Brazilian Journal of Probability and Statistics
2012, Vol. 26, No. 1, 56–70
DOI: 10.1214/10-BJPS122

On data-based selection of summary measures from repeated measurements

Ib M. Skovgaard and Torben Martinussen

University of Copenhagen

Abstract. Univariate analysis of variance of a good summary measure, or two, may provide a simple and effective way of analyzing repeated measurements. It is shown here that selection of a linear summary measure on the basis of inspection of the total sample of response curves, leads to valid F-tests in the subsequent analysis of variance. The selection may also be based on residuals from a base model, rather than on the raw data. The treatments should, however, be blinded in this summary measure selection step, that is, the inspection of the sample of curves (or residuals) and the selection of the summary measure may not rely on which responses stem from which treatment groups. It is advocated as a convenient and often effective method to use the first principal component from the total sample of curves as the first summary measure. The main mathematical result of the paper is a simple proof of the validity of the F-tests for linear summary measures selected in this way, provided data are multivariate normally distributed. Alternatively, permutation tests may be used to provide a distribution free reference distribution for the F-statistic. Two examples illustrate the method.

1 Introduction

Repeated measurements, or longitudinal data, arise when the response measured on each experimental unit in an experiment is a series of observations of the same kind, for example the concentration of a certain hormone from blood samples taken at specific times after a treatment. Thus, we are dealing with a multivariate response, or a response vector. A full model-based analysis of an experiment resulting in repeated measurements may cause various kinds of trouble related to the choice of covariance structure within the series. Aside from technical difficulties, these comprise, in particular, high sensitivity of assessment of treatment effects to the choice of covariance model, and low power of tests when a model with freely varying covariance matrix is used for long series. A specific dilemma is that more intensive measuring on each experimental unit may lead to poorer results because of the many covariance parameters that must be estimated in a full multivariate analysis.

Key words and phrases. Longitudinal data, permutation tests, principal components, repeated measurements, selection bias, summary measure.

Received July 2009; accepted April 2010.

An alternative way is to use a so-called summary measure calculated from each of the experimental units as a specific function of the response vector. This univariate result is then subsequently used in an ordinary analysis of variance (ANOVA). If two or three summary measures are used, they may be analyzed separately in an ANOVA for each of them, jointly in a multivariate analysis of variance, or first univariately for the first summary measure, then univariately for the second summary given the first, in an analysis of covariance, etc.

A well-chosen summary measure often provides a good and simple way of assessing treatment effects and other effects of interest. The problem for the researcher is, however, to choose a good summary measure. It is the purpose of the present article to advocate a default method of choosing the summary measure(s) based on principal components, and to prove that certain data-based selection methods are valid in the sense of leading to F-distributed test statistics in the ANOVA, despite the initial data-based selection. This holds when the vectors of responses are multivariate normal with any covariance matrix. If this assumption is dubious, the reference distribution may alternatively be created by permutations, again without introducing selection bias.

The key point is that it is perfectly valid to use the total sample of curves, regarded as a single population, as a basis for selection of the summary measure. In particular, the first one or two principal components may be used as summary measures. The validity of such a procedure may not be immediately evident as the principal components depend on the data. The use of principal components is based on the reasoning that when treatment effects of sufficiently large size are present, they are likely to dominate also the total variation and cause the pattern of the treatment effect to reflect itself in the summary measure.

In more general terms the result is that such data based selection of a linear function of the response vector as a summary measure does not lead to selection bias in the subsequent F-test in the analysis of variance. This is, of course, in sharp contrast to a selection of summary measure based on variation between treatment groups. The result is hardly surprising and, indeed, the proof is short, but our purpose was to give precise conditions allowing more flexible use of the approach than just the basic one, calculating the principal components from the given data.

The theorem and the approach advocated here only serve for testing treatment effects and several other effects. If such effects seem to be present, methods beyond the scope of the present paper should be used for estimation of their size and character.

Example 1 (Activity of rats). Thirty cages, each with two young rats were randomly divided into three treatment groups. The treatment was daily exposure to White Spirit levels 0 (control), 1 and 2, respectively, in relative units. The rats were then tested monthly for activity by counting how many times the two rats crossed a beam of light during a 57 hours test period. The rats were followed in

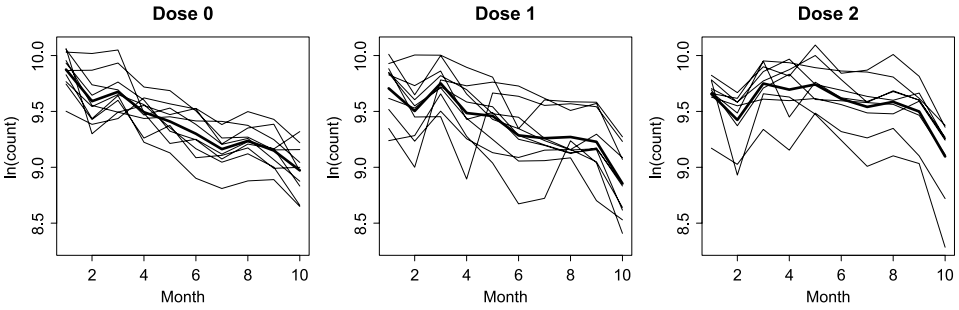

Figure 1 *The monthly log-counts for the activity of rats example. The three treatments* (*doses*) *are displayed in one plot each. Points from the same cage are joined and the thick line represents mean.*

this way for 10 months. Counts varied between 3962 and 24,197. The experimental unit is a cage, and for each of these the 10-dimensional response vector used here is the natural logarithm of the count for each of the 10 months. Results for one cage at the high dose are excluded here for reasons that are not relevant for the present paper.

The data are shown in Figure 1, where all response profiles are shown for each dose. These data, kindly provided by Grete Østergaard, are part of a series of experiments some of which are described in Østergaard et al. (1992).

The structure of this experiment is simple, but the choice of model and method for analysis may not be obvious since there are many possibilities. Some of these are briefly shown in Section 4. The present proposal, to analyze the scores from the first principal component (or first eigenvector) in a univariate analysis, automatically picks up the main features, at least in the present case. Important for the validity of the analysis, though, is that the principal component analysis is carried out on the total sample of response vectors, rather than separating the treatment groups as is done in the figure. It is a key point of the present paper that the test of treatment effect thereby remains valid. Thus, in the present case, a principal component analysis of the 29 10-dimensional vectors of logarithmic counts is carried out, ignoring groups but with unspecified 10-dimensional mean and variance–covariance. The 29 scores for the first component are then used in a one-way analysis of variance for testing group effect.

Example 2 (Glucose). The effect of infusion of a peptide (PYY) on appetite, energy balance and metabolism was investigated on 12 obese and 12 normal-weight persons; see Sloth et al. (2007) for further details. Here we use the plasma glucose measurements which were taken just before infusion (time 0) and then every 30 minutes for three hours. The infusion was discontinued after 90 minutes and a meal was given two hours later. The peptide was given in two forms, here referred to as treatment A and B, and each subject tried both treatments in a randomized crossover design. The total of 48 response curves are seen in Figure 2. The data,

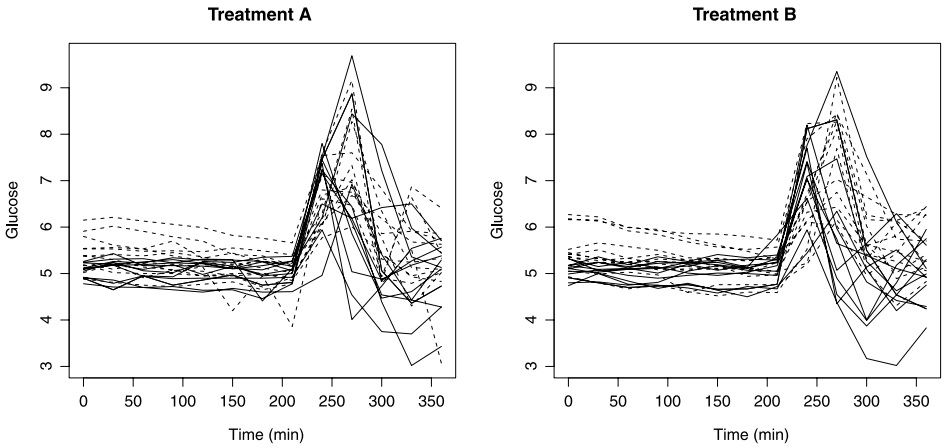

Figure 2 *Plasma glucose levels from Example* 2 *for* 24 *subjects*, 12 *normal-weight* (*solid lines*), *and* 12 *obese* (*dashed lines*), *each trying two treatments*. *The two treatments* (*forms of PYY*) *are displayed in one plot each*.

kindly provided by Birgitte Sloth, are part of the data material described in Sloth et al. (2007).

Apart from the difficulty in seeing any treatment effects, it is seen that the response changes dramatically when the meal is given at time 210 minutes. Care should be taken not to assume homogeneity of variance over time when modeling the data and furthermore some thought should be given to use an analysis focusing mainly on the relevant part of the time period. Again the method suggested here, using the first principal component, automatically puts more weight on the last part of the period in the present case.

Use of principal components in conjunction with series of measurements is advocated by many authors in the field of functional data analysis. Thus, Jones and Rice (1992) use principal components to display important features of large collection of similar curves showing for example the curves corresponding to the minimum, median and maximum of the scores associated with the first principal component. Rice and Silverman (1991) and Hall et al. (2006), and references therein, focus mostly on the estimation and representation of the function. Another application is the choice of covariates for covariance adjustment in growth curve analysis; see Soler and Singer (2000). Combination with parametric effects is investigated in, for example, Capra and Müller (1997) and Silverman (1995), but in a full and more complicated modeling approach. The approach discussed in the present paper is simpler and less ambitious by not trying to estimate the functions, but rather to extract the key features with the purpose of investigating their possible dependence on various factors. Even this point is an obvious idea and, at least, implicit in the literature on longitudinal and functional data. The present paper

adds, first, the conditions and proof for this procedure not to cause selection bias, second, suggests that variation due to less relevant effects may be removed in the calculation of principal components by use of the residuals from a basic model, third, advocates this method as a good alternative to more complicated methods, and fourth, recommends data inspection to be used in a proper way to select summary measures—a point often overlooked.

2 Main result

Let y_i be the p-dimensional response vector from individual $i = 1, \ldots, n$, and let Y denote the corresponding $n \times p$ matrix. A linear summary measure is a one-dimensional linear function,

$$Z_i = c^T y_i,$$

weighing the response coordinates by the coefficients specified as the vector c. Frequently used linear summary measures are the simple mean, area under the curve, which is often almost equivalent to the mean, the average slope of the response with time, and increase in response over a certain interval. Once the summary measure has been chosen and calculated for each individual, the analysis may continue using analysis of variance, or linear models more generally. In particular, an F-test may be obtained for treatment effect or for any other effect under consideration, provided that it may be expressed as a linear hypothesis in a univariate linear model with a univariate response for each individual.

The main result states that the selection of the vector, c, of coefficients defining the summary measure may be based, in any way, on the variation between the sample of curves without altering the F-distribution that would be obtained by a univariate analysis of variance of the summary measure, if the coefficients were given beforehand. This result requires a multivariate normal distribution of the response vector and independence between individuals.

As our base model we use

$$Y = X_b \beta_b + e_b, \tag{2.1}$$

where X_b is an $n \times q$ (design) matrix, β_b is a $q \times p$ parameter matrix, and the remainder e has independent rows with $p \times p$ variance matrix Σ_b. The maximum likelihood estimates are denoted $\hat{\beta}_b$ and $\hat{\Sigma}_b$. The residuals from this model are

$$\hat{e}_b = Y - X_b \hat{\beta}_b,$$

and it is the idea behind the base model that the variation between these residuals are to be used for selection of the summary measure. This variation is defined as the $p \times p$ matrix

$$SS_b = \hat{e}_b^T \hat{e}_b$$

consisting of sums of squares and cross products. The base model should be thought of as a model containing only the effects that we want to correct for when testing the hypothesis. Thus, the use of the residuals from the base model rather than the raw observations permits elimination of certain uninteresting types of effects, for example of sex, before numerical or graphical inspection of the curves for interesting features of variation is carried out to find good summary measures.

Next we consider our hypothesis model

$$Y = X_0\beta_0 + e_0 \tag{2.2}$$

which should contain the base model and which we wish to test against the model

$$Y = X_a\beta_a + e_a, \tag{2.3}$$

where the subscript stands for "alternative." Similar notation as above is used for these models.

Now the summary measure is supposed to be chosen by inspection of the variation between the residual vectors from the base model, that is,

$$Z_i(SS_b) = c(SS_b)^T Y_i,$$

where the vectors Z_i as well as c are written as functions of (SS_b), while the dependence of Z_i on Y_i is notationally depressed.

Let $F(SS_b)$ denote the usual univariate F-test with response variables $Z_i(SS_b)$, $i = 1, \ldots, n$, for reduction of the linear model (2.3) to the model (2.2), both models with $p = 1$. An expression for this F-statistic is

$$F(SS_b) = \frac{c^T(SS_0 - SS_a)c/(\mathrm{DF}_0 - \mathrm{DF}_a)}{c^T(SS_a)c/\mathrm{DF}_a},$$

where SS_0 and SS_a are defined analogously to SS_b, $c = c(SS_b)$, $\mathrm{DF}_a = n - \mathrm{rank}(X_a)$ and so on.

Theorem 2.1. *Let three nested models be given as above, with X_b spanning a subspace contained in that spanned by X_0 which again is contained in that spanned by X_a. Let $c(SS_b)$ be any nonvanishing p-vector with coefficients that are functions of SS_b. If the hypothesis (2.2) holds with a multivariate normal distribution for e_0, then $F(SS_b)$ has an F-distribution with $(\mathrm{DF}_0 - \mathrm{DF}_a, \mathrm{DF}_a)$ degrees of freedom, unaffected by the choice of $c(SS_b)$.*

Proof. Under the null-model (2.2) the statistic $(\hat{\beta}_0, SS_0)$ is complete sufficient and hence, by Basu's theorem, independent of any ancillary statistic. We continue by proving that for any fixed coefficient vector c, the corresponding F-statistic F_c is ancillary and consequently independent of SS_b which is a function of SS_0 and $\hat{\beta}_0$, since $\hat{\beta}_b$ is a function of $\hat{\beta}_0$. To prove the ancillarity of F_c note that the distribution of F_c remains the same when the response vector is transformed by

$$Y \to YA + X_0 b$$

for any nonsingular $p \times p$ matrix A and any p-vector b. The ancillarity then follows because this group of transformations generates the entire model (2.2). This completes the proof. □

From the proof it appears immediately that the result holds for any class of (test) statistics of the form $T(v)$ indexed by v in some arbitrary index set V, provided that:

- $T(v)$ is ancillary for the model (2.2),
- $T(v)$ has the same distribution, P_T say, for any $v \in V$,
- the selected statistic, $T_s(y)$, say, is defined as a choice of v depending only on the sufficient statistic under the model (2.2), that is, $T_s(y) = T(v(\hat{\beta}_0, SS_0))$.

The result is that the selected statistic, $T_s(y)$ then also has distribution P_T. This extension is useful when the model includes random effects, for example, so that test statistics may no longer follow the F-distribution.

3 Principal components as summary measures

The first principal component is that linear function of the response vector that has the largest sample standard deviation relative to the length of the vector of coefficients. It may also be characterized as the first eigenvector of the sample covariance matrix of the response vectors, or rather as the projection on this vector. If treatment effect is substantial it may be expected to affect the total sample variance; thus the first principal component is a reasonable candidate if a single summary measure is to be chosen automatically.

Sometimes, however, the primary variation is due to differences between the general response level of individuals, while effects of interest are rather reflected in the shape than in the level of the curves. Then it may be useful to try also the second principal component as summary measure. Since this is generally not independent of the first component, the subsequent linear model analysis of the second component should include the first component as a covariate to avoid having the same treatment features showing up twice, or the two components should be analyzed as a bivariate response in a multivariate analysis of variance, for example. Whether to use one, two or possibly even three principal components is a matter of judgement on the basis of inspection of the patterns of the response profiles, again without regard to treatment groups, the number of profiles observed, and knowledge of the response variable under investigation.

4 Examples

In the present section we briefly indicate how the suggested method works in the two introductory examples. For comparison we start each of the examples with

some results from more conventional analyses. It is not attempted to give a full analysis of the data. In particular there are other aspects of the analysis than the test of treatment effect.

Example 1 (Activity of rats (continued)). At least four or five different tests for treatment effect fall within mainstream approaches. Starting with the summary measures, the average activity over the entire period or the slope of a regression line might be calculated for each cage, reflecting, respectively, the general level of activity and the rate of decline of activity. One-way analysis of variance leads to the P-value 0.068 for the activity level and 0.0062 for the slope, with the rate of decline in the high-dose group being less than half of those of the other two groups. An analysis of covariance of the slope with level as covariate may be preferred; this gives the P-value 0.057 but should, of course, be seen in conjunction with the analysis of the level.

Turning to models for the full longitudinal data a reasonable attempt may be the multivariate one-way analysis of variance (MANOVA). The test for treatment effect gives the P-value 0.0043. Supplementary month-by-month ANOVA reveal that the differences occur mainly in months 6–9. This model uses a covariance matrix with 55 parameters, which may be excessive. A more parsimonious model may be preferred assuming variance homogeneity and modeling the correlation with one or two parameters. For example, an autoregressive model of first order run in the procedure Mixed in SAS/STAT Software Version 9.1 of the SAS System for Windows gives the P-value 0.049 for main effect of treatment and 0.0027 for the interaction between treatment and month.

In any case the results point towards a treatment effect which is most clearly seen by the analysis of the slopes. A full analysis should be more thorough and validate the longitudinal models carefully, but would lead to the same result. The point here is that there are many choices to make, and the analysis may not always be straightforward.

In comparison the first principal component from the sample of 29 10-dimensional vectors represents 63% of the variation and has coefficients (or loadings)

$$0.05, 0.11, 0.12, 0.24, 0.36, 0.40, 0.41, 0.36, 0.39, 0.41,$$

on the respective point in time. The summary measure obtained as the linear function of each response vector with these coefficients, enhances the period when treatment differences are more pronounced. A one-way ANOVA using this summary measure gives the P-value 0.0167. This approach leads easily and directly to a reasonable conclusion, although perhaps not to the same depth of investigation as a more complete modeling. The analysis could, however, be extended with the second principal component which has loadings

$$0.52, 0.67, 0.35, 0.15, -0.09, -0.13, -0.13, -0.18, -0.15, 0.19,$$

so that the summary measure resembles a slope and gives the P-value 0.021 when the first principal component is used as covariate.

Example 2 (Glucose (continued)). Of importance here is not only whether the treatment effects differ, but whether they are the same in the two groups, that is, whether treatment and group interact. We focus here on testing for such an interaction. The crossover structure of this experiment is experimentally sensible and also very common in nutritional studies, but complicates the analysis. It is not easy to come up with a reasonable model for the covariance matrix of the vector of observations from the same individual, since it consists of two series of measurements. The obvious variance heterogeneity due to the peak period makes it almost impossible to find satisfactory models implemented in commonly used software. Furthermore, the first measurement (at time 0) is a baseline measurement, unaffected by the treatment, which we would like to use as a covariate in the model. The conditioning on this first measurement further affects the covariance structure in a heterogeneous manner. Some failed attempts were made to achieve convergence for models with an unspecified covariance matrix for the series of measurements combined with a random effect of subject. Discouraged by these difficulties the two series from each subject were replaced by their differences. These difference series, each representing treatment A versus treatment B and eliminating subject main effects, were then analyzed in a MANOVA with baseline difference as covariate and with the two factors group and time together with their interaction.

The MANOVA resulted in the P-value 0.034 for the interaction between time and group, and (after elimination of the interaction) in the P-value 0.20 for the main effect of group. Notice that since these are effects on the treatment differences, they correspond to interactions with treatment in the original data. Taken together these two tests give some indication of a group-dependent treatment effect varying over time. These P-values were calculated using the default method in the procedure Mixed in SAS/STAT Software Version 9.1 of the SAS System for Windows, for approximating degrees of freedom in the F-distribution. Use of Satterthwaite's method (Satterthwaite (1946)) gave very similar results, but the Kenward–Roger correction (Kenward and Roger (1997)) gave the P-value 0.30 for interaction between group and time, followed by 0.39 for main effect of group, thus giving no indication of any group effect at all.

As an alternative, avoiding these difficulties, principal components were calculated from the residuals from the base model

$$Y = \mu(\text{subject}) + \alpha(\text{treatment}) + \beta \cdot \text{baseline} + e, \tag{4.1}$$

fitted for each time point separately. Here e denotes the residual, μ and α are arbitrary functions of the factor in question, and β is a constant. The first principal component of the resulting residual vectors has coefficients

$$0.01, 0.00, 0.02, 0.03, 0.04, 0.04, -0.02, -0.17, 0.66, 0.65, 0.31, 0.14$$

for time 30, 60, . . . , 360 minutes. As expected this coefficient puts main weight on the peak period seen in Figure 2.

The scores, obtained by multiplying each residual vector with the coefficient vector, were then analyzed in a univariate analysis, and in particular an F-test was calculated for the interaction between group and treatment. In the general terms of Section 2 this means that the null model (2.2) was identical to the base model (2.1), while the alternative model (2.3) also included the interaction term. The P-value for the interaction, using a linear normal model, was 0.048, giving some indication of a group-treatment interaction around the peak-time results.

From the estimates obtained from the univariate model for the first principal component it appears that the obese subjects have a more pronounced peak with treatment B than with treatment A, whereas the reverse is true for the normal-weight persons. This could, of course, be a random finding since the P-value is not very small.

5 Permutation tests

Once a summary measure has been chosen and a test statistic computed we need a reference distribution for assessing the test statistic. The F-distribution is often adequate, but if the assumptions leading to this distribution are in doubt a permutation test may be used. For comparison of treatment groups this entails drawing a number of random permutations of the response vectors so that they are allocated at random to the treatment groups. As discussed below, permutation test may be applied in more complex situations as well.

Since the permutation distribution is, by construction, conditional on the total sample of response curves, the selection of the test statistic may clearly be done on the basis of this total sample. Thus, even nonlinear statistics may be selected for the permutation test, such as the ratio between the values at two specific time points found as typical ‘peak times’ by inspection of the total sample of response curves.

For the rat activity example consider again the F-statistic based on the scores corresponding to the first principal component. The P-value obtained from 10,000 random permutations of these scores across the treatment groups was 0.0164, in close agreement with the P-value 0.0167, obtained from the F-distribution. Notice that the result is the same whether the eigenvectors, scores and F-statistics are calculated before or after the permutation of the 29 rats.

For the glucose example the permutation test was carried out by randomly permuting the group allocation (obese or normal-weight) on the 24 subjects while keeping the remaining variables unchanged. For the F-test statistic for interaction between group and treatment in the previous section the permutation P-value was 0.046 based on 10,000 random permutations, again in good agreement with the P-value 0.048 from the normal distribution model.

Permuting only the group factor always means that the hypothesis tested is that group allocation has no impact, no matter how the test statistic is chosen, but the choice of test statistic is decisive for which alternatives the test is directed towards and has power to detect. In the present case, the choice was to use the first principal component as response vector, after elimination of effects of baseline measurement, subject and treatment, and to consider the test for combined effect of treatment and group. Since any main effect of group is absorbed into the subject effect, a significant result points towards an interaction between group and treatment on the response profiles as represented by the first principal component.

6 Discussion

There are several possibilities of adapting the principal component method to investigate different effects of the various factors. Thus, if at some stage of the analysis the shape of the response series is of interest rather than their level, we do not want main effects of between-subject factors to affect the result, assuming for the moment that there is one series per subject. For that purpose we replace each series by the incremental series obtained as the successive differences and then proceed as before. A treatment effect on the first principal component, for example, then means that treatment interacts with time, because neither an additive effect of time nor of treatment would affect the summary measures by more than a common constant. Similarly it was seen in Example 2 (the glucose data) how an interaction between two between-series factors may be tested by suitable choice of the hypothesis model. Finally, the choice of base model may eliminate uninteresting effects, such as that of baseline, before the summary measure is derived.

Summarizing a series of measurements into one or two measures before further analysis may seem like a waste of data and important information may, indeed, be lost. There is also a potential gain, however, because well-chosen measures focus on the key features of the data, thereby avoiding loss of power due to the use of an omnibus test or due to estimation of many covariance parameters. The price is that there may be features of the response profiles that are affected by treatment or other factors, but which happen not to show up in the first principal component, even asymptotically as the number of series increases. In practice, however, the number of series is fixed and there is no way to guarantee that any perceivable treatment effect is found.

Alternative methods may provide a more complete modeling of the data, but they are not without problems as seen in Example 2 in Section 4. The hardest problem is often the choice of covariance structure. A model assuming variance homogeneity may well be unrealistic, especially if an initial (baseline) measurement is used as covariate. More rich and realistic models may involve many parameters, especially with long or dense series of measurements, and this may in

turn cause loss of power. On top of this, the theory of distributional approximations to the test statistics is not sufficiently well developed to give reliable results, as was also seen in Section 4.

The arguments above are not intended to imply that the method of summary measures, defined as principal components or otherwise, should always be used. Varying time points between subjects is one situation for which principal components are not suited, although summary measures still give valid results if time points are the same, but some observations are missing at random. Another type of situation, where principal components should not be used, occurs when the interesting feature of each curve, say a peak, is displaced in time from one series to another. A third type occurs when there is a sufficient number of replications to analyze the curves more fully. The use of a principal component summary is, however, a perfectly sound and often quite good alternative to more complicated methods, the reliability of which may be hard to judge even for experts.

Appendix: Data used in the examples

Table 1 *The data from Example* 1: *Activity of rats. Each line represents a cage with two rats. First column* (*T*) *is treatment* (*exposure in ppm*), *the subsequent columns contain the monthly counts. Data originate from Grete Østergaard, University of Copenhagen*

	Month									
T	1	2	3	4	5	6	7	8	9	10
0	20584	15439	17376	14785	11189	10366	8725	9974	9576	6849
0	23265	16956	16200	12934	13763	11893	9949	10490	8674	7153
0	17065	12429	14757	10524	11783	8828	9016	9635	8028	8099
0	19265	19316	20598	16619	16092	13422	10532	10614	9466	9494
0	21062	14095	13267	12543	12734	12268	12219	11791	10379	8463
0	23456	10939	13270	14089	12986	13723	11878	13338	12442	10094
0	13383	11899	12531	15081	14295	13650	9988	11518	11915	7844
0	22717	22434	23151	13163	10029	10408	9119	10188	9549	11153
0	17437	13950	15535	14199	11540	9568	8481	9143	8117	5765
0	18546	12520	15394	10137	9218	7343	6702	7173	7257	5708
400	18536	16827	19185	12445	13227	10412	9855	9169	9639	6853
400	18831	14043	16493	12562	10397	8568	8599	8818	6011	5062
400	15016	13765	16648	14537	13929	10778	9897	9225	9491	5523
400	22276	15497	22024	15616	12440	11454	10290	9456	9567	7003
400	18943	14834	18403	16232	13085	12679	10489	9495	10896	8836
400	13598	10233	13392	10457	9236	8847	9445	9501	8509	5656
400	20498	22136	22094	19825	18157	11452	14809	14564	14503	10643
400	19586	12710	12745	7294	15757	15296	14097	14308	13933	10210
400	11474	8108	17714	16795	17364	16766	15016	13475	14349	8698
400	10284	10760	15628	10692	8420	5842	6138	10271	8435	4486

Table 1 (*Continued*)

	Month									
T	1	2	3	4	5	6	7	8	9	10
800	18459	15805	19924	18337	24197	18790	19333	22234	18291	11595
800	16186	11750	16470	18637	14862	14695	14458	14228	12909	9079
800	9614	8319	11375	9446	13157	11153	10540	11476	8976	6123
800	15688	15016	20929	12706	17351	15089	14605	15952	14795	10434
800	15864	13169	20991	20655	19763	19180	19003	18172	15025	11790
800	17721	14489	19085	21333	17011	16148	15280	14762	15745	10477
800	17606	7558	15646	15194	13036	10316	8172	8977	8378	3962
800	15189	14046	14909	14713	14999	14201	13184	13073	14639	10330
800	16388	14538	17548	19416	22034	17761	14488	16068	14773	10595

Table 2 *The data from Example 2: Glucose. Each line contains half-hourly plasma glucose concentrations (mmol/ℓ) for one subject on one day. S denotes subject number and T denotes treatment (variant of PYY). Data originate from Birgitte Sloth, Department of Human Nutrition, University of Copenhagen*

			Time after start of infusion (min)												
S	Group	*T*	0	30	60	90	120	150	180	210	240	270	300	330	360
1	lean	A	4.90	4.65	4.95	4.75	4.86	4.97	4.75	4.81	7.46	8.87	5.06	4.36	4.73
2	lean	A	4.78	4.71	4.70	4.65	4.60	4.68	4.60	4.61	4.96	6.92	5.41	4.43	4.28
3	lean	A	5.10	5.20	5.25	5.30	5.22	5.11	5.25	5.06	7.80	5.86	4.47	4.61	5.10
4	lean	A	5.05	5.24	5.06	5.02	5.24	5.16	4.95	5.00	7.45	6.10	4.54	4.43	5.29
5	lean	A	4.91	4.86	4.70	4.72	4.64	4.65	4.46	4.78	7.22	6.44	4.39	3.02	3.43
6	lean	A	5.24	5.31	5.32	5.23	5.31	5.30	4.41	5.11	7.37	4.01	4.73	5.54	5.76
7	lean	A	5.09	5.18	5.19	5.23	5.21	5.10	5.17	5.27	7.15	6.18	4.85	5.27	5.57
8	lean	A	5.23	5.20	5.09	5.19	4.98	5.16	4.98	5.09	6.50	6.19	6.42	6.50	5.71
9	lean	A	5.12	5.26	5.23	5.34	5.29	5.29	5.26	5.16	6.20	8.44	7.78	5.98	5.44
10	lean	A	5.11	5.18	4.94	4.89	4.90	4.90	4.90	4.92	5.94	4.54	3.75	3.70	4.28
11	lean	A	5.26	5.43	5.17	5.17	5.16	5.11	5.08	5.27	7.24	5.04	4.88	5.18	5.38
12	lean	A	4.92	4.79	4.95	4.93	5.26	5.24	5.35	5.37	7.52	9.69	7.22	5.36	5.13
13	obese	A	5.17	5.18	5.24	5.20	5.15	5.23	5.12	5.02	6.69	6.86	5.31	5.24	5.33
14	obese	A	5.39	5.30	5.14	5.08	4.95	5.00	5.01	4.85	5.77	6.01	5.42	4.95	5.08
15	obese	A	5.37	5.41	5.38	5.45	5.36	5.19	5.26	5.22	7.11	6.96	5.86	5.91	5.72
16	obese	A	5.02	5.20	4.91	5.07	5.05	4.95	4.71	4.91	6.01	8.27	6.34	5.13	5.09
17	obese	A	5.91	6.02	5.91	5.79	5.70	5.41	5.38	5.18	6.18	7.33	4.93	6.64	5.44
18	obese	A	6.15	6.21	6.13	6.06	5.99	5.82	5.76	5.66	7.54	7.60	6.82	5.91	5.86
19	obese	A	5.54	5.56	5.53	5.38	5.28	5.13	5.32	5.15	6.10	6.68	6.18	5.68	5.63
20	obese	A	5.53	5.53	5.48	5.49	5.47	5.55	5.49	5.40	6.29	8.54	4.78	5.52	3.07
21	obese	A	4.88	4.81	4.75	4.72	4.77	4.82	4.65	4.78	7.59	9.16	4.82	6.87	6.41
22	obese	A	5.80	5.61	5.53	5.70	5.33	5.44	5.25	5.23	6.80	6.75	5.77	4.30	5.02
23	obese	A	5.25	5.09	5.22	5.06	5.14	4.20	4.95	3.86	6.61	6.47	4.90	4.79	5.02
24	obese	A	5.19	5.02	5.00	5.00	4.95	4.95	4.99	4.92	6.07	7.02	5.43	4.93	4.83

Table 2 (*Continued*)

			Time after start of infusion (min)												
S	Group	*T*	0	30	60	90	120	150	180	210	240	270	300	330	360
1	lean	B	4.91	4.80	4.70	4.73	4.76	4.66	4.67	4.77	8.12	8.31	5.50	4.54	4.24
2	lean	B	5.11	5.01	4.84	4.68	4.79	4.64	4.50	4.72	5.44	6.34	5.09	4.20	4.75
3	lean	B	5.17	5.09	5.11	5.01	5.20	5.07	5.18	5.23	8.20	5.97	4.82	4.42	4.29
4	lean	B	4.99	5.08	4.96	4.90	4.96	5.01	5.00	4.96	7.38	4.55	3.87	4.54	5.30
5	lean	B	4.80	4.82	4.75	4.81	5.05	5.17	5.07	5.05	7.09	7.47	5.26	4.63	4.99
6	lean	B	5.18	5.24	5.15	5.22	5.24	5.19	5.11	5.32	7.70	4.35	5.14	5.51	4.92
7	lean	B	5.32	5.06	5.10	5.14	5.01	4.96	5.03	5.17	7.03	5.74	4.00	5.18	5.75
8	lean	B	5.32	5.15	5.24	5.12	5.16	5.28	5.14	5.83	6.62	4.90	3.99	4.80	5.95
9	lean	B	5.17	5.36	5.19	5.36	5.26	5.41	5.34	5.40	7.41	5.65	5.39	5.11	4.92
10	lean	B	4.74	5.00	4.77	4.75	4.69	4.58	4.72	4.67	5.94	4.38	3.17	3.02	3.83
11	lean	B	5.10	5.24	5.12	5.26	5.22	5.24	5.12	5.25	7.05	5.07	5.65	6.29	5.61
12	lean	B	5.05	5.03	5.09	5.11	5.24	5.29	5.31	5.29	7.76	9.35	7.52	6.11	5.11
13	obese	B	5.42	5.33	5.24	5.26	5.24	5.33	5.02	5.24	6.14	6.75	5.62	5.46	5.26
14	obese	B	5.41	5.51	5.33	5.21	5.15	5.11	5.09	4.99	6.32	6.43	5.61	5.02	5.41
15	obese	B	6.18	6.13	6.00	5.92	5.86	5.85	5.80	5.72	8.23	8.26	5.60	5.81	6.44
16	obese	B	5.45	5.32	5.20	5.14	5.06	5.18	4.94	4.97	6.55	7.03	6.65	6.27	6.26
17	obese	B	6.27	6.22	5.99	5.95	5.76	5.61	5.48	5.31	6.62	8.46	6.16	6.13	5.70
18	obese	B	6.16	6.15	5.92	5.84	5.68	5.57	5.52	5.57	7.90	8.43	7.14	5.74	5.70
19	obese	B	5.52	5.66	5.58	5.50	5.34	5.32	5.28	5.35	6.11	6.03	4.96	5.44	6.29
20	obese	B	5.36	5.22	5.25	5.31	5.29	5.15	5.11	5.05	6.72	8.28	6.02	4.89	5.14
21	obese	B	4.93	4.80	4.67	4.75	4.61	4.52	4.60	4.59	5.52	9.22	6.26	6.14	6.41
22	obese	B	5.36	5.28	5.33	5.41	5.29	5.32	5.27	5.38	6.53	8.10	5.30	6.10	5.17
23	obese	B	5.13	5.25	5.20	5.08	5.10	5.09	5.12	5.04	5.22	7.69	5.09	4.31	4.83
24	obese	B	5.05	4.99	5.03	4.97	4.59	4.76	4.76	4.76	5.20	6.15	5.41	5.26	5.06

Acknowledgments

Grete Østergaard and Birgitte Sloth, both from University of Copenhagen, are gratefully acknowledged for providing the data of the two examples. We thank three anonymous referees for constructive criticism and suggestions that led to an improved version of the paper.

References

Capra, W. B. and Müller, H. G. (1997). An accelerated-time model for response curves. *J. Amer. Statist. Assoc.* **92**, 72–83. MR1436099

Jones, M. C. and Rice, J. A. (1992). Displaying the important features of a large collection of similar curves. *The American Statistician* **46**, 140–145.

Hall, P., Müller, H.-G. and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34**, 1493–1517. MR2278365

Kenward, M. G. and Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**, 983–997.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53**, 233–243. MR1094283

SAS/STAT Statistical Software for Windows, Version 9.1. Copyright © 2002-2003 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biom. Bull.* **2**, 110–114.

Silverman, B. W. (1995). Incorporating parametric effects into functional principal components analysis. *J. Roy. Statist. Soc. Ser. B* **57**, 673–689. MR1354074

Sloth, B., Holst, J. J., Flint, A., Gregersen, T. and Astrup, A. (2007). Effects of PYY1-36 and PYY3-36 on appetite, energy intake, energy expenditure, glucose and fat metabolism in obese and lean. *American Journal of Physiology: Endocrinology and Metabolism* **292**, 1062–1068.

Soler, J. M. P. and Singer, J. M. (2000). Optimal covariance adjustment in growth curve models. *Comput. Statist. Data Anal.* **33**, 101–110.

Østergaard G., Lam H. R., Ladefoged, O. and Arlien-Søborg, P. (1993). Effects of six months white spirit inhalation exposure in adult and old rats. *Pharmacol. Toxicol.* **72**, 34–39.

Department of Basic Sciences
and Environment
Faculty of Life Science
University of Copenhagen
Thorvaldsensvej 40
DK-1871 Frederiksberg C
Denmark
E-mail: ims@life.ku.dk
tmartinussen@health.sdu.dk