

# BIOMETRICS INFORMATION

(You're 95% likely to need this information)

PAMPHLET NO. # 39

DATE: June 18, 1992

SUBJECT: A Repeated Measures Example

This pamphlet describes an example analysis performed on a set of repeated measures data. The method of analysis is based on the method proposed by Meredith and Stehman. This paper was discussed in our recent "Repeated Measures Workshop".

In this simplified experiment, thirty seedlings from each of three families were randomly placed in a greenhouse. There were three blocks, with ten seedlings per family in each block. Seedling heights were measured every two weeks for twenty-six weeks, skipping week 24.

The purpose of the experiment was:

- to determine whether the final heights (cm) of the seedlings varied among families; and
- to determine whether the time (week) of maximum growth rates differed among families.

This is a repeated measures experiment because the same seedlings are measured at each time. The steps for analyzing this type of data are as follows:

1. for each treatment units (seedlings), fit a regression line (linear or non-linear) of the response variable with time; and
2. run ANOVA or MANOVA tests on the fitted parameters.

Figure 1 below shows the mean heights for each of the three families in Block 1.

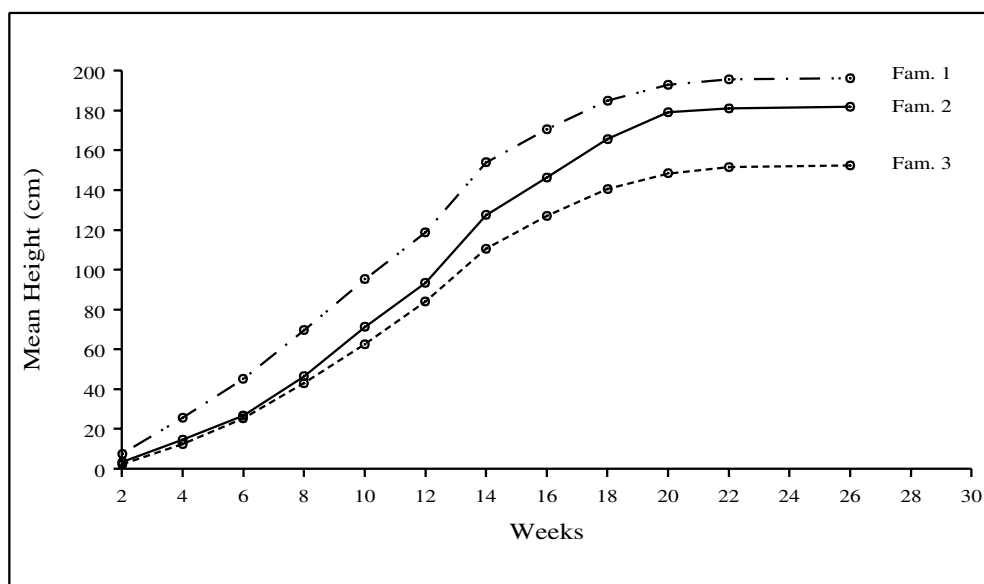


Figure 1. Mean height for each of the three families in Block 1

The S-shape of the curves suggest that the generalized logistic equation

$$y = \frac{a}{d + e^{b-cx}} \quad (1)$$

could be used in step 1. For this example,  $y$  is seedling height and  $x$  is time. The SAS procedure, PROC NLIN, can be used to fit this non-linear regression to the data. Since PROC NLIN is an iterative procedure, starting values for the parameters  $a$ ,  $b$ ,  $c$ , and  $d$  are required. The following examination of the logistic equation will help determine starting values and how the parameters are related to the experiment's objectives.

First, the logistic equation can be linearized. If we rearrange the terms and apply natural log (ln) to both sides of the equation, equation (1) can be rewritten as

$$\ln\left[\frac{a}{y} - d\right] = b - cx \quad (2)$$

The term on the left is linearly related to  $x$ . If we had estimates for  $a$  and  $d$ , we could transform  $y$  to  $\ln\left[\frac{a}{y} - d\right]$ , and run a linear regression using PROC REG on  $\ln\left[\frac{a}{y} - d\right]$  and  $x$  to get estimates for  $b$  and  $c$ . These estimates can then be used as starting values in PROC NLIN.

Second, the parameters in equation (1) describe different parts of the logistic curve. As the  $x$  value increases,  $b - cx$  becomes more negative for  $c > 0$ , and  $e^{b-cx}$  approaches zero. Therefore, for large values of  $x$ ,  $y$  has the value  $a/d$ , the final height. Also,  $x = b/c$  is the inflection point<sup>†</sup>. For our example, it is the time when the maximum growth rate occurs.

To directly test the final height and inflection point, we can re-parameterize equation (1) to get

$$y = \frac{A}{1 + De^{(B-x)C}} \quad (3)$$

where  $A = a/d$ ;  $B = b/c$ ;  $C = c$ ;  $D = 1/d$  (4)

Thus, we can test for differences in final height by testing  $A$ , and test for differences in inflection point by testing  $B$ . The following were the steps performed in this analysis.

1. For each seedling,  $a$  = height at week 26, and  $d = 1$  was used in equation (2) to estimate the starting values for  $b$  and  $c$ .
2. PROC NLIN was used to fit equation (3) to each seedling. The starting values for  $A$ ,  $B$ ,  $C$ , and  $D$  were computed from the estimated starting values  $a$ ,  $b$ ,  $c$ , and  $d$  according to equations (4). The coefficient of determination,  $R^2_{\ddagger}$ , of each curve was checked to ensure a reasonable fit was achieved.

---

<sup>†</sup>An inflection-point is where the shape of the curve changes from concave up to concave down or vice versa.

<sup>‡</sup>In this document,  $R^2$  is defined as:  $R^2 = \frac{\text{corrected total SS} - \text{residual SS}}{\text{corrected total SS}}$ .

3. PROC GLM was used to run ANOVA and MANOVA tests on the parameters.

The complete SAS program is shown in the appendix, and the test results are summarized below.

MANOVA results on the estimated parameter ( $\hat{A}$   $\hat{B}$   $\hat{C}$   $\hat{D}$ ) based on Wilks' Lambda statistics:

<u>Factor</u>	<u>p-value</u>
Family	0.5796
Block	0.0081
Block x Family	0.0001

ANOVA results on the individual parameters:

<u>Factor</u>	<u>p-value</u>			
	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>
Family	0.0052	0.2414	0.0801	0.3017
Block	0.1446	0.0091	0.3814	0.7590
Block x Family	0.0693	0.0011	0.3488	0.1150

The MANOVA for differences in the parameter (A B C D) showed a strong block by family interaction effect ( $p=0.0001$ ). This could mean that the interaction is significant for one or all of the parameters A, B, C, D.

The ANOVA for A showed non-significant interaction effect and Block effect, and significant Family effect. This implies that seedlings from different families reached different final heights; seedlings from the same family achieved similar final height regardless of where they were placed in the green house.

The ANOVA for B showed a significant interaction effect. This means the main effect tests are meaningless and we should focus on the test on interaction. Below is a table of the least square means of B.

Least Square Means of the Inflection-Points (B)

		<u>Block</u>			
		<u>1</u>	<u>2</u>	<u>3</u>	
Family	1	10.78	10.11	10.63	Standard Error = 0.177
	2	10.73	10.96	10.89	
	3	9.67	10.03	10.82	

The least significant difference for comparing two means, LSD, is given by the formula

$$\text{LSD} = t \sqrt{\frac{2 \cdot \text{MS}}{n}}$$

For B,  $n = 10$ ,  $\text{MS}(\text{Error}) = 0.3134$ ; if  $\alpha = 0.05$ , the required t-value is  $t = 2.262$ . Therefore, to examine the family by block interaction effect on B,

$$\text{LSD} = 2.262 \sqrt{\frac{2 \cdot 0.3134}{10}} = 0.57$$

The following is a plot of the least square means for the inflection-points with the LSD indicated.

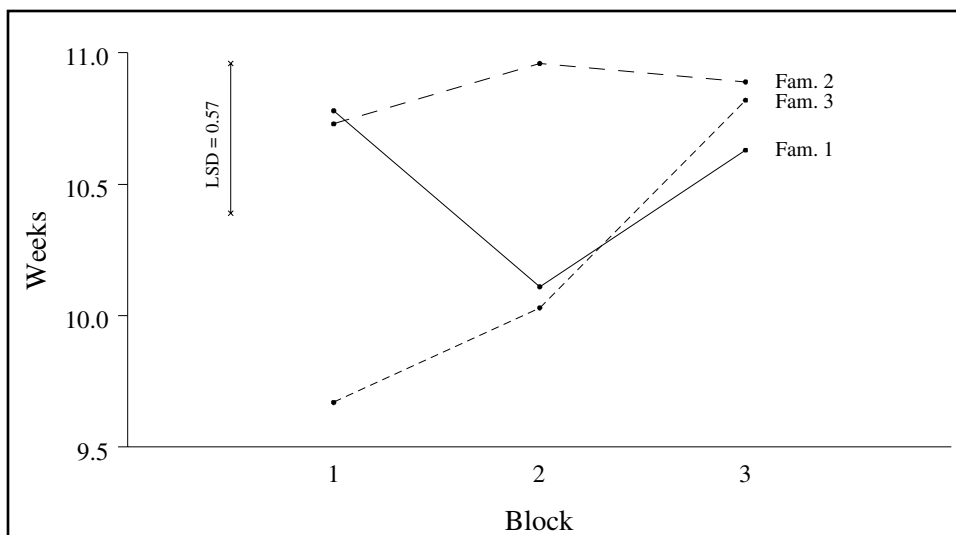


Figure 2. Plot of mean inflection-point by family and block

The plot shows that the maximum growth rate occurred at different times for the three families in the three blocks. All three families reached their maximum growth rate at about the same time when they grew in Block 3.

## Reference

Meredith, M.P. and S.V. Stehman. 1991. Repeated measures experiments in forestry; focus on analysis of response curves. *CJFR*, 21: pp 957-965.

CONTACT: Vera Sit  
356-0435

Appendix: SAS program for Analysing Repeated Measures Data

```

data height;
  input family block tree wk ht;

proc sort;
  by family block tree descending wk ;

/* do log transformation to height data */
/* a1 = maximum height observed (at week 26) */
/* = starting value for parameter a in proc nlin */

data new;
  set height;
  if wk=26 then a1=ht+0.1; /* add 0.1 to avoid computing */
  if ht>a1 then a1=ht+0.1; /* log(0) when ht=a1 */
  retain a1;
  y=log(a1/ht -1);
  output;
  keep family block a1 y ht wk tree ;

/* compute starting values for parameters */
/* b and c in proc nlin. */

proc reg outest=stats noprint;
  by family block tree;
  model y=wk ;

/* data param contains the original heights and the starting */
/* values for parameters a1 b1 c1 */
/* starting value for d is 1 */

data stats;
  set stats;
  b1 = intercep;
  c1 = wk;
  keep family block tree b1 c1;

data param;
  merge stats new;
  by family block tree;
  keep family block tree ht wk a1 b1 c1;

run;

/* non-linear regression with individual starting values */

proc nlin data=param outest=stats ;
  by family block tree;
  parameters a=220 b=4.323973 c=0.777747 d=1;
  if _iter_=0 then do; /* assign starting values */
    a=a1/d; b=-1*b1/c1; c=-1*c1; /* for each seedling */
  end; /* when iteration=0 */
  ebxc=exp(c*(b-wk));
  den=1+d*ebxc;
  aden2=a/(den**2);

```

```

model ht=a/den;
  der.a=1/den; /* derivatives of model */
  der.b=-1*aden2*c*d*ebxc; /* with respect to */
  der.c=-1*aden2*d(b-wk)*ebxc; /* parameters a, b, c, d*/
  der.d=-1*aden2*ebxc;
  output out=pred p=pht sse=sse parms= a b c d;
/* data pred contains final estimates for */
/* each family, block and tree */
data pred;
  set pred;
  by family block tree;
  if last.tree;
  keep family block tree a b c d;

proc glm data=pred;
  class family block;
  model a b c d=family|block/ss3;
  test h=block family e=family*block;
  lsmeans family|block / pdiff stderr;
  manova h=block family e=family*block;
  manova h=family*block;

run;

```