix} \quad \text{with 10 d.f.}$$

(b) There is no obvious indication that the covariance matrices are not homogeneous.

$$M = \sum_{i=1}^2 (n_i - 1) \log(|S|) - \sum_{i=1}^2 (n_i - 1) \log(|S_i|) = 3.0175$$

$$C^{-1} = 1 - \frac{(2)2^2 + (3)(2) - 1}{6(2+1)(2-1)} \left(\frac{1}{7-1} + \frac{1}{5-1} - \frac{1}{6+4} \right) = 0.7713$$

$$\text{Then, } MC^{-1} = 2.33 < \chi_{3,.05}^2 \quad \text{and} \quad p\text{-value} > .05$$

- (c) There is no clear indication that either the mean tail length or the mean wing length is different for males and females.

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{X}_1 - \bar{X}_2)' S^{-1} (\bar{X}_1 - \bar{X}_2)$$

$$= \frac{n_1 n_2}{n_1 + n_2} (-1 \ -1) \begin{bmatrix} 0.8 & 1.1 \\ 1.1 & 2.4 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ -1 \end{bmatrix} = 4.108$$

and

$$F = \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} T^2 = \frac{12 - 3}{(10)(2)} (4.108) = 1.85 \text{ on } (2, 9) \text{ d.f. (p-value} > .10)$$

(d)

Contrast	Formula	Confidence Interval
$\mu_{1F} - \mu_{1M}$	$(5 - 6) \pm (2.634)(.5237)$	$(-2.38, 0.38)$
$\mu_{2F} - \mu_{2M}$	$(5 - 6) \pm (2.634)(.5237)$	$(-3.39, 1.39)$

4. (a) profile plot

- (b) Wilks $\Lambda = 0.246$ $F = 3.0236$ d.f. = (25, 165) p-value = .0001
At least two of the profiles are not the same.

- (c) Wilks $\Lambda = 0.343$ $F = 2.8591$ d.f. = (20, 150.2) p-value = .0001
At least two of the profiles are not parallel

- (d) This parameterization is what SAS would use when the combinations of dose and manufacturer are simply used as six different treatment groups and the no intercept option (NOINT) is specified in the MODEL statement of PROC GLM. We have

$$\begin{bmatrix} X_{111} & X_{211} & X_{311} & X_{411} & X_{511} \\ X_{112} & X_{212} & X_{312} & X_{412} & X_{512} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{119} & X_{219} & X_{319} & X_{419} & X_{519} \\ X_{121} & X_{221} & X_{321} & X_{421} & X_{521} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{129} & X_{229} & X_{329} & X_{429} & X_{529} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{161} & X_{261} & X_{361} & X_{461} & X_{561} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{169} & X_{269} & X_{369} & X_{469} & X_{569} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{111} & \mu_{211} & \mu_{311} & \mu_{411} & \mu_{511} \\ \mu_{112} & \mu_{212} & \mu_{312} & \mu_{412} & \mu_{512} \\ \mu_{113} & \mu_{213} & \mu_{313} & \mu_{413} & \mu_{513} \\ \mu_{121} & \mu_{221} & \mu_{321} & \mu_{421} & \mu_{521} \\ \mu_{122} & \mu_{222} & \mu_{322} & \mu_{422} & \mu_{522} \\ \mu_{123} & \mu_{223} & \mu_{323} & \mu_{423} & \mu_{523} \end{bmatrix} + \begin{bmatrix} \text{error} \\ \text{matrix} \end{bmatrix}$$

or $X_{54 \times 5} = A_{54 \times 6} \beta_{6 \times 5} + \epsilon_{54 \times 5}$ and n hypotheses are written in the form
 $H_0: C\beta M = 0$.

(i)

$$C = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \quad M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix} \quad F=1.47 \quad df=(12,119.35) \quad p\text{-value}=.3293$$

$$\text{(ii)} \quad C = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \quad M = I_{5 \times 5} \quad F = 1.34 \quad df=(15, 121.87) \quad p\text{-value}=.3337$$

$$\text{(iii)} \quad C = \begin{bmatrix} 1 & -1 & 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 & 1 & -1 \end{bmatrix} \quad M = I_{5 \times 5} \quad F=1.42 \quad df=(10, 88) \quad p\text{-value}=.1858$$

(iv) Averaging across manufacturers you would use

$$C = \begin{bmatrix} 1 & -1 & 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 & 1 & -1 \end{bmatrix} \quad M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix} \quad F=5.82 \quad df=(8, 90) \quad p\text{-value}<.0001$$

Within manufacturers you would use

$$C = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \quad M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix} \quad F=3.46 \quad df=(16, 138.11) \quad p\text{-value}<.0001$$

$$(v) \quad C = I_{6 \times 6} \quad M = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \quad F=5.27 \quad df=(18, 130.6) \quad p\text{-value}<.0001$$

(vi) This hypothesis cannot be expressed in the form $H_0: C\beta M = 0$

E.

$$X_{ijkl} = \mu + M_i + D_j + MD_{ij} + \delta_{ijl} + \tau_k + M\tau_{ik} + D\tau_{jk} + MD\tau_{ijk} + \varepsilon_{ijkl}$$

where $i=1,2$ denotes the manufacturer

$j=1,2,3$ denote the dosage levels

$k=1,2,3,4,5$ denotes the time levels

$l=1,2,\dots,9$ denotes the rabbits within the 6 treatment groups

$\delta_{ijl} \sim NID(0, \sigma_\delta^2)$ is a random rabbit effect

$\varepsilon_{ijkl} \sim NID(0, \sigma_\varepsilon^2)$ is a random error

and δ_{ijl} is independent of any ε_{ijkl}

Source of variation	d.f.	Sums of Squares	Mean Squares	F	Mixed model d.f.	Conservative d.f.
Manuf.	1	164.89	164.89	0.35	(1, 48)	(1, 48)
Dose	2	15715.62	7857.81	16.60	(2, 48)	(2, 48)
M*D int.	2	1314.27	617.14	1.39	(2, 48)	(2, 48)
rabbits	48	22724.71	473.43			
Time	4	69674.65	17418.66	305.19	(4,192)	(1,48)
T*M	4	87.94	21.88	0.39	(4,192)	(1,48)
T*D	8	2224.30	278.04	4.87	(8,192)	(2,48)
T*M*D	8	499.10	62.44	1.09	(8,192)	(2,48)
error	192	10958.40	57.08			
Cor. total	269	123364.3				

The test for Mauchly's condition yields a chi-squared value of $X^2=85.39$ with 9 d.f. and $p\text{-value}<0.0001$. Hence, this condition is violated and some adjustment to degrees of freedom is needed in the lower part of the ANOVA table. This is also indicated by the 0.50 value of the estimated Geisser-Greenhouse correction on the SAS output. Using conservative degrees of freedom does not change the inferences in this case. There is evidence of a manufacturer effect. Both time and dosage level have significant effects and there is a significant interaction between these two factors.

- (g) The polynomial analysis suggests that a 4-th degree polynomial is needed to model the effects of time and dosage level, i.e.,

$$X_{ijkl} = \beta_0 + \beta_1(\text{time}) + \beta_{2j}(\text{time})^2 + \beta_3(\text{time})^3 + \beta_4(\text{time})^4 + \text{error}$$

Note that a different coefficient is applied to the $(\text{time})^2$ term for each dosage level. Some researchers may consider a fourth degree polynomial as too awkward and search for a more elegant model.

- (h) Conservative degrees of freedom are listed in the table in part (e). Conclusions are not affected in this case.

5.

(a).	F=2.91	df=(3,27)	p-value=0.0527
(b).	F=2.94	df=(4,26)	p-value=0.0394
(c).	F=2.58	df=(3,56)	p-value=0.0626
(d).	F=2.16	df=(4,55)	p-value=0.0861

- (e). If we ignore positive correlation, variances of difference between husbands and wives will be overestimated. This may reduce the values of F-test statistics and increase the corresponding p-values, and we may lose some of the power of test.