

Repeated Measures → why?

Hao Zhang

- Introduction
- Statistical Modelling and Analysis
 - The Univariate Analysis of Variance Approach
 - Modelling Covariance Structure
 - Modeling Time As a Regression Variable

★ A: You measure the same experimental units (subjects, plots, machines) multiple times — e.g. blood pressure at weeks 0, 4, 8. Responses on the same unit are correlated.

Introduction

In many applications, multiple measurements are made on the same experimental units over a period of time. Such data are called repeated measures. An example is growth curve data such as daily weights of chicks on different diets. The design for repeated measures could be a completely randomized design or other standard design. For example, three diet treatments are randomly assigned to the chicks according to a completely randomized design. The experiment units are the chicks and each unit is observed weekly for some weeks. The treatment factor is diet and is often referred to as the between-subjects factor. Time is also regarded as a factor and referred to as within-subject factor. The experimental units are often called .

In repeated measures experiments, interest centers on

1. How treatments change over time; and
2. How treatment differences change over time, i.e., is there a treatment by time interaction?

These questions arise in any factorial experiment and there is nothing peculiar about the objectives of a repeated measures experiment. What makes the repeated measures data analysis distinct is the covariance structure of the observed data—those data from the same subject are correlated and the correlation may decrease as the time lag increases.

Statistical Modelling and Analysis

The modelling and analysis of repeated measures are a complex topic. In this section, we only highlight some models and analyses by looking at some real data sets.

The Univariate Analysis of Variance Approach

Example 1. (Alzheimer's Data, Hand and Taylor, 1987, Table G.1) Two groups of patients with Alzheimer's disease were compared, one of which had 26 patients and received placebo, and the other had 22 and was treated with lecithin. The response variable is the number of words that a patient can recall from lists of words. The response variable was measured at time units 0, 1, 2, 4, and 6. Plots of the data are given in Figure 1.

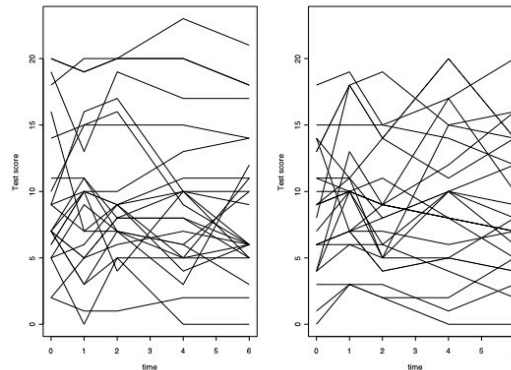


Figure 1: Alzheimer study response profiles: Placebo group on right, lecithin group on left.

From the graph, we can see differences between subjects within each group as well as differences between the two groups. In general, we will regard subject effects as random effects. In some analyses, the repeated measures from the same subject are assumed to be independent. If we take this position, we will have the **univariate analysis of variance approach**. The corresponding statistical model for this experiment is

$$y_{ijk} = \mu + \alpha_i + \underbrace{d_{j(i)}}_{\text{individual} \rightarrow \text{random}} + \tau_k + (\alpha\tau)_{ik} + \epsilon_{ijk}, \quad (1)$$

A
 WP
 B
 $A \times B$

where α_i , τ_k and $(\alpha\tau)_{ik}$ are fixed effects of treatment i , time k , and their interaction, respectively, $d_{j(i)}$ is the random effect associated with the j^{th} subject in group i , ϵ_{ijk} is random error associated with the j^{th} subject in group i at time k , $d_{j(i)}$ are i.i.d. $N(0, \sigma_s^2)$ and ϵ_{ijk} are i.i.d. $N(0, \sigma^2)$. Note that

$$E(y_{ijk}) = \mu + \alpha_i + \tau_k + (\alpha\tau)_{ik}, \quad Var(y_{ijk}) = \sigma_s^2 + \sigma^2,$$

and the covariance between any two different observations on the same subject is

$$Cov(y_{ijk}, y_{ijk'}) = Var(d_{j(i)}) = \sigma_s^2, \quad j \neq j'.$$

Such a covariance structure is called compound symmetric. Note also compound symmetry implies that $var(y_{ij} - y_{ij'})$ is a constant for any $j \neq j'$. Such a condition is called $\{$. Many computer programs report the results of the Mauchly test of sphericity though it seems this test is not powerful for detecting small departures from sphericity. Some adjusted F-tests for non-sphericity exist. Model (1) is similar to the model we used for split-plot designs since subjects are nested within the treatment groups.

We can use a very flexible SAS procedure `proc mixed` for model (1).

```
proc mixed;
class group subj time;
model response=group time group*time;
random subj(group);
run;
```

Treat "time" as a factor in a one-way ANOVA with blocks = subjects.

The `model` statement specifies three fixed effects in the model and the `random` statement specifies the random effect(s).

We see this model is similar to the model for a split-plot design.

Modelling Covariance Structure

As we said before, repeated measures from the same subject are usually dependent. Consider the Alzheimer experiment again. The measurements from the same subject on 5 occasions might be correlated. In this scenario, the model will be essentially the same but the error terms ϵ_{ijk} for the same subject are correlated. We should model this correlation structure. There are three commonly used covariance structures: compound symmetric, autoregression of order one (AR(1)) and unstructured.

1. Compound Symmetry

$$Var(\epsilon_{ijk}) = \sigma^2, Cov(\epsilon_{ijk}, \epsilon_{ijk'}) = \rho\sigma^2, k \neq k'$$

correlation is a constant

2. AR(1) $\epsilon_{ijk}, k = 1, 2, \dots$ is assumed to be an AR(1) process. Therefore, $Cov(\epsilon_{ijk}, \epsilon_{ijk'}) = \sigma^2 \rho^{|k-k'|}$.

Assume error term are dependent

3. Unstructured Covariance No mathematical pattern is imposed on the covariance matrix and the covariance structure of the repeated measures is estimated using the facts that this covariance structure remains the same for every subjects, and measurements taken from different subjects are independent.

SAS Program

We use the `repeated` statement in `proc mixed` with options `type` to specify one of the three covariance structures. For example, if we use the compound symmetric covariance structure for the Alzheimer experiment, the SAS program is

```
proc mixed;
class group subj time;
model response=group time group*time;
repeated/type=cs sub=subj(group) r rcorr;
```

In the `repeated` statement, `type=cs` specifies the covariance structure type to be compound symmetric, `sub` specifies that the compound symmetric structure pertains to submatrices corresponding to each subjects in each group. The options `r` and `rcorr` request printing of covariance matrix and correlation matrix.

If we were to use AR(1), we would change the `repeated` statement to

```
repeated/type=ar(1) sub=subj(group) r rcorr;
```

Note, this program is not appropriate for the experiment since the repeated measures were taken at unequally spaced time intervals. Use `type=sp(pow)` for unequally spaced measures.

If we use unstructured covariance, we change the `repeated` statement to

```
repeated/type=un sub=subj(group) r rcorr;
```

Some criteria exist for choosing the covariance structure, among which are Akaike's Information Criterion (AIC) and Schwarz's Bayesian Criterion (SBC). Both penalize the log likelihood function by addition a penalty term which increases with the number of parameters. We then choose the structure that maximizes a penalized log likelihood.

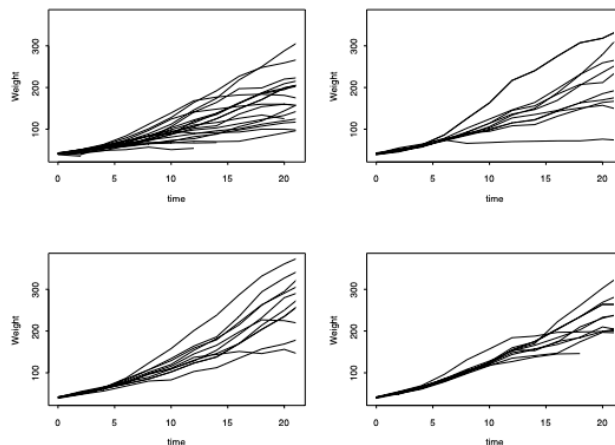


Figure 2: Growth curves of chicks on four different protein diets.

Modeling Time As a Regression Variable

Consider the study on body weights of chicks on different diets. There are four groups, each on different protein diet. Body weights are measured on alternate days. The body weights for the four groups are plotted in Figure 2.

time as a
continuous variable (t)

From the plots, we can see the differences between the groups. In addition, there are between-chicks differences within each group. For each chick, the growth curve can be reasonably modeled as a quadratic function of time. A reasonable model would be

$$y_{ijt} = \mu + \alpha_i + \underline{t}\beta_i + \underline{t^2}\gamma_i + \underline{t}b_{j(i)} + \underline{t^2}c_{j(i)} + \epsilon_{ijt},$$

Quadratic term (t^2) also may depend on subject
time effect also depend on subject

where μ , α_i , β_i and γ_i are fixed parameters, which explain for between-group differences, $b_{j(i)}$ and $c_{j(i)}$ are random coefficients, and $b_{j(i)}$ are i.i.d. $N(0, \sigma_{i,b}^2)$, $c_{j(i)}$ are i.i.d. $N(0, \sigma_{i,c}^2)$. The two random coefficients explain the between-subject differences within a group. $b_{j(i)}$ and $c_{j(i)}$ can be correlated.