Chapter 5: Multivariate Analysis and Repeated Measures

Multivariate — More than one dependent variable at once. Why do it? Primarily because if you do parallel analyses on lots of outcome measures, the probability of getting significant results just by chance will definitely exceed the apparent $\alpha = 0.05$ level. It is also possible in principle to detect results from a multivariate analysis that are not significant at the univariate level.

The simplest way to do multivariate analysis is to do a univariate analysis on each dependent variable separately, and apply a Bonferroni correction. The disadvantage is that testing this way is less powerful than doing it with real multivariate tests.

Another advantage of a true multivariate analysis is that it can "notice" things missed by several Bonferroni–corrected univariate analyses, because ...

Under the surface, a classical multivariate analysis involves the construction of the unique linear combination of the dependent variables that shows the strongest relationship (in the sense explaining the remaining variation) with the independent variables.

The linear combination in question is called the first **canonical variate** or **canonical variable**.

> The number of canonical variables equals the number of dependent variables (or IVs, whichever is fewer).

> The canonical variables are all uncorrelated with each other. The second one is constructed so that it has as strong a relationship as possible to the independent variables — subject to the constraint that it have zero correlation with the first one, and so on.

This why it is not optimal to do a principal components analysis (or factor analysis) on a set of dependent variables, and then treat the components (or factor scores) as dependent variables. Ordinary multivariate analysis is already doing this, and doing it much better.

Assumptions

As in the case of univariate analysis, the statistical assumptions of multivariate analysis concern *conditional distributions* — conditional upon various configurations of independent variable **X** values. Here we are talking about the conditional *joint* distribution of several dependent variables observed for each case, say $Y_1, ..., Y_k$. These are often described as a "vector" of observations. It may help to think of the collection of DV values for a case as a point in k–dimensional space, and to imagine an arrow pointing from the origin $(0, ...,0)$ to the point $(Y_1, ..., Y_k)$; the arrow is literally a vector. As I say, this may help. Or it may not.

The classical assumptions of multivariate analysis depend on the idea of a **population covariance**. The population covariance between Y_2 and Y_4 is denoted $\sigma_{2,4}$, and is defined by $\sigma_{2,4} = \rho\, \sigma_2\, \sigma_4$, where

$\quad\quad \sigma_2$ is the population standard deviation of Y_2,

$\quad\quad \sigma_4$ is the population standard deviation of Y_4, and

$\quad\quad \rho$ is the population correlation between Y_2 and Y_4

$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ (that's the Greek letter rho).

The population covariance can be estimated by the sample covariance, defined in a parallel way by $s_{2,4} = r\, s_2\, s_4$, where s_2 and s_4 are the sample standard deviations and r is the Pearson correlation coefficient.

Whether we are talking about population parameters or sample statistics, it is clear that zero covariance means zero correlation and vice versa.

We will use Σ (the capital Greek letter sigma) to stand for the population **variance-covariance matrix**. This is a k by k rectangular array of numbers with variances on the main diagonal, and covariances on the off–diagonals. For 4 dependent variables it would look like this:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 & \sigma_{3,4} \\ \sigma_{1,4} & \sigma_{2,4} & \sigma_{3,4} & \sigma_4^2 \end{bmatrix}$$

With this background, the assumptions of classical multivariate analysis are that (conditional on the **X** values)

Sample vectors $\mathbf{Y} = (Y_1, ..., Y_k)$ represent **independent observations** for different cases.

Each conditional distribution is **multivariate normal.**

Each conditional distribution has the **same population variance-covariance matrix** Σ.

These assumptions are directly parallel to those of classical univariate regression. Also parallel to univariate analysis is a linear model for each population mean (now we have k of them).

$$E[\mathbf{Y}|\mathbf{x}] = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix} = \begin{bmatrix} E[Y_1|\mathbf{x}] \\ E[Y_2|\mathbf{x}] \\ \vdots \\ E[Y_k|\mathbf{x}] \end{bmatrix} = \begin{bmatrix} \beta_{0,1} + \beta_{1,1}x_1 + \cdots + \beta_{p\pm1,1}x_{p\pm1} \\ \beta_{0,2} + \beta_{1,2}x_1 + \cdots + \beta_{p\pm1,2}x_{p\pm1} \\ \vdots \\ \beta_{0,k} + \beta_{1,k}x_1 + \cdots + \beta_{p\pm1,k}x_{p\pm1} \end{bmatrix}$$

There are **k different sets of regression coefficients** –– one for each dependent variable.

There is only **one set of independent variables** –– the same for each DV. Dummy variables, interactions etc. are exactly as in univariate regression.

Estimation: The least squares estimates of those doubly–subscripted betas are **exactly what one would get from k separate univariate analyses**. Since the estimated regression coefficients are the same, so are the $\hat{Y}$ values and so are the residuals. All methods for univariate residual analysis apply.

Only the tests and confidence intervals (probability statements) are different for univariate and multivariate analysis.

Testing: In univariate analysis, different standard methods for deriving tests (these are hidden from you) all point to Fisher's F test. In multivariate analysis there are four major test statistics, **Wilks' Lambda**, **Pillai's Trace**, **the Hotelling-Lawley Trace**, and **Roy's Greatest Root**.

When there is only one dependent variable, these are all equivalent to F. When there is more than one DV they are all about equally "good" (in any reasonable sense), and conclusions from them generally agree —— but not always. Sometimes one will designate a finding as significant and another will not. In this case you have borderline results and there is no conventional way out of the dilemma.

The four multivariate test statistics all have F approximations that are used by SAS and other stat packages to compute p–values. Tables are available in textbooks on multivariate analysis. For the first three tests (Wilks' Lambda, Pillai's Trace and the Hotelling-Lawley Trace), the F approximations are very good. For Roy's greatest root the F approximation is lousy. This is a problem with the cheap method for getting p–values, not with the test itself. One can always use tables.

When a multivariate test is significant, many people then follow up with ordinary univariate tests to see "which dependent variable the results came from." More conservative (and better) is to **follow up with Bonferroni-corrected univariate tests**. When you do this, there is no guarantee that any of the Bonferroni–corrected tests will be significant.

It is also possible, and in some ways very appealing, to follow up a significant multivariate test with Scheffe tests. For example, Scheffe follow–ups to a significant one–way multivariate ANOVA would include adjusted versions of all the corresponding univariate one–way ANOVAs, all multivariate pairwise comparisons, all univariate pairwise comparisons, and lots of other possibilities — all simultaneously protected at the 0.05 level.

You can also try interpret a significant multivariate effect by looking at the canonical variates, but there is no guarantee they will make sense.

```
/******************** senicmv96a.sas ***********************/
options linesize=79;
title 'Senic data: SAS glm & reg multivariate intro';

%include 'senicdef.sas';  /* senicdef.sas reads data, etc.
                             Includes reg1-reg3, ms1 & mr1-mr3 */


/* First a nice two-factor MANOVA on infrisk & stay */

proc glm;
    class region medschl;
    model infrisk stay = region|medschl;
    manova h = _all_;
```

The glm output starts with full univariate output for each DV. Then (for each effect tested) some multivariate output you ignore,

```
                    General Linear Models Procedure
                    Multivariate Analysis of Variance

         Characteristic Roots and Vectors of: E Inverse * H, where
        H = Type III SS&CP Matrix for REGION   E = Error SS&CP Matrix

        Characteristic    Percent      Characteristic Vector  V'EV=1
            Root
                                            INFRISK          STAY

           0.14830859     95.46         -0.00263408      0.06067199
           0.00705986      4.54          0.08806967     -0.03251114
```

Followed by the interesting part.

```
                  Manova Test Criteria and F Approximations for
                     the Hypothesis of no Overall REGION Effect
            H = Type III SS&CP Matrix for REGION    E = Error SS&CP Matrix

                          S=2     M=0     N=51

  Statistic                      Value          F       Num DF    Den DF  Pr > F

  Wilks' Lambda                0.86474110     2.6127        6        208   0.0183
  Pillai's Trace               0.13616432     2.5570        6        210   0.0207
  Hotelling-Lawley Trace       0.15536845     2.6672        6        206   0.0163
  Roy's Greatest Root          0.14830859     5.1908        3        105   0.0022

          NOTE: F Statistic for Roy's Greatest Root is an upper bound.
                 NOTE: F Statistic for Wilks' Lambda is exact.
```

```
...

                    Manova Test Criteria and Exact F Statistics for
                      the Hypothesis of no Overall MEDSCHL Effect
              H = Type III SS&CP Matrix for MEDSCHL    E = Error SS&CP Matrix


                            S=1     M=0     N=51

        Statistic                    Value         F      Num DF    Den DF   Pr > F

        Wilks' Lambda              0.92228611    4.3816        2       104   0.0149
        Pillai's Trace             0.07771389    4.3816        2       104   0.0149
        Hotelling-Lawley Trace     0.08426224    4.3816        2       104   0.0149
        Roy's Greatest Root        0.08426224    4.3816        2       104   0.0149

              NOTE: F Statistic for Roy's Greatest Root is an upper bound.

...

                    Manova Test Criteria and F Approximations for
                     the Hypothesis of no Overall REGION*MEDSCHL Effect
            H = Type III SS&CP Matrix for REGION*MEDSCHL    E = Error SS&CP Matrix


                            S=2     M=0     N=51

        Statistic                    Value         F      Num DF    Den DF   Pr > F

        Wilks' Lambda              0.95784589    0.7546        6       208   0.6064
        Pillai's Trace             0.04228179    0.7559        6       210   0.6054
        Hotelling-Lawley Trace     0.04387599    0.7532        6       206   0.6075
        Roy's Greatest Root        0.04059215    1.4207        3       105   0.2409

              NOTE: F Statistic for Roy's Greatest Root is an upper bound.
                 NOTE: F Statistic for Wilks' Lambda is exact.
```

Remember the output started with the univariate analyses. We'll look at them here (out of order) — just Type III SS, because that's parallel to the multivariate tests. We are tracking down the significant multivariate effects for Region and Medical School Affiliation. Using Bonferroni correction means only believe it if $p < 0.025$.

```
Dependent Variable: INFRISK    prob of acquiring infection in hospital

Source                   DF      Type III SS     Mean Square  F Value    Pr > F

REGION                    3       6.61078342      2.20359447     1.35    0.2623
MEDSCHL                   1       6.64999500      6.64999500     4.07    0.0461
REGION*MEDSCHL            3       5.32149160      1.77383053     1.09    0.3581
```

```
Dependent Variable: STAY    av length of hospital stay, in days

Source                    DF      Type III SS      Mean Square  F Value    Pr > F

REGION                     3      41.61422755      13.87140918     5.19    0.0022
MEDSCHL                    1      22.49593643      22.49593643     8.41    0.0045
REGION*MEDSCHL             3       0.92295998       0.30765333     0.12    0.9511
```

We conclude that the multivariate effect comes from a univariate relationship between the IVs and stay. Question: If this is what we were going to do in the end, why do a multivariate analysis at all? Why not just two univariate analyses with a Bonferroni correction?

The command file senicmv96a.sas continues as follows;

```
/* Now do it with proc reg. Syntax is the same, except list more
   than one dependent variable, and say "mtest" instead of "test." */

proc reg;
    model infrisk stay = reg1-reg3 ms1 mr1-mr3;
    regtest:  mtest reg1=reg2=reg3=0;
    mstest:   mtest ms1=0;
    m_by_r:   mtest mr1=mr2=mr3=0;
```

This gives us exactly the same results we got from proc glm. The point is that multivariate analysis of variance is just a special case of multivariate regression; you can do it either way. Proc reg can give you a little more control over the details, but at the cost of setting up your own dummy variables.

Repeated measures

In certain kinds of experimental research, it is common to obtain **repeated measurements of a variable from the same individual at several different points in time**. Usually it is unrealistic to assume that these repeated observations are uncorrelated, and it is very desirable to build their inter–correlations into the statistical model.

Sometimes, an individual (in some combination of experimental conditions) is measured under essentially the same conditions at several different points in time. In that case we will say that time is a **within-subjects factor**, because each subject contributes data at more than one value of the IV "time." If a subject experiences only one value of an IV, it is called a **between subjects factor**.

.Sometimes, an individual (in some combination of other experimental conditions) experiences more than one experimental treatment — for example judging the same stimuli under different background noise levels. In this case the order of presentation of different noise levels would be *counterbalanced* so that time and noise level are unrelated (not confounded). Here noise level would be a within–subjects factor. The same study can definitely have more than one within–subjects factor and more than one between subjects factor.

The meaning of main effects and interactions, as well as their graphical presentation, is the same for within and between subjects factors.

We will discuss three methods for analyzing repeated measures data. In an order that is convenient but not chronological they are

        1.   The multivariate approach.
        2.   The classical univariate approach.
        3.   The covariance structure approach.

The multivariate approach to repeated measures

First, note that any of the 3 methods can be multivariate, in the sense that several dependent variables can be measured at more than one time point. We will start with the simpler case in which a single dependent variable is measured for each subject on several different occasions.

The basis of the multivariate approach to repeated measures is that **the different measurements conducted on each individual should be considered as multiple dependent variables.**

If there are k dependent variables, regular multivariate analysis allows for the analysis of up to k linear combinations of those DVs, *instead of* the original dependent variables.

All the multivariate approach does is to set up those linear combinations to be *meaningful* in terms of representing the repeated measures structure of the data.

For example, suppose that men and women in 3 different age groups are tested on their ability to detect a signal under 5 different levels of background noise. There are 10 women and 10 men in each age group for a total n = 60. Order of presentation of noise levels is randomized for each subject, and the subjects themselves are tested in random order. This is a three–factor design. Age and sex are between subjects factors, and noise level is a within–subjects factor.

Let Y_1, Y_2, Y_3, Y_4 and Y_5 be the "Detection Scores" under the 5 different noise levels.

Let Y_1, Y_2, Y_3, Y_4 and Y_5 be the "Detection Scores" under the 5 different noise levels. Their population means are $\mu_1, \mu_2, \mu_3, \mu_4$ and μ_5 respectively.

Now construct 5 linear combinations of the Y's, as follows.

$$W_1 = (Y_1+Y_2+Y_3+Y_4+Y_5) / 5 \qquad E(W_1) = (\mu_1+\mu_2+\mu_3+\mu_4+\mu_5) / 5$$
$$W_2 = Y_1 - Y_2 \qquad\qquad\qquad E(W_2) = \mu_1 - \mu_2$$
$$W_3 = Y_2 - Y_3 \qquad\qquad\qquad E(W_3) = \mu_2 - \mu_3$$
$$W_4 = Y_3 - Y_4 \qquad\qquad\qquad E(W_4) = \mu_3 - \mu_4$$
$$W_5 = Y_4 - Y_5 \qquad\qquad\qquad E(W_5) = \mu_4 - \mu_5$$

All the population means are of course *conditional* on the values of some independent variables. We will adopt a linear model for each one, as in the usual multivariate setup. In this case the independent variables (the weights for the linear combinations of β's) are dummy variables for the categorical independent variables sex & age, and the product terms for their interactions.

Between-subjects effects: The main effects for age and sex, and the age by sex interaction, are just analyses conducted as usual on a single linear combination of the DVs, that is, on W_1. This is what we want; we are just averaging across within−subject values.

Within-subject effects: Suppose that (for each configuration of **X** values)

$$E(W_2) = E(W_2) = E(W_2) = E(W_2) = 0$$

This means $\qquad \mu_1 = \mu_2, \ \mu_2 = \mu_3, \ \mu_3 = \mu_4, \ \mu_4 = \mu_5.$

That is, no difference among noise level means, i.e., no main effect for the within−subjects factor.

Interactions of between and within-subjects factors are between-subjects effects tested simultaneously on the dependent variables representing differences among within-subject values –– W_2 through W_5 in this case. For example, a significant sex difference in W_2 through W_5 means that the pattern of differences in mean discrimination among noise levels is different for males and females. Conceptually, this is exactly a noise level by sex interaction.

Similarly, a sex by age interaction on W_2 through W_5 simultaneously means that the pattern of differences in mean discrimination among noise levels depends on special combinations of age and sex –– a three−way (age by sex by noise) interaction.

Note: There is nothing in this discussion that limits us to dummy variables for categorical independent variables. Thus, multiple regression with repeated measures is completely reasonable and presents no special difficulties.

Here is noise.dat. Order of vars is

ident, interest, sex, age, noise level, time noise level presented, discrim score

```
esc> less noise.dat
  1   2.5   1   2   1   4   50.7
  1   2.5   1   2   2   1   27.4
  1   2.5   1   2   3   3   39.1
  1   2.5   1   2   4   2   37.5
  1   2.5   1   2   5   5   35.4
  2   1.9   1   2   1   3   40.3
  2   1.9   1   2   2   1   30.1
  2   1.9   1   2   3   5   38.9
  2   1.9   1   2   4   2   31.9
  2   1.9   1   2   5   4   31.6
  3   1.8   1   3   1   2   39.0
  3   1.8   1   3   2   5   39.1
  3   1.8   1   3   3   4   35.3
  3   1.8   1   3   4   3   34.8
  3   1.8   1   3   5   1   15.4
  4   2.2   0   1   1   2   41.5
  4   2.2   0   1   2   4   42.5
```

```
/**************** noise96a.sas *********************/
options linesize=79 pagesize=250;
title 'Repeated measures on Noise data:  Multivariate approach';
proc format;        value sexfmt     0 = 'Male'  1 = 'Female' ;

data loud;
      infile 'noise.dat';   /* Multivariate data read */
     input ident  interest  sex  age  noise1 time1 discrim1
           ident2 inter2    sex2 age2 noise2 time2 discrim2
           ident3 inter3    sex3 age3 noise3 time3 discrim3
           ident4 inter4    sex4 age4 noise4 time4 discrim4
           ident5 inter5    sex5 age5 noise5 time5 discrim5 ;
     format sex sex2-sex5 sexfmt.;
     /* noise1 = 1, ... noise5 = 5. time1 = time noise 1 presented etc.
        ident, interest, sex & age are identical on each line */
     label interest = 'Interest in topic (politics)';

proc glm;
     class age sex;
     model discrim1-discrim5 = age|sex;
     repeated noise profile/ short summary;
```

First we get univariate analyses of discrim1−discrim5 –– not the transformed
vars yet. Then,

General Linear Models Procedure
Repeated Measures Analysis of Variance
Repeated Measures Level Information

Dependent Variable   DISCRIM1 DISCRIM2 DISCRIM3 DISCRIM4 DISCRIM5

Level of NOISE   1   2   3   4   5

Manova Test Criteria and Exact F Statistics for
the Hypothesis of no NOISE Effect
H = Type III SS&CP Matrix for NOISE   E = Error SS&CP Matrix

S=1   M=1   N=24.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.45363698	15.3562	4	51	0.0001
Pillai's Trace	0.54636302	15.3562	4	51	0.0001
Hotelling-Lawley Trace	1.20440581	15.3562	4	51	0.0001
Roy's Greatest Root	1.20440581	15.3562	4	51	0.0001

Manova Test Criteria and F Approximations for

the Hypothesis of no NOISE*AGE Effect
        H = Type III SS&CP Matrix for NOISE*AGE   E = Error SS&CP Matrix

                        S=2   M=0.5   N=24.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.84653930	1.1076	8	102	0.3645
Pillai's Trace	0.15589959	1.0990	8	104	0.3700
Hotelling-Lawley Trace	0.17839904	1.1150	8	100	0.3597
Roy's Greatest Root	0.16044230	2.0857	4	52	0.0960

          NOTE: F Statistic for Roy's Greatest Root is an upper bound.
              NOTE: F Statistic for Wilks' Lambda is exact.

              Manova Test Criteria and Exact F Statistics for
                  the Hypothesis of no NOISE*SEX Effect
        H = Type III SS&CP Matrix for NOISE*SEX   E = Error SS&CP Matrix

                        S=1   M=1   N=24.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.93816131	0.8404	4	51	0.5060
Pillai's Trace	0.06183869	0.8404	4	51	0.5060
Hotelling-Lawley Trace	0.06591477	0.8404	4	51	0.5060
Roy's Greatest Root	0.06591477	0.8404	4	51	0.5060

              Manova Test Criteria and F Approximations for
                  the Hypothesis of no NOISE*AGE*SEX Effect
       H = Type III SS&CP Matrix for NOISE*AGE*SEX   E = Error SS&CP Matrix

                        S=2   M=0.5   N=24.5

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.84817732	1.0942	8	102	0.3735
Pillai's Trace	0.15679252	1.1058	8	104	0.3654
Hotelling-Lawley Trace	0.17313932	1.0821	8	100	0.3819
Roy's Greatest Root	0.12700316	1.6510	4	52	0.1755

          NOTE: F Statistic for Roy's Greatest Root is an upper bound.
              NOTE: F Statistic for Wilks' Lambda is exact.

```
                    General Linear Models Procedure
                  Repeated Measures Analysis of Variance
                Tests of Hypotheses for Between Subjects Effects

Source                   DF      Type III SS      Mean Square  F Value    Pr > F

AGE                       2      1751.814067      875.907033      5.35    0.0076
SEX                       1        77.419200       77.419200      0.47    0.4946
AGE*SEX                   2       121.790600       60.895300      0.37    0.6911

Error                    54      8839.288800      163.690533
```

Then we are given "Univariate Tests of Hypotheses for Within Subject Effects"
We will discuss these later. After that in the lst file, ...

```
            Repeated measures on Noise data:  Multivariate approach

                    General Linear Models Procedure
                  Repeated Measures Analysis of Variance
                   Analysis of Variance of Contrast Variables

             NOISE.N represents the nth successive difference in NOISE

Contrast Variable: NOISE.1

Source                   DF      Type III SS      Mean Square  F Value    Pr > F

MEAN                      1      537.00416667    537.00416667      5.40    0.0239
AGE                       2       10.92133333      5.46066667      0.05    0.9466
SEX                       1       45.93750000     45.93750000      0.46    0.4996
AGE*SEX                   2       83.67600000     41.83800000      0.42    0.6587

Error                    54     5370.09100000     99.44612963

Contrast Variable: NOISE.2

Source                   DF      Type III SS      Mean Square  F Value    Pr > F

MEAN                      1      140.14816667    140.14816667      1.36    0.2489
AGE                       2      106.89233333     53.44616667      0.52    0.5985
SEX                       1       33.90016667     33.90016667      0.33    0.5688
AGE*SEX                   2      159.32233333     79.66116667      0.77    0.4670

Error                    54     5569.94700000    103.14716667
```

```
Contrast Variable: NOISE.3

Source                   DF      Type III SS     Mean Square  F Value    Pr > F

MEAN                      1       50.41666667     50.41666667     0.72    0.4012
AGE                       2       56.40633333     28.20316667     0.40    0.6720
SEX                       1      195.84266667    195.84266667     2.78    0.1012
AGE*SEX                   2      152.63633333     76.31816667     1.08    0.3456

Error                    54     3802.61800000     70.41885185

Contrast Variable: NOISE.4

Source                   DF      Type III SS     Mean Square  F Value    Pr > F

MEAN                      1      518.61600000    518.61600000     7.77    0.0073
AGE                       2      449.45100000    224.72550000     3.37    0.0418
SEX                       1       69.55266667     69.55266667     1.04    0.3118
AGE*SEX                   2      190.97433333     95.48716667     1.43    0.2479

Error                    54     3602.36600000     66.71048148
```

The classical univariate approach to repeated measures

The univariate approach to repeated measures is chronologically the oldest. It
can be derived in a clever way from the multivariate tests involving within
subjects factors. It's what you get at the end of the default glm output — before
the analysis of transformed variables, which you have to request specially.

```
                    General Linear Models Procedure
                  Repeated Measures Analysis of Variance
             Univariate Tests of Hypotheses for Within Subject Effects

Source: NOISE
                                                           Adj   Pr > F
     DF        Type III SS      Mean Square   F Value   Pr > F    G – G     H – F
      4      2289.31400000     572.32850000     14.12   0.0001   0.0001    0.0001

Source: NOISE*AGE
                                                           Adj   Pr > F
     DF        Type III SS      Mean Square   F Value   Pr > F    G – G     H – F
      8       334.42960000      41.80370000      1.03   0.4134   0.4121    0.4134
```

(The adj. G – G business will be explained later)

```
Source: NOISE*SEX
                                                                    Adj  Pr > F
      DF        Type III SS        Mean Square   F Value   Pr > F   G – G    H – F
       4        142.42280000       35.60570000      0.88   0.4777  0.4722   0.4777

Source: NOISE*AGE*SEX
                                                                    Adj  Pr > F
      DF        Type III SS        Mean Square   F Value   Pr > F   G – G    H – F
       8        345.66440000       43.20805000      1.07   0.3882  0.3877   0.3882

Source: Error(NOISE)

      DF        Type III SS        Mean Square
     216       8755.83320000       40.53626481

                      Greenhouse-Geisser Epsilon = 0.9356
                            Huynh-Feldt Epsilon = 1.1070
```

To explain the classical univariate approach to repeated measures, we need to talk
about random effects and nested designs.

Nested effects. Suppose a company runs computer training schools in three
different cities. One of the cities has 2 schools, the second city has 3 schools, and
the third city also has 3 schools. In each school, 4 instructors are selected for
evaluation (students' knowledge is measured somehow).

There are three factors in this study, city, school and instructor. But each school
is of course only in one city, and let's also say that an instructor teaches in only
one school. We say that school is *nested* within city, and instructor is nested
within school. There is a good dummy variable coding scheme for nested
designs, but we'll skip it. Proc glm uses the syntax
            model learn = city school(city) instr(school);
Designs can have some factors that are nested, and others that are not (these are
called "crossed"). The patterns can be complex, and the designs can be very
useful, very relevant to certain types of research.

Random effects. The models we have been dealing with until now have included

only fixed effects. In a random effects model, **the <u>values</u> of the independent variable represent a random sample from some population of values**. In the computer school example, if instructors were just designated for inclusion in the study, instructor would be a fixed effect (we are comparing Chris to Pat). If they were randomly sampled from a population of instructors (this is a big company), instructor would be a random effect. A model that contains both fixed and random effects is called "mixed."

Significance tests in random and mixed models use F statistics, but the denominator is not always MSE, as it is for purely fixed effects models. Sometimes it is an interaction term. Choosing the right error term for mixed models can be complicated job, guided by expected values of the mean square (SS/df) terms; these are called *expected mean squares*. Sometimes there is no right error term and certain hypotheses are untestable with this technology. Fortunately the whole process can be automated, and SAS does a good job. **When the design is unbalanced, usually none of the error terms is useful, and the expected mean squares approach breaks down.**

Random effects, like fixed effects, can either be nested or not; it depends on the logic of the design. An interesting case of nested and purely random effects is provided by **sub-sampling**. For example, we take a random sample of towns, from each town we select a random sample of households, and from each household we select a random sample of individuals to test, or measure, or question.

In such cases the population variance of the DV can truly be partitioned into pieces — the variance due to towns, the variance due to households within towns, and the variance due to individuals within households. These components of variance can be estimated, and they are, by a program called proc nested, a specialized tool for just exactly this design. All effects are random, and each is nested within the preceding one.

Another example: Suppose we are studying waste water treatment, specifically the porosity of "flocks," nasty little pieces of something floating in the tanks. We randomly select a sample of flocks, and then cut each one up into very thin slices. We then randomly select a sample of slices (called "sections") from each flock, look at it under a microscope, and assign a number representing how porous it is (how much empty space there is in a designated region of the section). The independent variables are flock and section. The research question is whether section is explaining a significant amount of the variance in porosity — because if not, we can use just one section per flock, and save considerable time & expense.

The SAS syntax for this would be

```
proc sort; by flock section; /* Data must be sorted */
proc nested;
      class flock section;
      var por;
```

The F tests on the output are easy to locate. The last column of output ("Percent of total") is estimated percent of total variance due to the effect. It's fairly close to R^2, but not the same. To include a covariate (say "window"), just use `var window por;` instead of `var por;`. You'll get an analysis of por with window as the covariate (which is what you want) and an analysis of window with por as the covariate (which you should ignore).

Anyway, **the classical univariate approach to repeated measures is to treat "subjects" as a random effect that is nested within the between-subjects factors, and which does not interact with any other factors**. Interactions between subjects and various factors may be formally computed, but actually these are error terms; they are not tested.

In the noise level example, we could do

```
/**************** noise96b.sas *********************/
options linesize=79 pagesize=250;
title 'Repeated measures on Noise data:  Univariate approach';
proc format;      value sexfmt    0 = 'Male'  1 = 'Female' ;

data loud;
     infile 'noise.dat';  /* Univariate data read */
     input ident interest sex age noise time discrim ;
     format sex sexfmt.;
     label interest = 'Interest in topic (politics)'
           time     = 'Order of presenting noise level';

proc glm;
     class age sex noise ident;
     model discrim = ident(age*sex) age|sex|noise;
     random ident(age*sex) / test;
```

Notice the univariate data read! We are assuming n = number of observations, not number of cases.

The results are identical to the univariate output produced as a by−product of the multivariate approach to repeated measures —– if you know where to look.

The overall test, and tests associated with Type I & Type III SS are all invalid.

There are expected mean squares, which you should probably ignore.

There are also repeated warnings that "`This test assumes one or more other fixed effects are zero.`" SAS is buying testability of the hypotheses by assuming that you're only interested in an effect if all the higher−order interactions involving the effect are absent.

Why do it this way at all? Time−varying covariates.

The univariate approach to repeated measures has some real virtues, sometimes.

Because n = the number of observations rather than the number of cases, it is possible to have more parameters in a model than cases, or even more measurements than cases. In this situation the multivariate approach just blows up.

(Statistical methods should not be a Procrustean bed.)

The univariate approach may assume n is the number of observations, but it does not assume those observations are independent. In fact, the observations that come from the same subject are assumed to be correlated, as follows.

The "random effect" for subjects is a little piece of random error, characteristic of an individual. We think of it as random because the individual was randomly sampled from a population. If, <u>theoretically</u>, the only reason that the measurements from a case are correlated is that each one is affected by this same little piece of under–performance or over–performance, the univariate approach represents a very good model.

The "random effect for a subject" idea implies a variance–covariance matrix of the DVs (say Y_1, ..., Y_4) with a **compound symmetry** structure.

$$\Sigma = \begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}$$

Actually, compound symmetry is sufficient but not necessary for the univariate repeated F tests to be valid. All that's necessary is "**sphericity**," which means the covariances of all differences among Y's within a case are the same.

Another virtue of the univariate approach is that it allows time–dependent covariates. Standard multivariate analysis has the same X values for each dependent variable.

Now some **weak points** of the classical univariate approach:

The model is good if the *only* reason for correlation among the repeated measures is that one little piece of individuality added to each measurement by a subject. However, if there are other sources of covariation among the repeated measures (like learning, or fatigue, or memory of past performance), there is too much chance rejection of the null hypothesis. In this case the multivariate approach, with its unknown variance–covariance matrix, is more conservative.

Even more conservative (overly so, if the assumptions of the multivariate approach are met) is the Greenhouse–Geisser correction, which compensates for the problem by reducing the error degrees of freedom.

If the design is unbalanced (non–proportional n's), the "F–tests" of the classical univariate approach do not have an F distribution (even if all the statistical assumptions are satisfied), and it is unclear what they mean, if anything.

Like the multivariate approach, the univariate approach to repeated measures analysis throws out a case if any of the observations are missing. Did somebody say "mean substitution?" Oh no!)

It has real trouble with unequally spaced observations, and with very natural and high quality data sets where (probably) different numbers of observations are collected for each individual.

The covariance structure approach to repeated measures.

In the covariance structure approach, the data are set up to be read in a univariate manner, and one of the variables is a case identification, which will be used to determine which observations of a variable come from the same case. Naturally, data lines from the same case should be adjacent in the file.

Instead of assuming independence or inducing compound symmetry within subjects by random effects assumptions, **one directly specifies the structure of the covariance matrix of the observations that come from the same subject**.

The following present no problem at all:

> Time–varying covariates (categorical, too)
> Unbalanced designs
> Unequally spaced observations*
> Missing or unequal numbers of observations within subjects*
> More variables than subjects (but not more parameters than subjects)

It's implemented with SAS proc mixed. Only SAS seems to have it.

- Lots of different covariance structures are possible, including compound symmetry and unknown.
- A good number of powerful features will not be discussed here.
- Everything's still assumed multivariate normal.

---

* Provided this is unrelated to the variable being repeatedly measured. Like if the DV is how sick a person is, and the data might be missing because the person is too sick to be tested, there is a problem.

```
/**************** noise96c.sas ********************/
options linesize=79 pagesize=250;
title 'Repeated measures on Noise data:  Cov Struct Approach';
proc format;     value sexfmt    0 = 'Male'  1 = 'Female' ;

data loud;
     infile 'noise.dat';  /* Univariate data read */
     input ident  interest  sex  age  noise  time  discrim ;
     format sex sexfmt.;
     label interest = 'Interest in topic (politics)'
           time     = 'Order of presenting noise level';

proc mixed method = ml;
     class age sex noise;
     model discrim = age|sex|noise;
     repeated / type = un subject = ident r;
     lsmeans age noise;

proc mixed method = ml;
     class age sex noise;
     model discrim = age|sex|noise;
     repeated / type = cs subject = ident r;
```

Now part of noise95c.lst

The MIXED Procedure

Class Level Information

Class	Levels	Values
AGE	3	1 2 3
SEX	2	Female Male
NOISE	5	1 2 3 4 5

ML Estimation Iteration History

Iteration	Evaluations	Objective	Criterion
0	1	1521.4783527	
1	1	1453.7299937	0.00000000

Convergence criteria met.

```
                        R Matrix for Subject 1

Row          COL1          COL2          COL3          COL4          COL5

   1   54.07988333   17.08300000   21.38658333   17.91785000   24.27668333
   2   17.08300000   69.58763333   15.56748333   29.98861667   21.71448333
   3   21.38658333   15.56748333   54.37978333   25.15906667   21.00126667
   4   17.91785000   29.98861667   25.15906667   59.31531667   27.58265000
   5   24.27668333   21.71448333   21.00126667   27.58265000   55.88941667


                  Covariance Parameter Estimates (MLE)

     Cov Parm           Estimate      Std Error       Z   Pr > |Z|

     DIAG UN(1,1)      54.07988333    9.87359067     5.48     0.0001
          UN(2,1)      17.08300000    8.22102992     2.08     0.0377
          UN(2,2)      69.58763333   12.70490550     5.48     0.0001
          UN(3,1)      21.38658333    7.52577602     2.84     0.0045
          UN(3,2)      15.56748333    8.19197469     1.90     0.0574
          UN(3,3)      54.37978333    9.92834467     5.48     0.0001
          UN(4,1)      17.91785000    7.66900119     2.34     0.0195
          UN(4,2)      29.98861667    9.15325956     3.28     0.0011
          UN(4,3)      25.15906667    8.01928166     3.14     0.0017
          UN(4,4)      59.31531667   10.82944565     5.48     0.0001
          UN(5,1)      24.27668333    7.75870531     3.13     0.0018
          UN(5,2)      21.71448333    8.52518917     2.55     0.0109
          UN(5,3)      21.00126667    7.61610965     2.76     0.0058
          UN(5,4)      27.58265000    8.24206793     3.35     0.0008
          UN(5,5)      55.88941667   10.20396474     5.48     0.0001
     Residual           1.00000000        .          .         .


                  Model Fitting Information for DISCRIM

          Description                      Value

          Observations                   300.0000
          Variance Estimate                1.0000
          Standard Deviation Estimate      1.0000
          Log Likelihood                -1002.55
          Akaike's Information Criterion -1017.55
          Schwarz's Bayesian Criterion  -1045.32
          -2 Log Likelihood              2005.093
          Null Model LRT Chi-Square        67.7484
          Null Model LRT DF                14.0000
          Null Model LRT P-Value            0.0000
```

```
                    Tests of Fixed Effects

          Source          NDF   DDF   Type III F  Pr > F

          AGE               2     54       5.95   0.0046
          SEX               1     54       0.53   0.4716
          AGE*SEX           2     54       0.41   0.6635
          NOISE             4    216      18.07   0.0001
          AGE*NOISE         8    216       1.34   0.2260
          SEX*NOISE         4    216       0.99   0.4146
          AGE*SEX*NOISE     8    216       1.30   0.2455
```

From the multivariate approach we had F = 5.35, p < .001 for age & approx F = 15.36 for noise.

```
               Least Squares Means

      Level           LSMEAN      Std Error   DDF        T  Pr > |T|

      AGE 1       38.66100000   1.21376060    54    31.85    0.0001
      AGE 2       35.24200000   1.21376060    54    29.04    0.0001
      AGE 3       32.76700000   1.21376060    54    27.00    0.0001
      NOISE 1     39.82166667   0.94938474   216    41.94    0.0001
      NOISE 2     36.83000000   1.07693727   216    34.20    0.0001
      NOISE 3     35.30166667   0.95201351   216    37.08    0.0001
      NOISE 4     34.38500000   0.99427793   216    34.58    0.0001
      NOISE 5     31.44500000   0.96513744   216    32.58    0.0001
```

Now for the second mixed run we get the same kind of beginning, and then for compound symmetry structure,

```
                    Tests of Fixed Effects

          Source          NDF   DDF   Type III F  Pr > F

          AGE               2     54       5.95   0.0046
          SEX               1     54       0.53   0.4716
          AGE*SEX           2     54       0.41   0.6635
          NOISE             4    216      15.69   0.0001
          AGE*NOISE         8    216       1.15   0.3338
          SEX*NOISE         4    216       0.98   0.4215
          AGE*SEX*NOISE     8    216       1.18   0.3096
```

From the univariate approach we had F = 14.12 for noise.

Now proc glm will allow easy examination of residuals no matter which approach
you take to repeated measures, provided the data are read in a univariate manner.

```
/*************** noise96d.sas **********************/
options linesize=79 pagesize=60;
title 'Repeated measures on Noise data: Residuals etc.';
proc format;      value sexfmt    0 = 'Male'  1 = 'Female' ;

data loud;
    infile 'noise.dat';  /* Univariate data read */
    input ident  interest  sex  age  noise  time  discrim ;
    format sex sexfmt.;
    label interest = 'Interest in topic (politics)'
          time     = 'Order of presenting noise level';

proc glm;
    class age sex noise;
    model discrim = age|sex|noise;
    output out=resdata predicted=predis  residual=resdis;

/* Look at some residuals */
proc sort; by time;
proc univariate plot;
   var resdis; by time;
proc plot;
    plot resdis * (ident interest);

/* Include time */
proc mixed method = ml;
    class age sex noise time;
    model discrim = time age|sex|noise;
    repeated / type = un subject = ident r;
    lsmeans time age noise;
```

(Then I generated residuals from this new model using glm, and plotted again.
Nothing.)

```
                              Variable=RESDIS

       |
    25 +
       |
       |
       |
    20 +                         0                          |
       |                                                    |
       |                                                    |
       |                           |              |         |
    15 +                           |              |         |
       |                           |              |         |
       |                           |              |         |
    10 +            |              |              |      +-----+
       |            |              |              |      |     |
       |            |              |           +-----+   *-----*
     5 +            |              |           |     |   |     |
       |            |           +-----+        |     |   |  +  |
       |            |           |     |        |     |   |     |
       |            |           |     |        *-----*   |     |
     0 +            |           |     |        |  +  |    |     |
       |         +-----+        *--+--*        |     |    +-----+
       |         |     |        |     |        |     |       |
       |         |     |        |     |        |     |       |
    -5 +         |  +  |        +-----+        +-----+       |
       |         *-----*        |                           |
       |         |     |        |                           |
       |         |     |        |                           |
   -10 +         +-----+        |                           |
       |            |           |                           |
       |            |           |                           |
       |            |           |                           |
   -15 +            |           |                           |
       |            |           |                           |
       |            |           0                           |
   -20 +            |           0
       |            |
       |            |
   -25 +            |
         ----------+-----------+-----------+-----------+-----------
  TIME             1           2           3           4
```
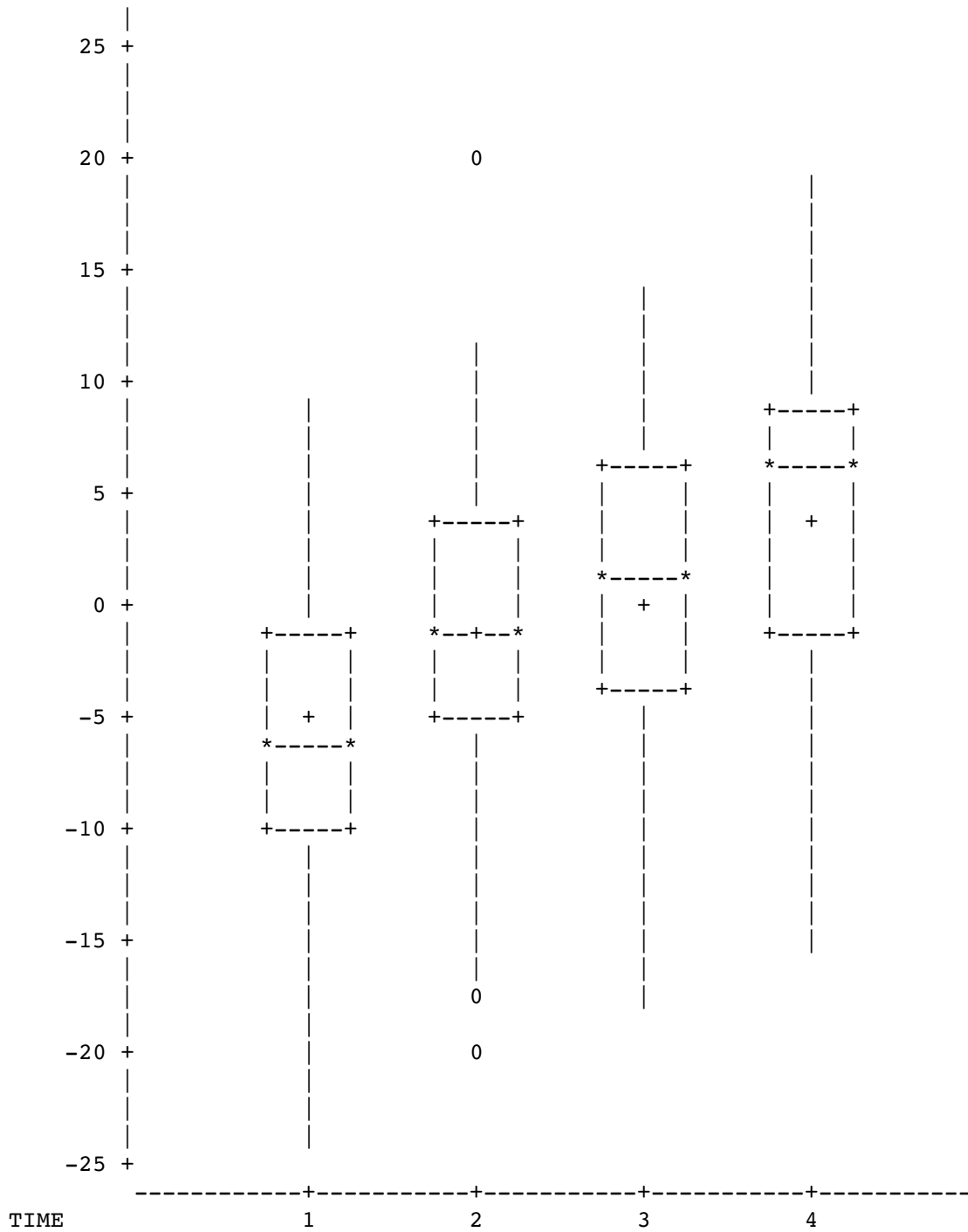
Unfortunately time = 5 wound up on a separate page. .

When time is included the results get stronger but conclusions don't change.

```
               Tests of Fixed Effects

        Source           NDF   DDF  Type III F  Pr > F

        TIME              4    266     17.67    0.0001
        AGE               2    266     18.45    0.0001
        SEX               1    266      1.63    0.2027
        AGE*SEX           2    266      1.28    0.2789
        NOISE             4    266     10.95    0.0001
        AGE*NOISE         8    266      0.51    0.8488
        SEX*NOISE         4    266      0.44    0.7784
        AGE*SEX*NOISE     8    266      0.74    0.6573

                 Least Squares Means

  Level          LSMEAN      Std Error   DDF       T  Pr > |T|

  TIME 1    29.54468242    0.91811749   266   32.18    0.0001
  TIME 2    34.61557451    0.91794760   266   37.71    0.0001
  TIME 3    36.18863723    0.92819179   266   38.99    0.0001
  TIME 4    39.72344496    0.91838886   266   43.25    0.0001
  TIME 5    37.71099421    0.93376736   266   40.39    0.0001
  AGE 1     38.66100000    0.68895774   266   56.12    0.0001
  AGE 2     35.24200000    0.68895774   266   51.15    0.0001
  AGE 3     32.76700000    0.68895774   266   47.56    0.0001
  NOISE 1   39.69226830    0.89132757   266   44.53    0.0001
  NOISE 2   36.80608879    0.89274775   266   41.23    0.0001
  NOISE 3   35.35302821    0.89130480   266   39.66    0.0001
  NOISE 4   34.12899017    0.89502919   266   38.13    0.0001
  NOISE 5   31.80295787    0.89180628   266   35.66    0.0001
```

Some good covariance structures are available in proc mixed.

Variance Components: type = vc

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

Compound Symmetry: type = cs

$$\Sigma = \begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}$$

Unknown: type = un

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 & \sigma_{3,4} \\ \sigma_{1,4} & \sigma_{2,4} & \sigma_{3,4} & \sigma_4^2 \end{bmatrix}$$

Banded: type =

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_5 & 0 & 0 \\ \sigma_5 & \sigma_2^2 & \sigma_6 & 0 \\ 0 & \sigma_6 & \sigma_3^2 & \sigma_7 \\ 0 & 0 & \sigma_7 & \sigma_4^2 \end{bmatrix}$$

First order autoregressive: type = ar(1)

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^4 & \rho & 1 \end{bmatrix}$$

There are more, including Toeplitz, banded Toeplitz & spatial (covariance is a function of Euclidian distance).