7   Drawbacks and limitations of classical methods

7.1   Introduction

It is worth noting that both the univariate and multivariate "classical" methods we have discussed so far may be extended to more complicated situations. For example

- The **group** designations may in fact be the result of a **factorial arrangement**, e.g. in an experiment to compare the change over time of body weight of rats, the groups may be defined by the (2×3) factorial arrangement of genders and drugs. Interest may focus on how the rate of change in body weight over time differs across genders averaged over drugs and doses averaged across genders. Interest may also focus on whether the way this change differs across drugs is different for the two genders (the drug by gender interaction). These are "between-unit" comparisons.

- The "**time**" factor may in fact be the result of a factorial arrangement, e.g. in an agricultural study, plots may be randomized to different rates of fertilizer. Then, at each of 4 different time points, core samples are taken from each plot at 3 different depths, and a measurement of nutrient content is recorded for each. Here, then, each plot is seen under $4 \times 3 = 12$ different conditions.

We do not discuss these extensions; see, for example, Vonesh and Chinchilli (1997, section 3.3).

The fact that these fancy extensions are possible still does not alter the fact that the "classical" models and methods have some serious limitations, some of which we have remarked upon in our development so far. Now that we are familiar with these so-called "classical" methods and the statistical models underlying them, we are in a position to be more specific about these limitations.

7.2   Assumptions and restrictions of classical methods

Here, we provide a "laundry list" of the assumptions made by classical methods and the restrictions that they impose. The rest of the course will be devoted to statistical models and associated analysis methods that seek to address some or all of these restrictions.

1. *BALANCE.* A prominent feature both of the univariate and multivariate classical models and methods is the requirement that all units be observed at the **same** n "time" points. That is, not only must each data vector $\boldsymbol{Y}_i$ be of the **same** length, n, for all units, but each element Y_{ij}, $j = 1, \ldots, n$ must have been observed at the **same** set of times $t_1, \ldots, t_n$, say.

- In some situations, this may not be much of a restriction. For example, in agricultural or industrial experimentation where it is possible to have a good deal of control over experimental conditions, an experiment may be carefully planned and executed. It may thus be perfectly reasonable to expect that observations expected to be taken at certain times would be available.

- However, even in the best of situations, it is often the case that things may go awry. For example, suppose that the Y_{ij} are are responses on plots planted with different varieties of soybean over the growing season. At a given time, 3 plants from a plot are sampled, their leaves are harvested, aggregated, and ground up, and the resulting leaf sample is assayed for concentration of a particular chemical substance. It is an unfortunate fact of life that samples may be misplaced or mistakenly discarded or that error may be made in conducting the assay, leading to erroneous measurements. In such circumstances, measurements may thus be unavailable at certain time points for certain plots, thus destroying the **balance** necessary for classical models and methods to be applied.

- When the units are **humans**, this becomes even more of a problem, even if a study is carefully designed. For example, suppose that a study is conducted to compare several cholesterol-lowering drugs. Subjects are randomly assigned to take regular doses of one of the drugs and are required to return at 3 month intervals for 2 years so that a measure of serum cholesterol may be taken from blood samples drawn at each visit. Thus, if "time" for each subject is measured from the subject's entry into the study, the subject should have observations on serum cholesterol at $n = 8$ times 3, 6, 9, 12, . . . , 21, and 24 months. However, reality may cause this ideal set-up to be compromised.

  – Subjects may move away during the course of the study, so that only measurements up to their last visit before moving are available.

  – A subject may be out of town and miss his 9 month visit but come to the clinic at 10.5 months instead.

– Blood samples may be mislabelled or dropped in the lab, so that observations on serum cholesterol for some times for some subjects may be impossible to obtain.

– Errors by technicians in performing the analytic laboratory techniques required to measure the cholesterol level may render other measurements erroneous or unavailable.

The bottom line is that real life often conspires to make **balance** an unachievable ideal for many longitudinal studies. Although some researchers have discussed ways to "adjust" the classical approaches to handle some types of imbalance, just as with the "adjusted" F tests in univariate analysis, these "fix-ups" skirt the **real** issue, which is that a model that requires balance may simply be too restrictive to represent real life!

2. *FORM OF COVARIANCE MATRIX.* Both the "classical" univariate and multivariate procedures we have discussed assume that the covariance matrix of each data vector $\boldsymbol{Y}_i$, $i = 1, \ldots, m$ is the **same** for all i, regardless of group membership or anything else; we discuss this assumption below. Provided we believe this assumption is reasonable, and take $\boldsymbol{\Sigma}$ to be this common $(n \times n)$ covariance matrix, we are still faced with the issue of what we assume about the structure of $\boldsymbol{\Sigma}$.

- The univariate methods make the assumption of **compound symmetry**, which implies a very specific pattern of **correlation** among observations taken on the same unit at different times, one that may be quite unrealistic for longitudinal data. This model says that the correlation among **all** observations on a given unit is **the same regardless** of how near or far apart the observations are taken in time. Thus, the univariate methods are based on an assumption about the covariance structure that may be **too restrictive** if **within-unit** sources of correlation are not negligible.

- The multivariate methods make **no assumption** about the structure of $\boldsymbol{\Sigma}$. Thus, these methods do not attempt to take into account at all the way in which observations arise in the longitudinal setting. There are two acknowledged sources of variation:

  – Random (biological) variation among units

  – Within-unit variation due to the way in which measurements are taken on a unit (error in measuring device, correlation due to time separation, etc)

The model underlying the multivariate methods does not explicitly recognize these two distinct sources. Rather, the methods allow for the possibility that the covariance structure could be virtually **anything**, thus including as possibilities structures that are unlikely to represent data subject to the two distinct sources above. Thus, the multivariate methods are based on an assumption about the covariance structure that is likely **too vague**.

3. *COMMON COVARIANCE MATRIX.* Both the univariate and multivariate approaches assume that the covariance matrix of a data vector is **the same** for all units, regardless of group or anything else. (This is akin to making the usual assumption in linear regression or scalar analysis of variance that variance is the same for all scalar observations.) This is often adopted without much thought; however, it is quite reasonable to expect that this assumption may be incorrect.

For example, suppose the units are human subjects and the groups are determined by assignment to either a particular hypertension medication or placebo. A common observation with such data is that subjects with "high" systolic blood pressure tend to exhibit much more variability in their within-individual measured pressures than do subjects with "low" systolic blood pressure. That is, in terms of the conceptual model in Chapter 4, the within-subject "flucutations" for subjects with high blood pressure tend to be of greater magnitude than those for subjects with low blood pressure. More formally, $\text{var}(e_{1ij}$ is **smaller** for subjects with low blood pressure than for those with high blood pressure. This would lead to **overall** variance of Y_{ij} that is smaller for lower values of Y_{ij}.

Suppose the drug is quite effective in lowering systolic blood pressure. We would thus expect observations on subjects in the drug group, particularly toward the end of the study, to be "lower" than those for the placebo group. In symbols, if $\boldsymbol{Y}_i$ is a data vector for a subject in the drug group (1), we might expect

$$\boldsymbol{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in} \end{pmatrix}, \quad \text{var}(Y_{in}) = \sigma^2_{n(1)},$$

while for a subject in the placebo group (0), we might expect

$$\boldsymbol{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in} \end{pmatrix}, \quad \text{var}(Y_{in}) = \sigma^2_{n(0)},$$

$$\sigma^2_{n(1)} < \sigma^2_{n(0)}.$$

Under these conditions, assuming that $\boldsymbol{Y}_i$ from both groups have the **same** covariance matrix $\boldsymbol{\Sigma}$ would be inappropriate, because we would doubt that the (n,n) element is the **same** for data vectors from both groups. A **better** model would say that there are **two** different covariance matrices, i.e. $\text{var}(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}_0$ is subject i is in the placebo group, and $\text{var}(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}_1$ is subject i is in the drug group.

It is possible to modify the classical models and methods to handle this situation. One common approach is to work on a **transformed** scale on which one believes variances may be similar; e.g. one may model the logarithmically transformed data. A problem with this approach is that the results may be difficult to interpret, as inferences about what happens on the **original** scale of measurement are of interest. Alternatively, methods such as Hotelling's T^2 may be modified to allow a different covariance matrix for each group. However, this may make statistical power even lower – now, we must estimate a **separate** covariance matrix for each group. Later in the course we will see methods that address the issue of lack of common covariance matrix in more realistic ways.

4. *INCORPORATION OF INFORMATION.* A characteristic shared both by the univariate and multivariate classical methods we have discussed is that, because **balance** is assumed, **time** itself does not appear explicitly in the model for the mean of a data vector. Rather, "time" enters the model only through the specification of separate parameters γ_j and $(\tau\gamma)_{\ell j}$. As will become clear when we study more flexible models, this can pose an obstacle to answering some key questions of interest (see 5. below, too). This problem may be partially addressed by inspecting, for example, orthogonal polynomial contrasts in time, but a more direct representation of time in the model is much more useful.

In addition, we may wish to incorporate other **covariate** information. For example, in the cholesterol study in 1. above, we may believe that a subject's **age** at the start of the study may play a role in how he/she responds to cholesterol-lowering medication. Or we may believe that this response over time may be affected by a subject's systolic blood pressure, which may **also** be changing over time. Just as ordinary analysis of variance is modified to incorporate covariates by analysis of covariance, one may wish to do something similar in the case of repeated measurements. Things are more complicated, however.

- In the first example, the **covariate**, **age at start of study**, is something that is **time-independent**, or **fixed** over the time points at which the unit is observed, being measured only **once** (at the start of the study). Both univariate and multivariate analyses may be modified to take account of time-independent covariates; these are discussed in sections 2.6 and 3.4 of Vonesh and Chinchilli (1997).

We do not discuss them here because, as discussed above, they still require **balance**; moreover, the way in which the covariates may be included in the model is limited. Models we will discuss later in the course allow more flexibility to address common questions about the effect of covariates.

- In the second example, the **covariate**, **systolic blood pressure**, may be measured at each of the same time points as the response, and thus is **time-dependent**, or **changing** with time. Incorporation of such covariate information poses difficult conceptual challenges. The models we have discussed represent the mean response at each time point as a function of information such as group membership; i.e. possibly different means for each group. If we consider models that incorporate **changing** information, important questions arise. For example, does the mean cholesterol at a particular time only depend on systolic blood pressure at **that** time? Or does it depend on systolic blood pressure at **several** previous times **as well**?

  We will return to this issue later; for now, note that although it is possible to introduce **time-dependent covariates** into modeling of repeated measurements, a key issue is this conceptual one. It is possible to modify the univariate analysis to incorporate time-dependent covariates; however, modification of the MANOVA analyses is not possible.

Still another issue arises in the inclusion of **group** information. Recall the guinea pig diet example. Here, dose groups were labelled "zero," "low," and "high." In the model, the parameters τ_ℓ and $(\tau\gamma)_{\ell j}$ incorporate different groups. Suppose, however, that the actual numerical dose values were available, say 0, 100, and 500 μg/g. As we discuss in 5. below, it might be useful if the actual dose levels rather than just classifications were incorporated in the model.

We will discuss other models and methods where inclusion of such covariate information is more direct and interpretable.

5. *QUESTIONS OF INTEREST AND INTERPRETATION.* The analysis based on "classical" methods focuses on **hypothesis testing**, i.e. general questions of interest are stated in terms of the model and the quality of the evidence in the data to refute the null hypothesis is assessed. A pronouncement is then made (we do or don't reject the null hypothesis).

However, in many situations, this does not really address the objectives of the investigator. For example, consider the cholesterol study described in 1. above. The investigators may wish to do more than just claim that the way in which cholesterol changes on average over time on the different drugs is different. They may actually wish to use the results of their study to make recommendations on how to treat **future patients**. Thus, they may wish to make more specialized inferences.

- **How** different is the rate of cholesterol lowering among the drugs? E.g. if they knew that Drug 1 lowered cholesterol at the rate of 5 mm Hg per month and Drug 2 lowered cholesterol at rate 15 mm Hg per month, this information might help them to decide which drug (mild Drug 1 or aggressive Drug 2) might be more appropriate for a certain patient. Thus, the investigators might be interested in actually **estimating** the **rate of change** in the mean response over time for each group!

- What would the cholesterol trajectory look like for a new male patient 45 years of age after 8 months on one of the drugs? That is, before treatment, the investigators might wish to be able to **predict** what the cholesterol profile might look like over 8 months for a patient with specific characteristics and what his cholesterol level might be at the end of that time based on his measurement at time zero. Note that 8 months is not even one of the time points (every 3 months) included in the original study.

Clearly, in order to address such questions, a more flexible model that incorporates time and rate of change in a more explicit way is needed.

A further illustration is provided by the guinea pig diet example as discussed in 4. above. Suppose the investigators would like to be able to understand how the rate of change in body weight of the pigs over time is associated with the actual numerical dose. Does rate of change increase as we change the dose? By how much per unit change of dose? If the actual dose **amount** could be incorporated explicitly in the model, these questions could be addressed.

It should be clear from this brief discussion that the "classical" models and methods have serious limitations with respect to these important issues. A serious drawback alone is that of the need for **balance**. Another is failure of the models to represent explicitly important features like **rate of change** with time. We begin our discussion in the next chapter with models and methods that seek to address these problems..