

women's
health
australia



the australian longitudinal
study on women's health

Changes Report 2:

Examples from the Australian Longitudinal Study
on Women's Health for Analysing Longitudinal Data

june 2005



Changes Report 2: Examples from the Australian Longitudinal Study on Women's Health for Analysing Longitudinal Data

June 2005

**Australian Longitudinal Study on Women's Health
Report to the Department of Health and Ageing**

This report is based on the collective work of the Statistics Group of the Australian Longitudinal Study on Women's Health, and was written by Jessica Ford, Annette Dobson, Anne Young, Virginia Wheway, Anne Russell, and Jennifer Powers.

CONTENTS

PART 1: Models for Longitudinal Data	- 2 -
1.1 The Australian Longitudinal Study on Women's Health	- 2 -
1.2 Introduction.....	- 4 -
1.3 Scope.....	- 4 -
1.4 Marginal Models.....	- 5 -
1.5 Random Effects Models.....	- 5 -
1.6 Transition Models	- 7 -
PART 2: Use of SAS for analysing longitudinal data.....	- 9 -
2.1 Statistical Analysis Software (SAS)	- 9 -
2.2 Data Structure	- 9 -
2.3 Coding TIME	- 11 -
2.3.1 Baseline Coding.....	- 11 -
2.3.2 Centred Coding	- 12 -
2.3.3 Endpoint coding.....	- 12 -
2.4 General SAS code.....	- 13 -
2.5 Computational issues in implementing SAS Procedures.....	- 14 -
2.5.1 Specification of random effects	- 14 -
2.5.2 Convergence	- 14 -
2.5.3 Methods to randomly select a smaller sample to deal with non-convergence....	- 16 -
2.5.4 Estimates of group means	- 17 -
2.5.5 Missing waves and missing items.....	- 17 -
PART 3: Example of continuous outcome – measured at two times	- 18 -
3.1 Introduction.....	- 18 -
3.2 Model SAS Code	- 20 -
3.3 MODEL 1: PROC GENMOD unadjusted.....	- 22 -
3.4 MODEL 2: PROC GENMOD adjusted.....	- 25 -
3.5 MODEL 3: PROC MIXED unadjusted	- 28 -
3.5 MODEL 3: PROC MIXED unadjusted	- 28 -
3.6 MODEL 4: PROC MIXED adjusted	- 31 -
3.7 Summary of least squares means	- 35 -
PART 4: Plotting results using SAS	- 37 -
PART 5: Sample Size.....	- 41 -
5.1 Single Time Point Studies.....	- 41 -
5.2 Extension to Longitudinal Studies	- 43 -
PART 6: References.....	- 49 -

PART 1: Models for Longitudinal Data

1.1 The Australian Longitudinal Study on Women's Health

The Australian Longitudinal Study on Women's Health (ALSWH) – widely known as Women's Health Australia - is a longitudinal population-based survey, funded by the Australian Government Department of Health and Ageing. The project began in 1996 and examines the health of over 40,000 Australian women.

The ALSWH involves three large, nationally representative, cohorts of Australian women representing three generations:

- The Younger women, aged 18-23 when first recruited in 1996 (n = 14247), are now in their late 20s, the peak years for relationship formation, childbearing, and establishing adult health habits (eg physical activity, diet) and paid and unpaid work patterns.
- The Mid-age women, initially aged 45-50 (n = 13716), are now experiencing menopause, as well as changes in household structure, family care giving, and impending retirement, which are common at this life stage. Some are showing early signs of age-related physical decline, while some are adopting new health behaviours in preparation for a healthy old age.
- The Older women, aged 70-75 when first recruited (n = 12432), are now in their 80s and facing the physical, emotional and social challenges of old age.

Features of the study design include:

- Women were randomly selected from the Medicare database and invited to participate in the longitudinal study.
- Women in rural and remote areas of Australia were intentionally over-sampled to ensure adequate numbers for statistical analysis.

- After Survey 1 in 1996, the three age cohorts have been surveyed sequentially, one cohort per year, on a rolling basis since 1998 (see Figure 1.1).

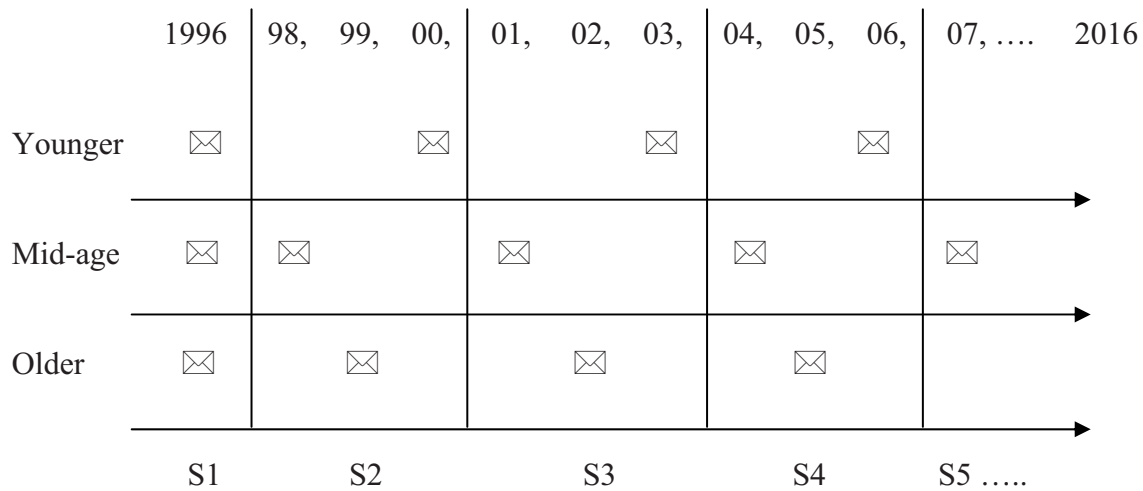


Figure 1.1: Timeline for main ALSWH Surveys

The study was designed to explore factors that influence health among women who are broadly representative of the entire Australian population. The study assesses:

- Physical and emotional health (including well-being, major diagnoses, symptoms)
- Use of health services (GP, specialist and other visits, access, satisfaction)
- Health behaviours and risk factors (diet, exercise, smoking, alcohol, other drugs)
- Time use (including paid and unpaid work, family roles, and leisure)
- Socio-demographic factors (location, education, employment, family composition)
- Life stages and key events (such as childbirth, divorce, widowhood).

The project provides a valuable opportunity to examine associations over time between aspects of women’s lives and their physical and emotional health. It provides an evidence base to the Australian Government Department of Health and Ageing – as well as other Australian and State/Territory Departments – for the development and evaluation of policy and practice in many areas of service delivery that affect women. An overview of the study and investigators, copies of the questionnaires, and abstracts of all publications and presentations, can be located on the Study’s website <http://www.sph.uq.edu.au/alswh>.

1.2 Introduction

This report deals mainly with the analysis of correlated data arising from repeated measurements, when the outcome measurements are continuous and are assumed to have a multivariate Normal distribution. Analogous methods are available for repeated measures of categorical outcome variables.

Generalised Estimating Equations (GEEs) and random effects models provide practical methods for these data.

1.3 Scope

This report summarises and illustrates the main statistical methods for analysing longitudinal data from the ALSWH.

The mathematical and statistical models generally applied to longitudinal data are described briefly. Examples from the ALSWH data demonstrate the implementation of these models in Statistical Analysis Software (SAS)¹. The GENMOD procedure can fit models to correlated responses by the GEE method. However, random effects models, fitted using the MIXED procedure, may be more appropriate for many situations. The interpretation and presentation of the results are also discussed.

Longitudinal analysis within the ALSWH typically involves an outcome variable Y_{ij} , measured for the i th subject at survey j . Measurements on the same subject at successive surveys are likely to be correlated because they refer to the same individual. They are also expected to be affected by a set of k covariates or explanatory variables $\{x_{ijk}\}$. For example, variable 1 x_{ij1} may be the weight of woman i at survey j ; x_{ij2} may be an indicator variable for whether or not she is a current smoker at survey j ; and $x_{i,3}$ may be an indicator variable denoting her country of birth, which does not vary across surveys. The responses Y_{ij} are assumed to have a distribution from the exponential family (which includes the Normal distribution for continuous scale measurements, the binomial distribution for dichotomous measures, the multinomial distribution for categorical variables with more than two

categories, and the Poisson distribution for count data). Any model will include the specification of the particular distribution for the Y_{ij} 's, an equation relating the expected value $E(Y_{ij})$ to the covariates, x_{ijk} , and some way of describing the correlation between repeated measurements on the same person.

There are three basic types of models: marginal models, random effects models and transition models. These can be summarised as follows:

1.4 Marginal Models

Marginal models are analogous to generalised linear models for independent data. Y_{ij} has some probability distribution f (e.g., f is Normal), with expected values $E(Y_{ij}) = \mu_{ij}$ which are linked to the covariates by the function $h(\mu_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta}$ where $\boldsymbol{\beta}$ is a vector of parameters of interest (i.e., the effects of the explanatory variables on the outcome Y_{ij}), \mathbf{x}_{ij} is the vector of covariates $\{x_{ijk}\}$ and responses for the same individual i are related by correlation terms $\rho(Y_{ij}, Y_{im}) = \rho_{ijm}$ (which are usually assumed to be the same for all individuals).

These models involve explicit specification of the probability distribution f , the link function h and the correlation structure. If the responses for the same individual were not correlated (i.e. $\rho_{ijm} = 0$) then the class of models called generalised linear models are used: these include analysis of variance and multiple regression if the Y_{ij} 's are continuous and various forms of logistic regression if the Y_{ij} 's are categorical. If the responses are correlated then the correlations have to be estimated and taken into account in estimating $\boldsymbol{\beta}$ – this involves a technique called Generalised Estimating Equations (GEEs). In SAS all of these models can be implemented using the procedure GENMOD.

1.5 Random Effects Models

An alternative way of coping with the correlation between responses from the same individual at different surveys is to define $h(\mu_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{U}_i$ where \mathbf{U}_i denotes unobserved “random” effects describing the characteristics of subject i which might be summarised over

all subjects by a probability distribution d . The terms z_{ij} are used to specify which components of the model are to be regarded as random effects that are not individually estimated, in contrast to the “fixed effects” described by the parameters β , which are estimated. The correlation between measurements Y_{ij} and Y_{im} on the same subject is “captured” in these models because the expected values μ_{ij} and μ_{im} both depend on the subject-specific random effects U_i .

Random effects models are also called mixed models because they actually involve both fixed effects β and random effects U_i . They are modelled in SAS using the procedure MIXED.

This quote is taken from the following website ²

http://www.umanitoba.ca/centres/mchp/concept/dict/Statistics/random_effects/random_effect_s.html

“Random effects models are also known as multi-level models, mixed models, random coefficient models, empirical Bayes models, and random regression models. Random effect models for longitudinal data are regression models in which the regression coefficients are allowed to vary across the subjects. These models have two components:

- Within-individual component: an individual’s change over time is described by a regression model with a population-level intercept and slope
- Between-individual component: variation in individual intercepts and slopes is captured.

For longitudinal studies, random effects models enable the analyst to not only describe the trend over time while taking account of the correlation that exists between successive measurements, but also to describe the variation in the baseline measurements and in the rate of change over time.

There are a number of techniques for analysing longitudinal data, unlike including univariate and multivariate ANOVAs and generalised linear models with generalised estimating equations (i.e., GEE models), in random effects analyses.

1. Subjects are not assumed to be measured on the same number of time points, and the time points do not need to be equally spaced;
2. Analyses can be conducted for subjects who may miss one or more of the measurement occasions, or who may be lost to follow-up at some point during the study.

Neither univariate nor multivariate analysis of variance can be easily applied to longitudinal data that contain time-varying covariates. Both random effects models and GEE models however allow for the inclusion of time-varying and time-invariant covariates. Time-varying covariates are independent variables that co-vary with the dependent variable over time.

Both random effects and GEE models allow the analyst to model the correlation structure of the data. Thus, the analyst does not need to assume that measurements taken at successive points in time are equally correlated, which is the correlation structure that underlies the ANOVA model. The analyst also does not need to assume measurements taken at successive points in time have an unstructured pattern of correlations, which is the structure that underlies the multivariate analysis of variance model. The former pattern is generally too restrictive, while the latter is too generic. With both random effects and GEE models, the analyst can fit a specific correlation structure to the data, such as an autoregressive structure, which assumes a decreasing correlation between successive measurements over time. This can result in a more efficient analysis, with improved power to detect significant changes over time.”

1.6 Transition Models

In the ALSWH we are interested in the effects of changes over time in covariates on the outcome variables in the same subjects. We are also interested in changes in the outcome variables themselves. If the outcome variable is categorical we usually define a new transition variable in terms of combinations of variables at individual surveys. If pet ownership at one survey is denoted by Y (yes) for owning a pet and N (no) for not owning a pet then the stability or change over two surveys can be defined by a categorical variable with four categories - YY, YN, NY, NN -, where for example NY is used to describe a subject

who did not own a pet at the first survey, but did own a pet at the second survey (and so became a pet owner between the surveys).

If the outcome variable is continuous (e.g., a scale measuring physical functioning) then, depending on the circumstances, we may use the measures at each survey or the difference between measures at different surveys (i.e., a measure of change in the model). It is important to also consider the initial value as a covariate to account for the effect known as regression to the mean.

If we are interested in changes or transitions in an explanatory variable then these can be included in the model explicitly as explanatory variables (such as change in weight, or change in pet ownership). The SAS procedures GENMOD and MIXED can also be used for these models but the specifications of terms may need particular care.

PART 2: Use of SAS for analysing longitudinal data

2.1 Statistical Analysis Software (SAS)

SAS is a statistical software package with extensive routines for statistical analysis, ranging from cross-tabulations and frequencies, to regression, analysis of variance (ANOVA), and longitudinal data analysis. SAS is a very flexible and powerful analytical tool which can accommodate a very large number of observations and variables and allows good control over statistical modelling.

SAS is syntax-orientated, but offers a limited menu-driven interface. SAS Online Documentation (<http://v8doc.sas.com/sashtml/>) offers extensive explanation of the syntax and procedures used throughout this document.

However, reliance on syntax makes SAS a slightly more difficult statistical package to use than Stata or SPSS.

2.2 Data Structure

Conventional data sets contain a single line of data (a record) for each subject or experimental unit. Fitting models for longitudinal data in SAS requires that each subject has as many lines of data as there are measurement occasions. For example, if data are recorded on n subjects for 3 surveys, the data should contain 3 records per subject. This is known as long format or stacked data. In the worked example below, 3 characteristics, each measured twice, for n subjects are transformed from standard format (n records and 7 variables – ID plus 2 variables for each characteristic) to long format ($2n$ records and 5 variables – ID, a single variable for each of the 3 characteristics plus a variable indicating the time that each characteristic was measured).

The example shows how to structure data for a continuous outcome and two categorical explanatory variables measured at two time points (in this example at surveys 2 and 3 of the older women). The outcome variable is a measure of physical functioning (PF) from the

Medical Outcomes Study Short-Form 36 (SF-36) Quality of Life questionnaire.³ Physical functioning is measured on a scale of 0-100, with higher scores representing better functioning. The explanatory variables are pet ownership (*pet*) which was collapsed into four categories (no pet; dog as a pet; cat as a pet; other pet) and ability to manage on income (*income*) which was also collapsed into four categories (easy; not too bad; sometimes difficult; difficult all the time and impossible).

The original data set (*petdata*) takes the conventional format, with 10 lines from 10 subjects, with one column of data for each variable at each survey.

Obs	IDalias	o2pf	o3pf	o2income	o3income	o2pet	o3pet
1	119	90.000	100.000	3 Not bad	3 Not bad	1 no pet	1 no pet
2	120	10.000	5.000	3 Not bad	1 Imp/Diff.always	2 dog pet	2 dog pet
3	121	20.000	0.000	2 Diff.smtmes	3 Not bad	1 no pet	1 no pet
4	122	55.000	61.111	2 Diff.smtmes	2 Diff.smtmes	1 no pet	1 no pet
5	123	70.000	65.000	3 Not bad	4 Easy	1 no pet	1 no pet
6	124	60.000	35.000	3 Not bad	2 Diff.smtmes	1 no pet	1 no pet
7	125	95.000	100.000	4 Easy	4 Easy	1 no pet	1 no pet
8	126	75.000	75.000	3 Not bad	3 Not bad	1 no pet	1 no pet
9	127	70.000	85.000	3 Not bad	4 Easy	4 other pet	4 other pet
10	128	90.000	90.000	3 Not bad	2 Diff.smtmes	1 no pet	1 no pet

The following SAS code transforms the dataset *pets* to long format.

```

data pets;
  set petdata;
  /*each 'set' of covariates is assigned an array*/
  array pfn {2} o2pf o3pf;
  array petn {2} o2pet o3pet;
  array incomen {2} o2income o3income;
  /*a do loop is then used to run through and rename each 'set' with
  one name and a 'time' value so that o2 and o3 'x' are merely
  distinguished by the value of time*/
  do time=1 to 2;
    pf = pfn{time};
    pet = petn{time};
    income = incomen{time};
    output;
  end;
  keep ID time pf pet income;
run;

/*sort the dataset by ID*/
proc sort data = pets; by ID; run;

```

The newly created data set, *petstack*, now appears on 20 lines, with one column for each variable and a time variable indicating the corresponding survey.

Obs	IDalias	income	pet	time	pf
1	119	3 Not bad	1 no pet	1	90.000
2	119	3 Not bad	1 no pet	2	100.000
3	120	3 Not bad	2 dog pet	1	10.000
4	120	1 Imp/Diff.always	2 dog pet	2	5.000
5	121	2 Diff.smtmes	1 no pet	1	20.000
6	121	3 Not bad	1 no pet	2	0.000
7	122	2 Diff.smtmes	1 no pet	1	55.000
8	122	2 Diff.smtmes	1 no pet	2	61.111
9	123	3 Not bad	1 no pet	1	70.000
10	123	4 Easy	1 no pet	2	65.000
11	124	3 Not bad	1 no pet	1	60.000
12	124	2 Diff.smtmes	1 no pet	2	35.000
13	125	4 Easy	1 no pet	1	95.000
14	125	4 Easy	1 no pet	2	100.000
15	126	3 Not bad	1 no pet	1	75.000
16	126	3 Not bad	1 no pet	2	75.000
17	127	3 Not bad	4 other pet	1	70.000
18	127	4 Easy	4 other pet	2	85.000
19	128	3 Not bad	1 no pet	1	90.000
20	128	2 Diff.smtmes	1 no pet	2	90.000

2.3 Coding *TIME*

When fitting models to analyse longitudinal data, the method selected for coding the time variable (*time*) will influence the interpretation of the model. Three options are discussed, with examples of data from the first four surveys of each ALSWH cohort.

The surveys were (and will be) conducted in the years shown in table 2.3.0.

Table 2.3.0: Timeline for the ALSWH Surveys

	Survey 1	Survey 2	Survey 3	Survey 4
Younger	1996	2000	2003	2006
Mid-age	1996	1998	2001	2004
Older	1996	1999	2002	2005

2.3.1 Baseline Coding

Code *time* so that the baseline measure has a value of zero and successive measurements are incremented by the number of years between surveys. Using this coding, the intercept in the model represents the mean value of the dependent variable at baseline.

Table 2.3.1: Baseline coding for the first four surveys

	Survey 1	Survey 2	Survey 3	Survey 4
Younger	0	4	7	10
Mid-age	0	2	5	8
Older	0	3	6	9

2.3.2 Centred Coding

Code *time* by centring the values. Using this format the intercept represents the dependent variable measurement at the midpoint of the period of analysis. In the ALSWH time could be centred around the mid-point between the first and last survey analysed.

Table 2.3.2: Centred coding for the first four surveys

	Year S1	Year S4	Middle S1-S4	Mid- point	Survey 1	Survey 2	Survey 3	Survey 4
Younger	1996	2006	2001	5	-5	-1	2	5
Mid-age	1996	2004	2000	4	-4	-2	1	4
Older	1996	2005	2000 / 2001	4.5	-4.5	-1.5	1.5	4.5

2.3.3 Endpoint coding

Code *time* so that the last measured time point has a value of zero and preceding measurements are decremented accordingly. Using this format, the intercept represents the mean value of the dependent variable at the endpoint.

Table 2.3.3 Endpoint coding for the first four surveys

	Survey 1	Survey 2	Survey 3	Survey 4
Younger	10	6	3	0
Mid-age	8	6	3	0
Older	9	6	3	0

It would be difficult to run a *DO* loop in the preceding data step using these non-consecutive numbers. The solution is to run a *DO* loop on the number of surveys you will use (i.e. for four surveys: *DO time = 1 to 4*) and then recode the time variable (in a new data set) to reflect one of the three coding choices from above.

2.4 General SAS code

Standardised SAS code for longitudinal analysis using either the GENMOD or MIXED procedure is shown below. The required SAS terms are given in upper case, whilst user specified elements are in lowercase.

Extensive explanation of the code is available via the SAS Online Documentation.

PROC GENMOD

```
PROC GENMOD DATA = data-set-name;  
CLASS variables;  
MODEL dependent-variable = <explanatory variables> </options>;  
REPEATED subject=ID </options eg. type=correlation-structure >;  
LSMEANS <explanatory variables> </options>;  
RUN;QUIT;
```

PROC MIXED

```
PROC MIXED DATA = data-set-name;  
CLASS variables;  
MODEL dependent-variable = <fixed effects></options>;  
RANDOM random-effects </options eg. type=correlation-structure >;  
LSMEANS fixed effects </options>;  
parms (value-for-covariance-parameter-estimate-INTERCEPT)  
      (value-for-covariance-parameter-estimate-RESIDUAL);  
RUN;QUIT;
```

The RANDOM statement in the MIXED procedure identifies the random effects and the MODEL statement identifies effects that are assumed to be fixed.

2.5 Computational issues in implementing SAS Procedures

2.5.1 Specification of random effects

An essential part of a mixed or random effects model is specifying which of the covariates or variables are random and which are fixed. Random effects are variables with values assumed to be from a random sample of subjects drawn from and representative of a larger population of values or observations. In contrast, fixed effects are variables which are the same across all subjects and are estimated in the model.

In the case of fixed categorical effects, we are usually interested in making explicit comparisons of one level against another, while for random factors the interest lies more in the extent to which the factor accounts for variance in the dependent variable. For example, in the analysis of the effect of pet ownership on physical functioning among a sample of 7000 women, pet ownership is taken as a fixed effect, and we are interested in comparing the mean physical functioning across the different levels of ownership and the unique ID for each of the 7000 women would be treated as a random factor. If the comparison of the levels of a variable against one another is the goal, then these variables should be treated as fixed effects.

Random effects can be included in a mixed model using the RANDOM statement. For example, a random intercept together with subject ID specification results in random effects being deviations from the fixed means. The line of code below specifies that each subject has their own intercept value.

```
RANDOM intercept / subject= ID;
```

2.5.2 Convergence

SAS may stop fitting a model early with warning '*stopped due to infinite likelihood*' if there are convergence problems.

Explanation of convergence problems is available via the SAS Online Documentation.

Convergence problems in PROC MIXED may arise from estimating the covariance parameters in the model, not the fixed effects. For example, when the covariance parameters are on a different scale, the algorithm in PROC MIXED may have trouble converging. Furthermore, if there is very little variability in the time effects the variance of the random slopes may approach 0, which may generate numerical difficulties.

Common causes of non-convergence:

- two of the covariance parameters are several orders of magnitude apart,
- data values are extremely large or extremely small,
- there is little variability in time effects,
- there is not enough data to estimate the specified covariance structure,
- linear dependencies exist among parameters,
- there is a mis-specified model or violation of model assumptions.

Recommendations to deal with non-convergence:

1. use the PARMS statement to input initial values;
2. rescale the data to improve stability;
3. specify the SCORING= option to invoke Fisher's scoring estimation method;
4. tune the maximum iteration (MAXITER= and MAXFUNC=) options in the PROC MIXED statement;
5. make sure no observations from the same subject are producing identical rows in the R or V matrix;
6. reduce the number of terms in the model.

In relation to the first two recommendations above:

1. Using the PARMS statement: Experience in the ALSWH suggests it may be practical to choose values based on results from an analysis of a sample of the women. The values are chosen from 'Covariance Parameter Estimates' – the first is the intercept, the second is the residual value estimate – these values give the model an initial starting point.

2. Centring the data to improve stability: this can be achieved, for example, with mental health (MH), by subtracting 80 (an approximate mean) from every value (this results in values more or less centred on zero). Then add 80 (or whatever value was used) to the final LSMEANS score.
3. Rescaling the data could also improve stability, that is subtracting a value that is approximately the mean and then by dividing by some value that is approximately the standard deviation.

2.5.3 Methods to randomly select a smaller sample to deal with non-convergence

As mentioned above, if you experience problems of non-convergence with your model it can be useful to select a smaller sample to obtain starting estimates. **This sample should be selected prior to creating the long format data structure.**

The SURVEYSELECT procedure in SAS provides a variety of methods for selecting probability-based random samples. The following code provides a basic outline for selecting a random sample of subjects.

```
proc surveyselect data = data-set-name
                 method=SRS
                 rep=1
                 sampsize=7000
                 seed=12345
                 out=output-data-set-name;
ID _all_;
run;
```

Method specifies the method for sample selection - SRS requests Simple Random Sampling, that is selection with equal probability and without replacement.

Rep specifies the number of separate samples to be taken.

Sample size specifies the number of units selected for the sample

Seed specifies the initial value for random number generation (useful for reproducing identical samples by using the same initial value).

Out names the output data set that contains the sample.

ID names variables from the input data set to be included in the output data set of selected units - here *_all_* indicates that all initial variables are to be kept.

Further information can be obtained from the SAS Online Documentation files on the SURVEY SELECT procedure.

2.5.4 Estimates of group means

By default LSMEANS in PROC GENMOD gives equal weights to all classification effects (that is, it assumes equal numbers in each group) and so the adjusted estimates of group means may be incorrect. Therefore we recommend PROC MIXED instead as it has the observed margins (OM) option, which weights means in proportion to group size.

2.5.5 Missing waves and missing items

If a subject does not participate at all in a particular survey (or “wave”) then this person’s data for all other surveys are omitted from the analysis if PROC GENMOD is used. In contrast PROC MIXED is not affected by a missing wave of data but the data are assumed to be missing at random (which may not be a reasonable assumption).

If a subject has missing data on one item (or more) from a particular survey then all the data from that survey are omitted from the analysis by both procedures.

PART 3: Example of continuous outcome – measured at two times

3.1 Introduction

Relationship between companion animals and physical health in older women.

Research Question: Do older women who have companion animals have better physical functioning on average over time than older women without companion animals?

In this analysis we examine both cross-sectional and longitudinal associations between pet ownership and physical health whilst controlling for ability to manage on current income.

To estimate the effects of pet ownership on Physical Functioning (PF), models were used to analyse the data from Surveys 2 and 3 simultaneously. Fitting models, using both the MIXED and GENMOD procedures, we estimated mean PF (and 95% confidence limits) for each level of pet ownership at each survey, adjusted for ability to manage on income.

Table 3.1.1 indicates the group size and mean Physical Functioning (mean PF) for each category of pet ownership and income, for the two times.

Table 3.1.1 Frequency and mean Physical Functioning for categories of Pet Ownership and Ability to manage on income.

Time	Pet Ownership			Ability to manage on income		
	Category	Frequency	Mean PF	Category	Frequency	Mean PF
1	No Pet	4450	64.0	Easy	1657	70.2
1	Dog Pet	1458	63.7	Not too bad	3628	63.8
1	Cat Pet	653	62.9	Difficult Sometimes	1364	57.8
1	Other Pet	439	60.2	Impossible	351	52.9
2	No Pet	4652	58.8	Easy	1705	64.6
2	Dog Pet	1352	57.7	Not too bad	3607	58.7
2	Cat Pet	613	58.9	Difficult Sometimes	1216	52.7
2	Other Pet	383	57.7	Impossible	472	50.7

For the transition analysis of change in pet ownership (*o23pet*), four categories were created: pet at surveys 2 and 3; no pet at surveys 2 and 3; pet at survey 2 but not at survey 3; and pet at survey 3 but not at survey 2.

Table 3.1.2 indicates the group size and mean Physical Functioning (mean PF) at two times for each of the categories in the change in pet ownership variable between the two surveys.

Table 3.1.2 Frequency and mean Physical Functioning for categories of Change in Pet Ownership

Change in Pet Ownership (<i>o23pet</i>)			
Category	Frequency	Mean PF	
		Time 1	Time 2
Yes → Yes	1992	63.4	58.7
No → No	4094	64.3	59.2
Yes → No	558	61.0	55.6
No → Yes	356	60.6	54.3

The SAS output for the four models (PROC GENMOD and PROC MIXED – each unadjusted and adjusted) are annotated in sections 3.3 to 3.6. Section 3.7 identifies the differences in least squares means (ls means) produced from the GEE (GENMOD) and random effects (MIXED) procedures.

3.2 Model SAS Code

Here is the SAS code for four different analyses using dataset *petstack*:

MODEL 1: GEE model to estimate the marginal means for physical functioning by pet ownership at each survey.

```
Proc genmod data = petstack;  
class IDalias time pet ;  
model pf= time time*pet / type3;  
repeated subject=IDalias / type=EXCH;  
lsmeans time time*pet / cl e;  
run; quit;
```

- i. **Type 3** specifies score statistics and p-values which adjust for all other terms in the model.
- ii. The **repeated** command indicates there are repeated data for each subject.
- iii. Type **EXCH** specifies an exchangeable working correlation matrix (which is the only option for just two measurement times, but there are other options for more than two repeated measures; see SAS Online Documentation).
- iv. The **lsmeans** command produces mean estimates of PF for each of the pet ownership categories at each time. .

MODEL 2: GEE model for marginal means for physical functioning by pet ownership, adjusted for the ability to manage on income variable at both times.

```
Proc genmod data = petstack;  
class IDalias time pet income;  
model pf= time time*pet time*income / type3;  
repeated subject=IDalias / type=EXCH;  
lsmeans time time*pet time*income / cl e;  
run; quit;
```

- v. The **lsmeans** command produces mean estimates of PF for each of the pet ownership categories at each time, adjusted for ability to manage on income. However the numbers of subjects in each category of

ability to manage on income are assumed to be equal (which they are not).

MODEL 3: Random effects models with pet ownership and time as fixed effects and subject as a random effect.

```
Proc mixed data = petstack ;
class IDalias time pet;
model pf= time time*pet / cl solution;
random intercept / type = VC subject=IDalias;
lsmeans time time*pet / om cl e;
run; quit;
```

- vi. **Solution** gives the estimates of fixed effects
- vii. **VC** gives the variance components corresponding to the random effects
- viii. The **om** option for the **lsmeans** command ensures that the correct means are estimated.

MODEL 4: Random effects model for physical functioning and pet ownership, adjusted for the ability to manage on income variable at both times.

```
proc mixed data = petstack ;
class IDalias time pet income;
model pf= time time*pet time*income / cl solution;
random intercept / type = VC subject=IDalias;
lsmeans time time*pet time*income / om cl e;
run; quit;
```

Models incorporating the change variable for pet ownership between two surveys use similar syntax – with the variable for change in pet ownership between surveys (*o23pet*) replacing the time specific values. For example, the random effects model, adjusted for ability to manage on income, would have the following SAS syntax:

```
proc mixed data = petstack2 ;
class IDalias time o23pet income;
model pf= time time*o23pet time*income / cl solution;
random intercept / type = VC subject=IDalias;
lsmeans time time*o23pet time*income / om cl e;
run; quit;
```

3.3 MODEL 1: PROC GENMOD *unadjusted*

The *model information* describes the model, the covariates that are included, and the method used to fit the model. Note that there are 14000 observations, but only 7000 subjects (IDalias) – this results from reformatting to long format over two time periods, so that the number of observations is two times 7000. There are 11 parameters to be estimated: the intercept, each of the two times and each of the four pet ownership categories at each of the two times.

Model Information			
Data Set	WORK.PETSTACK		
Distribution	Normal		
Link Function	Identity		
Dependent Variable	pf		
Number of Observations Read	14000		
Number of Observations Used	14000		
Class Level Information			
Class	Levels	Values	
IDalias	7000	300083 300088 300089 300090 300091 300093 300094 300095 300096 300100 300101 300102 300103 300107...	
time	2	1 2	
pet	4	1NoPet 2DogPet 3CatPet 4OtherPet	
Parameter Information			
Parameter	Effect	time	pet
Prm1	Intercept		
Prm2	time	1	
Prm3	time	2	
Prm4	time*pet	1	1NoPet
Prm5	time*pet	1	2DogPet
Prm6	time*pet	1	3CatPet
Prm7	time*pet	1	4OtherPet
Prm8	time*pet	2	1NoPet
Prm9	time*pet	2	2DogPet
Prm10	time*pet	2	3CatPet
Prm11	time*pet	2	4OtherPet

The preferred *criteria for assessing goodness of fit* are the scaled Deviance and scaled Pearson statistics. Both indicate reasonable fit if Value/DF is close to one.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	14E3	9621494.2472	687.6425
Scaled Deviance	14E3	14000.0000	1.0006
Pearson Chi-Square	14E3	9621494.2472	687.6425
Scaled Pearson x2	14E3	14000.0000	1.0006
Log Likelihood		-65594.0222	
Algorithm converged.			

Initial Parameter Estimates are an intermediate step in the model fitting and can usually be ignored. The parameter estimates comprise an overall mean (called the intercept) and corner point estimates of the other effects with “other pet” as the reference category. The reference category is set by default or can be specified on either theoretical grounds or sample size considerations.

Analysis of Initial Parameter Estimates								
Parameter	DF	Estimate	Standard Error	Wald	95% Confidence Limits	Chi-Square	Pr>ChiSq	
Intercept	1	57.7395	1.3395	55.1140	60.3650	1857.93	<.0001	
time	1	2.4719	1.8330	-1.1207	6.0645	1.82	0.1775	
time	2	0.0000	0.0000	0.0000	0.0000	.	.	
time*pet	1	1NoPet	3.8159	1.3115	1.2454	6.3863	8.47	0.0036
time*pet	1	2DogPet	3.4474	1.4272	0.6502	6.2447	5.83	0.0157
time*pet	1	3CatPet	2.6590	1.6180	-0.5123	5.8302	2.70	0.1003
time*pet	1	40thPet	0.0000	0.0000	0.0000	0.0000	.	.
time*pet	2	1NoPet	1.0667	1.3936	-1.6647	3.7981	0.59	0.4440
time*pet	2	2DogPet	-0.0679	1.5175	-3.0421	2.9063	0.00	0.9643
time*pet	2	3CatPet	1.1206	1.7075	-2.2260	4.4672	0.43	0.5116
time*pet	2	40thPet	0.0000	0.0000	0.0000	0.0000	.	.
Scale	1	26.2154	0.1567	25.9102	26.5243	.	.	

NOTE: The scale parameter was estimated by maximum likelihood.

GEE Model Information
Correlation Structure Exchangeable
Subject Effect IDAlias (7000 levels)
Number of Clusters 7000
Correlation Matrix Dimension 2
Maximum Cluster Size 2
Minimum Cluster Size 2
Algorithm converged.

The parameter estimates from the GEE model are shown below. The correlation value of 0.74 shows that the PF values for the same subjects at the two surveys are highly correlated.

Exchangeable Working Correlation								
Correlation	0.7431344935							
Analysis of GEE Parameter Estimates Empirical Standard Error Estimates								
Parameter	Estimate	Standard Error	Wald	95% Confidence Limits	Z	Pr > Z		
Intercept	58.3792	1.0349	56.3509	60.4076	56.41	<.0001		
time	1	3.5104	1.0769	1.3998	5.6210	3.26	0.0011	
time	2	0.0000	0.0000	0.0000	0.0000	.	.	
time*pet	1	1NoPet	1.9859	0.9415	0.1407	3.8312	2.11	0.0349
time*pet	1	2DogPet	1.6711	1.0578	-0.4023	3.7444	1.58	0.1142
time*pet	1	3CatPet	1.1058	1.1874	-1.2215	3.4331	0.93	0.3517
time*pet	1	4OtherPet	0.0000	0.0000	0.0000	0.0000	.	.
time*pet	2	1NoPet	0.2952	1.0495	-1.7618	2.3522	0.28	0.7785
time*pet	2	2DogPet	-0.2271	1.1889	-2.5574	2.1032	-0.19	0.8485
time*pet	2	3CatPet	0.0209	1.3605	-2.6457	2.6874	0.02	0.9878
time*pet	2	4OtherPet	0.0000	0.0000	0.0000	0.0000	.	.

Type 3 *score statistics* indicate the p-values for the null hypotheses, that there is no effect of time, or of pet ownership by time. The two p-values obviously indicate the strong effect of time on physical functioning ($p < 0.0001$), but no effect of pet ownership over time on physical functioning ($p = 0.4089$).

Score Statistics For Type 3 GEE Analysis			
Source	DF	Chi-Square	Pr > ChiSq
time	1	149.41	<.0001
time*pet	6	6.13	0.4089

The coefficients for the least squares means are shown in the output below, these are used to calculate the means. For example, for no pet at time 1 the estimate is $P1 * 58.3792 + P2 * 3.5104 + P3 * 0 + P4 * 1.9859 + P5 * 0 + P6 * 0 \dots = 63.876$. More generally the mean PF for

each of the 8 groups is obtained by multiplying the matrix of 1s and 0s below by the vector of parameter estimates in the box with *Analysis of GEE Parameter Estimates*.

Coefficients for time Least Squares Means												
Label	Row	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
time	1	1	1	0	0.25	0.25	0.25	0.25	0	0	0	0
time	2	1	0	1	0	0	0	0	0.25	0.25	0.25	0.25
time*pet	1	1	1	0	1	0	0	0	0	0	0	0
time*pet	2	1	1	0	0	1	0	0	0	0	0	0
time*pet	3	1	1	0	0	0	1	0	0	0	0	0
time*pet	4	1	1	0	0	0	0	1	0	0	0	0
time*pet	5	1	0	1	0	0	0	0	1	0	0	0
time*pet	6	1	0	1	0	0	0	0	0	1	0	0
time*pet	7	1	0	1	0	0	0	0	0	0	1	0
time*pet	8	1	0	1	0	0	0	0	0	0	0	1

The resulting estimates (and corresponding confidence limits, etc) are shown below. The chi-square tests are testing hypotheses that the mean PF values for each category are zero which are obviously rejected as the p-values are less than 0.0001. Note that the significant effect of 'time' on PF results in the means at time 2 being much lower than the means at time 1.

Least Squares Means										
Effect	time	pet	Estimate	Standard Error	DF	Chi-Square	Pr>ChiSq	Alpha	Confidence	Limits
time	1		63.0803	0.3801	1	27542	<.0001	0.05	62.3353	63.8253
time	2		58.4015	0.4376	1	17807	<.0001	0.05	57.5437	59.2593
time*pet	1	1NoPet	63.8756	0.3376	1	35794	<.0001	0.05	63.2138	64.5373
time*pet	1	2DogPet	63.5607	0.5636	1	12720	<.0001	0.05	62.4561	64.6652
time*pet	1	3CatPet	62.9954	0.7718	1	6661.3	<.0001	0.05	61.4826	64.5082
time*pet	1	40thPet	61.8896	0.9316	1	4413.8	<.0001	0.05	60.0638	63.7154
time*pet	2	1NoPet	58.6745	0.3812	1	23689	<.0001	0.05	57.9273	59.4216
time*pet	2	2DogPet	58.1521	0.6534	1	7921.2	<.0001	0.05	56.8715	59.4327
time*pet	2	3CatPet	58.4001	0.9335	1	3914.1	<.0001	0.05	56.5705	60.2296
time*pet	2	40thPet	58.3792	1.0349	1	3182.2	<.0001	0.05	56.3509	60.4076

3.4 MODEL 2: PROC GENMOD adjusted

The table below shows the 19 parameters that are fitted for time, time by pet ownership at each survey, and time by ability to manage on income at each survey (together with a mean effect or ‘intercept’).

Model Information				
Data Set		WORK.PETSTACK		
Distribution		Normal		
Link Function		Identity		
Dependent Variable		pf		
Number of Observations Read		14000		
Number of Observations Used		14000		
Class Level Information				
Class	Levels	Values		
Idalias	7000	300083 300088 300089 300090 300091 300093 300094		
		300095 300096 300100 300101 300102 300103 300107...		
time	2	1 2		
pet	4	1NoPet 2DogPet 3CatPet 4OtherPet		
income	4	1ImpDA 2DStime 3NotBad 4Easy		
Parameter Information				
Parameter	Effect	time	pet	income
Prm1	Intercept			
Prm2	time	1		
Prm3	time	2		
Prm4	time*pet	1	1NoPet	
Prm5	time*pet	1	2DogPet	
Prm6	time*pet	1	3CatPet	
Prm7	time*pet	1	4OtherPet	
Prm8	time*pet	2	1NoPet	
Prm9	time*pet	2	2DogPet	
Prm10	time*pet	2	3CatPet	
Prm11	time*pet	2	4OtherPet	
Prm12	time*income	1		1ImpDA
Prm13	time*income	1		2DStime
Prm14	time*income	1		3NotBad
Prm15	time*income	1		4Easy
Prm16	time*income	2		1ImpDA
Prm17	time*income	2		2DStime
Prm18	time*income	2		3NotBad
Prm19	time*income	2		4Easy

As discussed for the previous model, the *criteria for assessing goodness of fit*, the scaled Deviance and scaled Pearson statistics indicate very good fit with Value/DF close to one.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	14E3	9332649.4327	667.2851
Scaled Deviance	14E3	14000.0000	1.0010
Pearson Chi-Square	14E3	9332649.4327	667.2851
Scaled Pearson X2	14E3	14000.0000	1.0010
Log Likelihood		-65380.6577	
Algorithm converged.			

As in the previous model, initial parameter estimates for the GEE model are displayed by SAS and should not be confused with the final *Least Squares Means*. This is intermediate output, which can usually be ignored, and has been omitted.

The box below shows the final estimates.

```

GEE Model Information
Correlation Structure              Exchangeable
Subject Effect                    IDalias (7000 levels)
Number of Clusters                 7000
Correlation Matrix Dimension      2
Maximum Cluster Size              2
Minimum Cluster Size              2

Algorithm converged.

Exchangeable Working Correlation
Correlation    0.735338953

Analysis Of GEE Parameter Estimates Empirical Standard Error Estimates
                    Standard 95% Confidence
Parameter          Estimate Error Limits          Z Pr > |Z|
Intercept          61.9711  1.1202  59.7756  64.1667  55.32 <.0001
time               3.8259  1.1660  1.5406  6.1113  3.28  0.0010
time               0.0000  0.0000  0.0000  0.0000  .    .
time*pet           1 1NoPet  1.7158  0.9389  -0.1243  3.5559  1.83  0.0676
time*pet           1 2DogPet  1.8493  1.0560  -0.2205  3.9191  1.75  0.0799
time*pet           1 3CatPet  1.0192  1.1809  -1.2953  3.3337  0.86  0.3881
time*pet           1 4OtherPet 0.0000  0.0000  0.0000  0.0000  .    .
time*pet           2 1NoPet  0.1223  1.0555  -1.9466  2.1911  0.12  0.9078
time*pet           2 2DogPet -0.1345  1.1931  -2.4729  2.2039 -0.11  0.9102
time*pet           2 3CatPet  0.0262  1.3629  -2.6451  2.6975  0.02  0.9846
time*pet           2 4OtherPet 0.0000  0.0000  0.0000  0.0000  .    .
time*income        1 1ImpDA  -9.9168  1.1017 -12.0760 -7.7575 -9.00 <.0001
time*income        1 2DStime -6.9883  0.6835  -8.3280 -5.6487 -10.22 <.0001
time*income        1 3NotBad  -3.6771  0.5037  -4.6643 -2.6899 -7.30 <.0001
time*income        1 4Easy    0.0000  0.0000  0.0000  0.0000  .    .
time*income        2 1ImpDA  -7.6573  1.0816  -9.7772 -5.5374 -7.08 <.0001
time*income        2 2DStime -6.9578  0.7984  -8.5226 -5.3931 -8.72 <.0001
time*income        2 3NotBad  -3.4356  0.5717  -4.5562 -2.3151 -6.01 <.0001
time*income        2 4Easy    0.0000  0.0000  0.0000  0.0000  .    .

```

Type 3 *Score statistics* and p-values provide tests of the null hypotheses of no effect of pet ownership (adjusted for ability to manage on income), and the effect of ability to manage on income (adjusted for pet ownership). These results indicate that after considering the potential confounding of ability to manage on income, pet ownership is still not significant.

```

Score Statistics For Type 3 GEE Analysis
Source          DF   Chi-Square   Pr>ChiSq
time            1    93.74        <.0001
time*pet        6     5.26        0.5110
time*income     6    147.88       <.0001

```

The following box shows the multipliers used for the least squares means calculations. Note that each category of ability to manage on income is given an equal weight of 0.25, suggesting there are equal numbers of subjects in each category, which is incorrect, as shown in the table in Section 3.1.

Coefficients for time Least Squares Means																				
Label	Row	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19
time	1	1	1	0	0.25	0.25	0.25	0.25	0	0	0	0	0.25	0.25	0.25	0.25	0	0	0	0
time	2	1	0	1	0	0	0	0	0.25	0.25	0.25	0	0	0	0	0	0.25	0.25	0.25	0.25
time*pet	1	1	1	0	1	0	0	0	0	0	0	0	0.25	0.25	0.25	0.25	0	0	0	0
time*pet	2	1	1	0	0	1	0	0	0	0	0	0	0.25	0.25	0.25	0.25	0	0	0	0
time*pet	3	1	1	0	0	0	1	0	0	0	0	0	0.25	0.25	0.25	0.25	0	0	0	0
time*pet	4	1	1	0	0	0	0	1	0	0	0	0	0.25	0.25	0.25	0.25	0	0	0	0
time*pet	5	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0.25	0.25	0.25	0.25
time*pet	6	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0.25	0.25	0.25	0.25
time*pet	7	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0.25	0.25	0.25	0.25
time*pet	8	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0.25	0.25	0.25	0.25
time*income	1	1	1	0	0.25	0.25	0.25	0.25	0	0	0	0	1	0	0	0	0	0	0	0
time*income	2	1	1	0	0.25	0.25	0.25	0.25	0	0	0	0	0	1	0	0	0	0	0	0
time*income	3	1	1	0	0.25	0.25	0.25	0.25	0	0	0	0	0	0	1	0	0	0	0	0
time*income	4	1	1	0	0.25	0.25	0.25	0.25	0	0	0	0	0	0	0	1	0	0	0	0
time*income	5	1	0	1	0	0	0	0	0	0.25	0.25	0.25	0.25	0	0	0	1	0	0	0
time*income	6	1	0	1	0	0	0	0	0	0.25	0.25	0.25	0.25	0	0	0	0	1	0	0
time*income	7	1	0	1	0	0	0	0	0	0.25	0.25	0.25	0.25	0	0	0	0	0	1	0
time*income	8	1	0	1	0	0	0	0	0	0.25	0.25	0.25	0.25	0	0	0	0	0	0	1

The least squares adjusted means shown in the table below will be inaccurate if they do not correctly reflect the distribution of ‘ability to manage on income’ among ALSWH participants. The sizes of the differences among the marginal means are maintained although the absolute values may differ to those obtained using multipliers that correctly reflect the actual distribution of ‘ability to manage on income’.

Effect	Least Squares Means		Standard Error	DF	Chi-Square	Pr>ChiSq	Alpha	Confidence Limits		
	time	pet/income						Estimate	Lower	Upper
time	1		61.7976	0.4379	1	19913	<.0001	0.05	60.9393	62.6559
time	2		57.4620	0.4814	1	14247	<.0001	0.05	56.5184	58.4055
time*pet	1	1NoPet	62.3673	0.4065	1	23544	<.0001	0.05	61.5707	63.1640
time*pet	1	2DogPet	62.5008	0.5980	1	10923	<.0001	0.05	61.3287	63.6729
time*pet	1	3CatPet	61.6707	0.7979	1	5974.6	<.0001	0.05	60.1070	63.2345
time*pet	1	40thPet	60.6515	0.9555	1	4029.4	<.0001	0.05	58.7788	62.5242
time*pet	2	1NoPet	57.5807	0.4366	1	17395	<.0001	0.05	56.7250	58.4364
time*pet	2	2DogPet	57.3240	0.6775	1	7159.7	<.0001	0.05	55.9961	58.6518
time*pet	2	3CatPet	57.4847	0.9544	1	3627.6	<.0001	0.05	55.6140	59.3553
time*pet	2	40thPet	57.4585	1.0585	1	2946.5	<.0001	0.05	55.3838	59.5331
time*income	1	1ImpDA	57.0264	1.0418	1	2996.4	<.0001	0.05	54.9845	59.0682
time*income	1	2DStime	59.9548	0.5946	1	10168	<.0001	0.05	58.7894	61.1202
time*income	1	3NotBad	63.2660	0.4214	1	22536	<.0001	0.05	62.4400	64.0920
time*income	1	4Easy	66.9431	0.5279	1	16079	<.0001	0.05	65.9084	67.9779
time*income	2	1ImpDA	54.3174	1.0153	1	2862.2	<.0001	0.05	52.3274	56.3073
time*income	2	2DStime	55.0168	0.7001	1	6176.1	<.0001	0.05	53.6447	56.3889
time*income	2	3NotBad	58.5390	0.4943	1	14023	<.0001	0.05	57.5701	59.5079
time*income	2	4Easy	61.9746	0.6015	1	10615	<.0001	0.05	60.7957	63.1536

3.5 MODEL 3: PROC MIXED *unadjusted*

The *model information* describes the model and variables involved and the method used to fit it. *Fixed Effects SE Method – Model Based* describes the method used to compute the approximate standard errors for the Fixed-Effects parameter estimates. Note again 7000 subjects, and 14000 observations ($n = \text{times} * \text{subjects}$), and the absence of missing data is indicated by *number of subjects not used = 0*. The class level information lists the level of the fixed effect variables specified in the model.

Model Information	
Data Set	WORK.PETSTACK
Dependent Variable	pf
Covariance Structure	Variance Components
Subject Effect	IDalias
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment
Class Level Information	
Class	Levels Values
IDalias	7000 not printed
time	2 1 2
pet	4 1NoPet 2DogPet 3CatPet 4OtherPet
Dimensions	
Covariance Parameters	2
Columns in X	11
Columns in Z Per Subject	1
Subjects	7000
Max Obs Per Subject	2
Number of Observations	
Number of Observations Read	14000
Number of Observations Used	14000
Number of Observations Not Used	0

The *iteration history* describes the optimisation of the residual log likelihood and describes the number of iterations that the algorithm takes in order to minimise the objective function.

Iteration History			
Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	131177.34980423	
1	2	125570.33408300	0.00000001

The convergence criterion was met after two iterations. The estimates of the (co)variance components were - *intercept* variance (for IDalias) = 510.54 and the *residual* variance = 177.04 – showing that the between subject variance is larger than the within subject variance in this case.

NOTE: *these two covariance parameter estimates obtained from a subsample of the data can be used as starting points for the model if non-convergence is an issue. See Section 2.5.2 for more details and explanation.*

Convergence criteria met.		
Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Intercept	IDalias	510.54
Residual		177.04

The statistics that SAS uses to assess the goodness of fit of the model are Residual Log Likelihood, Akaike’s Information Criteria (AIC), the corrected form of Akaike’s Information Criteria (AICC) and Schwarz’s Bayesian Criterion (BIC). AIC, AICC and BIC are closely related to the log likelihood statistics with some adjustment for degrees of freedom (i.e. the number of parameters estimated) and they are used when comparing one model with another (a smaller values indicates a better fit).

Fit Statistics	
-2 Res Log Likelihood	125570.3
AIC (smaller is better)	125574.3
AICC (smaller is better)	125574.3
BIC (smaller is better)	125588.0

The solution (i.e., estimates) for the fixed effect parameters are shown in the following box for the fixed effect (pet), with approximate standard errors and corresponding confidence intervals. The Pr>|t| column contains the two-tailed *p*-value for a test of the null hypothesis that the parameter value is zero. Note that the t value and Pr>|t| here are comparable to the output z-value and corresponding probability in the PROC GENMOD unadjusted model (see *Analysis of GEE parameter estimates, empirical standard error estimates, Section 3.3*).

Solution for Fixed Effects										
Effect	time	pet	Estimate	Standard Error	DF	t Value	Pr> t	Alpha	Lower	Upper
Intercept			58.3780	0.9939	6999	58.74	<.0001	0.05	56.4297	60.3262
time	1		3.5098	1.0405	6993	3.37	0.0007	0.05	1.4702	5.5495
time	2		0							
time*pet	1	1NoPet	1.9880	0.9495	6993	2.09	0.0363	0.05	0.1268	3.8492
time*pet	1	2DogPet	1.6727	1.0685	6993	1.57	0.1175	0.05	-0.4220	3.7673
time*pet	1	3CatPet	1.1077	1.2050	6993	0.92	0.3580	0.05	-1.2544	3.4698
time*pet	1	4OtherPet	0							
time*pet	2	1NoPet	0.2967	1.0092	6993	0.29	0.7688	0.05	-1.6817	2.2751
time*pet	2	2DogPet	-0.2264	1.1321	6993	-0.20	0.8415	0.05	-2.4456	1.9929
time*pet	2	3CatPet	0.02252	1.2680	6993	0.02	0.9858	0.05	-2.4631	2.5081
time*pet	2	4OtherPet	0							

The *Type 3 Tests of Fixed Effects* indicate the F tests and p-values refer to tests of null hypotheses of no effects. Thus we see again that time is significant ($p < 0.0001$) but the interaction of time by pet is not ($p = 0.42$).

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
time	1	6993	177.99	<.0001
time*pet	6	6993	1.00	0.4222

Coefficients for pet ownership by time.

Coefficients for time Least Squares				
Effect	time	pet	Row1	Row2
Intercept			1	1
time	1		1	
time	2			1
time*pet	1	1NoPet	0.6357	
time*pet	1	2DogPet	0.2083	
time*pet	1	3CatPet	0.0933	
time*pet	1	4OtherPet	0.0627	
time*pet	2	1NoPet		0.6646
time*pet	2	2DogPet		0.1931
time*pet	2	3CatPet		0.0876
time*pet	2	4OtherPet		0.0547

Coefficients for time*pet Least Squares Means										
Effect	time	pet	Row1	Row2	Row3	Row4	Row5	Row6	Row7	Row8
Intercept			1	1	1	1	1	1	1	1
time	1		1	1	1	1				
time	2						1	1	1	1
time*pet	1	1NoPet	1							
time*pet	1	2DogPet		1						
time*pet	1	3CatPet			1					
time*pet	1	4OtherPet				1				
time*pet	2	1NoPet					1			
time*pet	2	2DogPet						1		
time*pet	2	3CatPet							1	
time*pet	2	4OtherPet								1

Least squares means for physical functioning by pet ownership at each time clearly indicate a bigger difference between the time periods than within the pet categories.

NOTE: these least square mean estimates are the same as the PROC GENMOD least squares means estimates (to two decimal places).

Least Squares Means										
Effect	time	pet	Standard Estimate	Error	DF	t Value	Pr> t	Alpha	Lower	Upper
time	1		63.6033	0.3134	6993	202.94	<.0001	0.05	62.9889	64.2177
time	2		58.5334	0.3134	6993	186.76	<.0001	0.05	57.9190	59.1478
time*pet	1	1NoPet	63.8758	0.3659	6993	174.59	<.0001	0.05	63.1586	64.5930
time*pet	1	2DogPet	63.5605	0.5869	6993	108.30	<.0001	0.05	62.4100	64.7110
time*pet	1	3CatPet	62.9955	0.8192	6993	76.90	<.0001	0.05	61.3896	64.6014
time*pet	1	4OtherPet	61.8878	0.9321	6993	66.39	<.0001	0.05	60.0606	63.7150
time*pet	2	1NoPet	58.6747	0.3603	6993	162.85	<.0001	0.05	57.9684	59.3809
time*pet	2	2DogPet	58.1516	0.6072	6993	95.77	<.0001	0.05	56.9613	59.3419
time*pet	2	3CatPet	58.4005	0.8441	6993	69.18	<.0001	0.05	56.7457	60.0553
time*pet	2	4OtherPet	58.3780	0.9939	6993	58.74	<.0001	0.05	56.4297	60.3262

3.6 MODEL 4: PROC MIXED adjusted

The class level information lists the level of the variables specified in the model, with IDalias as a random effect, and time, pet and ability to manage on income as fixed effects.

Model Information					
Data Set	WORK.PETSTACK				
Dependent Variable	pf				
Covariance Structure	Variance Components				
Subject Effect	IDalias				
Estimation Method	REML				
Residual Variance Method	Profile				
Fixed Effects SE Method	Model-Based				
Degrees of Freedom Method	Containment				
Class Level Information					
Class	Levels	Values			
IDalias	7000	not printed			
time	2	1 2			
pet	4	1NoPet	2DogPet	3CatPet	4OtherPet
income	4	1ImpDA	2DStime	3NotBad	4Easy
Dimensions					
Covariance Parameters	2				
Columns in X	19				
Columns in Z Per Subject	1				
Subjects	7000				
Max Obs Per Subject	2				
Number of Observations					
Number of Observations Read					14000
Number of Observations Used					14000
Number of Observations Not Used					0

The *iteration history* shows that four iterations were needed to reach the convergence criterion.

Iteration History				
Iteration	Evaluations	-2 Res Log Like	Criterion	
0	1	130740.67311213		
1	2	125402.97530480	0.00002000	
2	1	125402.53874102	0.00001152	
3	1	125401.95449849	0.00000001	

The estimates of the covariance components were - *intercept* variance = 492.56 and the *residual* variance = 178.29 – showing that again the between subject variance is larger than the within subject variance in this case.

Convergence criteria met.		
Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Intercept	IDalias	492.56
Residual		178.29

This model (including ability to manage on income) is a better fit, e.g. according to AIC (AIC = 125406), than the model with pet alone (AIC = 125574).

Fit Statistics	
-2 Res Log Likelihood	125402.0
AIC (smaller is better)	125406.0
AICC (smaller is better)	125406.0
BIC (smaller is better)	125419.7

The box below shows the fixed effects estimates with an overall mean (intercept) and effects relative to reference categories of 'other pet' and 'easy to manage on income'.

Solution for Fixed Effects										
Effect	time	pet/income	Estimate	Standard Error	DF	t Value	Pr> t	Alpha	Lower	Upper
Intercept			61.9767	1.0739	6999	57.71	<.0001	0.05	59.8715	64.0818
time	1		3.8254	1.1456	6987	3.34	0.0008	0.05	1.5797	6.0710
time	2		0							
time*pet	1	1NoPet	1.7185	0.9494	6987	1.81	0.0703	0.05	-0.1426	3.5795
time*pet	1	2DogPet	1.8529	1.0668	6987	1.74	0.0825	0.05	-0.2384	3.9442
time*pet	1	3CatPet	1.0222	1.2034	6987	0.85	0.3957	0.05	-1.3369	3.3813
time*pet	1	4OthPet	0							
time*pet	2	1NoPet	0.1240	1.0091	6987	0.12	0.9022	0.05	-1.8541	2.1021
time*pet	2	2DogPet	-0.1327	1.1303	6987	-0.12	0.9066	0.05	-2.3485	2.0831
time*pet	2	3CatPet	0.02898	1.2662	6987	0.02	0.9817	0.05	-2.4532	2.5111
time*pet	2	4OthPet	0							
time*income	1	1ImpDA	-9.9366	1.1094	6987	-8.96	<.0001	0.05	-12.1113	-7.7619
time*income	1	2DStime	-7.0030	0.7105	6987	-9.86	<.0001	0.05	-8.3959	-5.6102
time*income	1	3NotBad	-3.6845	0.5572	6987	-6.61	<.0001	0.05	-4.7768	-2.5922
time*income	1	4Easy	0							
time*income	2	1ImpDA	-7.6748	0.9862	6987	-7.78	<.0001	0.05	-9.6080	-5.7417
time*income	2	2DStime	-6.9720	0.7258	6987	-9.61	<.0001	0.05	-8.3948	-5.5492
time*income	2	3NotBad	-3.4427	0.5520	6987	-6.24	<.0001	0.05	-4.5248	-2.3606
time*income	2	4Easy	0							

The test for fixed effects provides us with a synopsis analogous to an analysis of variance table. It shows statistically significant effects for time and for ability to manage on income by time. We can also see that pet is not statistically significant in the model (similar results to PROC GENMOD).

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
time	1	6987	106.90	<.0001
time*pet	6	6987	0.85	0.5312
time*income	6	6987	27.62	<.0001

Due to the use of the observed margins function the coefficients shown below reflect the distributions of both the pet variable and the ability to manage on income variable. Compare the correct coefficients below, to those produced in the adjusted PROC GENMOD model (0.25 for all levels of pet and income).

Coefficients for time Least Squares Means Using WORK.PETSTACK Margins					
Effect	time	pet	income	Row1	Row2
Intercept				1	1
time	1			1	
time	2				1
time*pet	1	1NoPet		0.6357	
time*pet	1	2DogPet		0.2083	
time*pet	1	3CatPet		0.0933	
time*pet	1	4OtherPet		0.0627	
time*pet	2	1NoPet			0.6646
time*pet	2	2DogPet			0.1931
time*pet	2	3CatPet			0.0876
time*pet	2	4OtherPet			0.0547
time*income	1		1ImpDA	0.0501	
time*income	1		2DStime	0.1949	
time*income	1		3NotBad	0.5183	
time*income	1		4Easy	0.2367	
time*income	2		1ImpDA		0.0674
time*income	2		2DStime		0.1737
time*income	2		3NotBad		0.5153
time*income	2		4Easy		0.2436

Coefficients for time*pet Least Squares Means Using WORK.PETSTACK Margins										
Effect	time	pet/income	Row1	Row2	Row3	Row4	Row5	Row6	Row7	Row8
Intercept			1	1	1	1	1	1	1	1
time	1		1	1	1	1				
time	2						1	1	1	1
time*pet	1	1NoPet	1							
time*pet	1	2DogPet		1						
time*pet	1	3CatPet			1					
time*pet	1	4OtherPet				1				
time*pet	2	1NoPet					1			
time*pet	2	2DogPet						1		
time*pet	2	3CatPet							1	
time*pet	2	4OtherPet								1
time*income	1	1ImpDA	0.0501	0.0501	0.0501	0.0501				
time*income	1	2DStime	0.1949	0.1949	0.1949	0.1949				
time*income	1	3NotBad	0.5183	0.5183	0.5183	0.5183				
time*income	1	4Easy	0.2367	0.2367	0.2367	0.2367				
time*income	2	1ImpDA					0.0674	0.0674	0.0674	0.0674
time*income	2	2DStime					0.1737	0.1737	0.1737	0.1737
time*income	2	3NotBad					0.5153	0.5153	0.5153	0.5153
time*income	2	4Easy					0.2436	0.2436	0.2436	0.2436

Coefficients for time*income Least Squares Means Using WORK.PETSTACK Margins										
Effect	time	pet/income	Row1	Row2	Row3	Row4	Row5	Row6	Row7	Row8
Intercept			1	1	1	1	1	1	1	1
time	1		1	1	1	1				
time	2						1	1	1	1
time*pet	1	1NoPet	0.6357	0.6357	0.6357	0.6357				
time*pet	1	2DogPet	0.2083	0.2083	0.2083	0.2083				
time*pet	1	3CatPet	0.0933	0.0933	0.0933	0.0933				
time*pet	1	4OtherPet	0.0627	0.0627	0.0627	0.0627				
time*pet	2	1NoPet					0.6646	0.6646	0.6646	0.6646
time*pet	2	2DogPet					0.1931	0.1931	0.1931	0.1931
time*pet	2	3CatPet					0.0876	0.0876	0.0876	0.0876
time*pet	2	4OtherPet					0.0547	0.0547	0.0547	0.0547
time*income	1	1ImpDA	1							
time*income	1	2DStime		1						
time*income	1	3NotBad			1					
time*income	1	4Easy				1				
time*income	2	1ImpDA					1			
time*income	2	2DStime						1		
time*income	2	3NotBad							1	
time*income	2	4Easy								1

The box below shows *least squares means* obtained using the **om** function. Note that the LSMEANS are no longer the same as those obtained from PROC GENMOD.

Least Squares Means										
Effect	time	pet/income	Estimate	Standard Error	DF	t Value	Pr> t	Alpha	Lower	Upper
time	1		63.6033	0.3096	6987	205.46	<.0001	0.05	62.9965	64.2102
time	2		58.5334	0.3096	6987	189.08	<.0001	0.05	57.9265	59.1402
time*pet	1	1NoPet	63.7481	0.3625	6987	175.85	<.0001	0.05	63.0374	64.4587
time*pet	1	2DogPet	63.8825	0.5843	6987	109.32	<.0001	0.05	62.7370	65.0279
time*pet	1	3CatPet	63.0517	0.8162	6987	77.25	<.0001	0.05	61.4518	64.6517
time*pet	1	40thPet	62.0296	0.9304	6987	66.67	<.0001	0.05	60.2056	63.8535
time*pet	2	1NoPet	58.5981	0.3568	6987	164.23	<.0001	0.05	57.8986	59.2975
time*pet	2	2DogPet	58.3414	0.6040	6987	96.60	<.0001	0.05	57.1574	59.5254
time*pet	2	3CatPet	58.5030	0.8409	6987	69.57	<.0001	0.05	56.8546	60.1515
time*pet	2	40thPet	58.4740	0.9923	6987	58.93	<.0001	0.05	56.5288	60.4193
time*income	1	1ImpDA	57.4392	1.0066	6987	57.06	<.0001	0.05	55.4659	59.4124
time*income	1	2DStime	60.3728	0.5485	6987	110.08	<.0001	0.05	59.2976	61.4479
time*income	1	3NotBad	63.6913	0.3743	6987	170.16	<.0001	0.05	62.9575	64.4250
time*income	1	4Easy	67.3758	0.5171	6987	130.31	<.0001	0.05	66.3622	68.3894
time*income	2	1ImpDA	54.3612	0.8765	6987	62.02	<.0001	0.05	52.6430	56.0793
time*income	2	2DStime	55.0640	0.5742	6987	95.90	<.0001	0.05	53.9384	56.1895
time*income	2	3NotBad	58.5933	0.3752	6987	156.17	<.0001	0.05	57.8578	59.3288
time*income	2	4Easy	62.0360	0.5106	6987	121.51	<.0001	0.05	61.0352	63.0369



3.7 Summary of least squares means

Table 3.7.1 shows estimates of mean levels of physical functioning (PF) in relation to companion animal ownership at Surveys 2 and 3. At each time the differences between pet categories appear insignificant both before and after adjustment for income; however as expected among older women there is a clear decline in overall physical functioning in the 3 years between survey 2 and 3 (times 1 and 2 in section 3.6). In contrast ability to manage on income was strongly associated with differences in PF ($p < 0.0001$) – results not shown here.

The table also illustrates that the unadjusted means from both models are identical; however the adjusted means are lower from the GENMOD procedure than the MIXED procedure – due to the observed margins function used in the MIXED procedure.

Table 3.7.1 PETS ownership at Survey 2 and Survey 3

	PROC MIXED		PROC GENMOD	
	P=0.4222	p=0.5312	p=0.4089	p=0.5110
	Unadjusted	Adjusted	Unadjusted	Adjusted
Survey 2				
No Pet	63.9(63.2,64.6)	63.7(63.0,64.5)	63.9(63.2,64.5)	62.4(61.6,63.2)
Dog pet	63.6(62.4,64.7)	63.9(62.7,65.0)	63.6(62.5,64.7)	62.5(61.3,63.7)
Cat pet	63.0(61.4,64.6)	63.1(61.5,64.7)	63.0(61.5,64.5)	61.7(60.1,63.2)
Other pet	61.9(60.1,63.7)	62.0(60.2,63.9)	61.9(60.1,63.7)	60.7(58.8,62.5)
Survey 3				
No Pet	58.7(58.0,59.4)	58.6(57.9,59.3)	58.7(57.9,59.4)	57.6(56.7,58.4)
Dog pet	58.2(57.0,59.3)	58.3(57.2,59.5)	58.2(56.9,59.4)	57.3(56.0,58.7)
Cat pet	58.4(56.7,60.1)	58.5(56.9,60.2)	58.4(56.6,60.2)	57.5(56.6,59.4)
Other pet	58.4(56.4,60.3)	58.5(56.5,60.4)	58.4(56.4,60.4)	57.5(55.4,59.5)

Table 3.7.2 presents the results of analyses examining associations between change in companion animal ownership and mean levels of PF across the two surveys. The differences in PF between categories of the transition variable for changes in pet ownership (e.g., moving from not owning a pet at Survey 2 to owning a pet at Survey 3) were statistically significant,

even though they were small compared to the overall decline in physical functioning as a result of the passage of time (i.e., between surveys).

The same differences in mean Physical Functioning for change in pet ownership are apparent in table 3.7.2.

Table 3.7.2 PETS change in ownership (SAS output not shown)

	PROC MIXED		PROC GENMOD	
	P=0.0018	P=0.0058	P=0.0043	P=0.0133
Survey 2	Unadjusted	Adjusted	Unadjusted	Adjusted
Yes → Yes	63.38(62.2,64.5)	53.8(62.6,64.9)	63.4(62.3,64.5)	62.4(61.2,63.6)
No → No	64.3(63.5,65.1)	64.1(63.3,64.9)	64.3(63.6,65.1)	62.8(61.9,63.6)
Yes → No	61.0(58.9,63.2)	61.2(59.0,63.3)	61.0(58.9,63.2)	59.8(57.6,62.0)
No → Yes	60.6(57.9,63.3)	60.6(57.9,63.3)	60.6(58.1,63.1)	59.2(56.7,61.7)
Survey 3				
Yes → Yes	58.7(57.5,59.8)	58.9(57.8,60.1)	58.7(57.4,59.9)	57.9(56.6,59.2)
No → No	59.2(58.4,60.0)	59.1(58.3,59.9)	59.2(58.4,60.1)	58.1(57.1,59.0)
Yes → No	55.6(53.4,57.7)	55.7(53.6,57.9)	55.6(53.1,58.0)	54.7(52.3,57.2)
No → Yes	54.3(51.5,57.0)	54.4(51.8,57.1)	54.3(51.3,57.2)	53.4(50.5,56.4)

PART 4: Plotting results using SAS

A stock plot is a useful way of displaying averages and 95% confidence limits across time for different groups.

There are many alternative options available in SAS - for more options go to SAS online and SAS/GRAPH – SAS/GRAPH Statements or GPLOT procedure and search for additional options.

Data can be manually entered, or calculated from the *output* dataset in both MIXED and GENMOD procedures. Here, means and standard errors are entered manually and used to calculate confidence limits. Values for 'pet' categories are offset from one-another so that the plots of confidence intervals do not overlap. Do not make this offset any greater than necessary.

The results shown below come from the unadjusted MIXED model (Model 3).

```
data work.petsgraph;
  input time pet mean se;
  datalines;
1 1 63.8758 0.3659
1 2 63.5605 0.5869
1 3 62.9955 0.8192
1 4 61.8879 0.9321
2 1 58.6746 0.3603
2 2 58.1516 0.6072
2 3 58.4005 0.8441
2 4 58.3781 0.9938
;run;

data work.expand;
  set work.petsgraph;
  z=probit(0.975);
  if pet=2 then time=time+0.02;
  if pet=3 then time=time+0.04;
  if pet=4 then time=time+0.06;
  yaxis=mean; output;
  yaxis=mean+(z*se); output;
  yaxis=mean-(z*se); output;
run;
```

The options below resets any current settings and specifies the font and height of the text to be used.

```
goptions reset=all ftext=centx htext=2;
```

The four 'symbol' statements call the options for each level of the variable – i.e. no pet, dog pet, cat pet and other pet.

- **i** = option specifies the high/low/mean format for the graph
- **w** = width of the line
- **line** = line type (solid)
- **f** = (font) the marker font associates every letter of the alphabet with a marker (see below)
- **value** = specifies which marker to use;

```
symbol1 color=black i=hilocjt w=4 line=1 value=A f=marker;  
symbol2 color=blue i=hilocjt w=4 line=1 value=U f=marker;  
symbol3 color=red i=hilocjt w=4 line=1 value=W f=marker;  
symbol4 color=green i=hilocjt w=4 line=1 value=V f=marker;
```

The first axis statement describes the horizontal axis (*haxis* below)

- **minor** indicates presence or absence of minor tick marks are along the x-axis
- **offset** indicates the amount of offset from the y-axis – this helps for readability of the graph – the larger the value the greater the offset
- **order** indicates the scale and range for the x-axis
- **major** indicates the height of the major tick marks – the larger the value the bigger the marks;

```
axis1 label=('Time') minor=none offset=(5,0) major=(h=0.1)  
order=(1 to 2.5 by 0.5) value=('Survey 2' ' ' 'Survey 3' ' ');
```

The second axis statement describes the y-axis

- **a** (in label) indicates the angle of the labelling – '0' writes it horizontally at the top of the y-axis, '90' writes it vertically along the y-axis
- **minor** indicates presence or absence of minor tick marks are along the y-axis;

```
axis2 label=(a=90 'Physical Functioning') minor=none;
```

This statement is associated with the legend

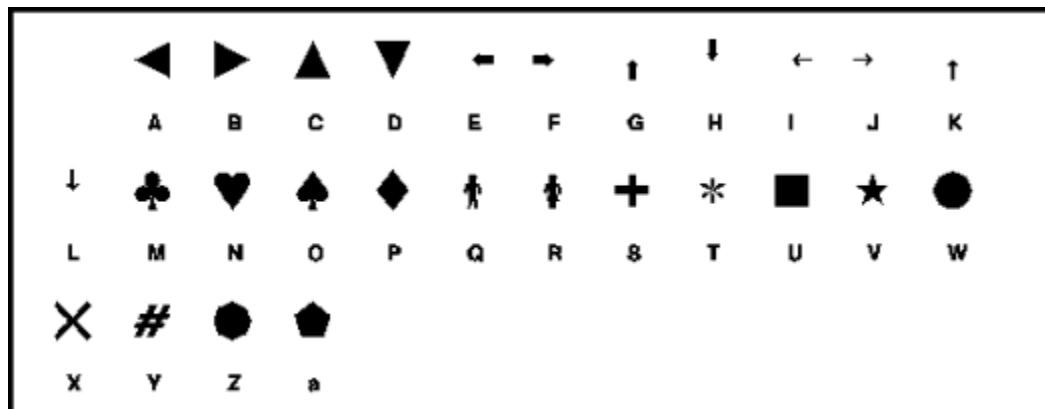
- **position** of the legend indicates bottom/top/middle of the window; left/centre/right; and outside/inside in relation to the axis area
- the **frame** draws a frame around the legend;

```
legend1 position=(bottom center outside) frame label=none  
value=(j=1 'No Pet' 'Dog Pet' 'Cat Pet' 'Other Pet');
```

The graph is created below in the GPLOT procedure. Note there is no frame around the axis area and there are statements specifying which axis is horizontal and which is vertical.

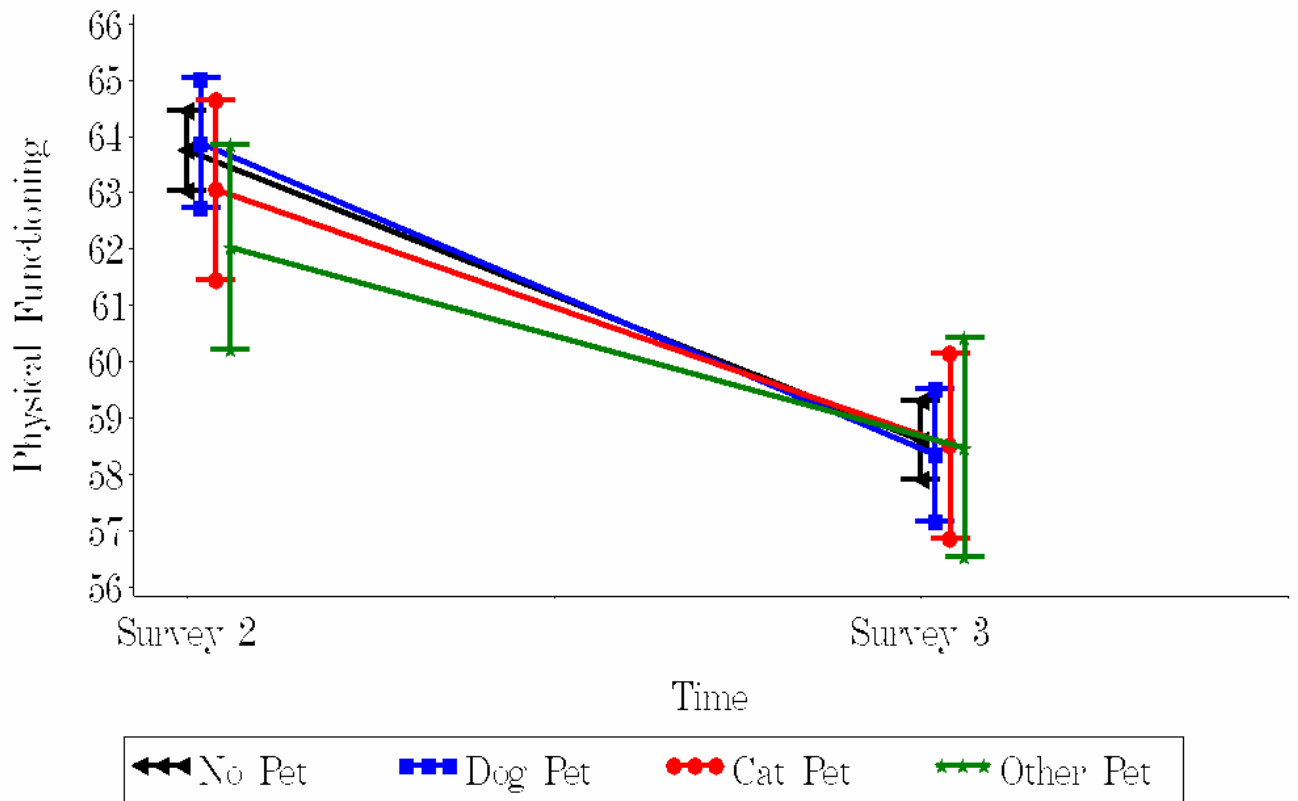
```
proc gplot data=work.expand;  
plot yaxis*time=pet / noframe haxis=axis1  
vaxis=axis2 legend=legend1;  
run; quit;
```

Marker Font



A good online page for colour schemes in SAS is <http://www.devenezia.com/docs/SAS/sas-colors.html>

Note that these do not always turn out as they appear.



The SAS graphs appear most clear in Microsoft Word if they are pasted in as a *Device Independent Bitmap* using *Paste Special*.

PART 5: Sample Size

Notes on sample size calculation for longitudinal studies with continuous response variables.

These notes outline the theory of sample size calculation for comparing two groups at a single time-point and then give extensions to the case of longitudinal data.

Calculating an appropriate sample size for a study requires background information on the study factors such as the magnitude and variability of the more important response variables and the size of the effects that would be biologically or socially significant.⁴

For standard situations, for example with cross-sectional data, there is good software available; some of it free, such as PS which can be downloaded from <http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize>

The main purpose of this note is to describe sample size calculations for longitudinal studies with continuous response variables.

5.1 Single Time Point Studies

The following sample size formula is applicable for a study comparing two groups, each with N subjects⁴. The null hypothesis states that there is no difference in the mean response variable between the groups, versus an alternative hypothesis that there is a difference. We want to test this two-tailed hypothesis at the level of significance α and with power β .

$$N = \frac{2\sigma^2(Z_{(1-\alpha/2)} + Z_{(1-\beta)})^2}{\nu^2} \dots\dots\dots[\text{Equation 1}]$$

where:

- N is the sample size per group
- σ is the standard deviation of the response variable, and is assumed to be equal in the two study groups (an estimate of σ , s may be substituted)
- $Z_{(1-\alpha/2)}$ is the $(1 - \alpha/2)$ percentile point of the standard Normal distribution.

- $Z_{(1-\beta)}$ is the $(1 - \beta)$ percentile point of the standard Normal distribution.
- ν is the size of the effect that we wish to have a good chance of detection and would be biologically/socially important; it is calculated as the difference between the estimated means of the two groups being studied.

Percentile points from the standard Normal distribution that are most often used in sample size calculations are shown in Table 5.1.

Table 5.1: Percentiles of Standard Normal Distributions for varying α , β .

$\alpha/2$	$Z_{(1-\alpha/2)}$	Power = $1 - \beta$	$Z_{(1-\beta)}$
0.05	1.645	0.8	0.84
0.025	1.960	0.9	1.28
0.005	2.575	0.95	1.645

If you need to design a study with unequal group sizes, for example because one group is easier to access than the other, then the following formula is relevant.

For unequal groups of size N_1 and N_2 , where $r = N_1/N_2$, the sample size formula is given by⁵:

$$N_1 = \frac{(\sigma_1^2 + \sigma_2^2 / r)(Z_{1-\alpha/2} + Z_{(1-\beta)})^2}{\nu^2} \text{ and } r = N_1/N_2 \quad \dots\dots\dots[\text{Equation 2}]$$

where:

- σ_1 is the standard deviation in group 1 and σ_2 the standard deviation in group 2 (the estimates s_1 and s_2 may be substituted for σ_1 and σ_2 respectively).

5.2 Extension to Longitudinal Studies

The sample size calculations for longitudinal studies are very similar to those from “standard” experimental studies which assume at most one response measurement. When repeated response measurements are made, and the aim of the research is to track the development of a response over time, then standard sample size formulae must be revised to incorporate the correlation between the repeated measures.

When analysing data from longitudinal studies, there are two types of comparisons we may want to make:

- 1) Comparing the rate of change of a response variable over time, between two different groups of subjects;
- 2) Comparing the average response for two groups.

It is important to differentiate between the two comparisons, as the formulae for calculating sample size differ.

The model for calculating sample size assumes that each person is measured at the same fixed times (for all subjects in all groups). Repeated measures from a subject are typically correlated. For simplicity in calculating sample size, assume the correlation between two variables remains the same across all time points.

If Y_{ijk} represents the response for the i -th subject at time j , in group k , then assume $Var(Y_{ijk}) = \sigma^2$ and $corr(Y_{ijk}, Y_{ilk}) = \rho$ for all $j \neq l$. i.e., ρ is the correlation coefficient between the repeated measures for the outcome variable of interest. The values for σ^2 and ρ must be estimated from previous data, or assumed.

1) Comparing the rate of change over time – between different groups of subjects

Rate of change over time can be estimated as the slope of a straight line drawn through the two or more time points at which the response is measured. So, differences in rate of change can be tested as differences in slopes.

Assuming an equal number of subjects in each group, then the sample size formula is given as^(5, 6):

$$N = \frac{2\sigma^2(1-\rho)(Z_{(1-\alpha/2)} + Z_{(1-\beta)})^2}{ms_t^2\nu^2} \dots\dots\dots[\text{Equation 3}]$$

where:

- N is the required sample size per group to be compared
- σ is the standard deviation of the response variable
- ρ is the correlation coefficient of the repeated measurements
- $Z_{(1-\alpha/2)}$ is the $(1 - \alpha/2)$ percentile point of the standard Normal distribution.
- $Z_{(1-\beta)}$ is the $(1 - \beta)$ percentile point of the standard Normal distribution.
- m is the number of response measurements for each subject.
- $s_t^2 = \sum_{j=1}^m (t_j - \bar{t})^2 / m$, with t_j being the time of measurement of the j -th repeated measurement $j = 1, \dots, m$.
- ν is the difference in mean value of the response variable between the groups

For example, if a series of measurements were undertaken in 1996, 2000, 2003, then $t_1 = 0$ (baseline), $t_2 = 4$ (years), $t_3 = 7$ (years) and $m = 3$.

Note that in this case, the sample size decreases as ρ increases. Also, the calculation assumes linear changes over time, and the formula does not allow for missing data, either due to dropouts or to missing responses at any particular time points.

2) Comparing the average response for 2 groups

If the outcome variable is continuous, then the equation for sample size when comparing means from two groups is as follows^(5,7,8,9):

$$N = \frac{2\sigma^2[1 + (T - 1)\rho](Z_{(1-\alpha/2)} + Z_{(1-\beta)})^2}{v^2T} \dots\dots\dots[\text{Equation 4}]$$

where:

- N is the sample size per group to be compared.
- σ is the standard deviation of the response variable.
- T is the number of **follow-up** measurements
- ρ is the correlation coefficient of successive measurements
- $Z_{(1-\alpha/2)}$ is the $(1 - \alpha/2)$ percentile point of the standard Normal distribution.
- $Z_{(1-\beta)}$ is the $(1 - \beta)$ percentile point of the standard Normal distribution.
- v is the difference in mean value of the response variable between the groups.

For example, if we are comparing changes over 3 surveys (S1, S2, S3), then $T = 2$ ($T = m - 1$). Note that in this case, the sample size increases as ρ increases.

This formula can be implemented in a spreadsheet. An example is given in Figure 5.1.

If we have groups of unequal size N_1 and N_2 , where $r = N_1/N_2$, the above formula is adjusted as follows:

$$N_1 = \frac{(r + 1)\sigma^2[1 + (m - 1)\rho](Z_{(1-\alpha/2)} + Z_{(1-\beta)})^2}{v^2rm} \dots\dots\dots[\text{Equation 5}]$$

A larger sample size will be required as:

- the standard deviation of the outcome increases
- the power of the test increases
- the significance level of the test decreases
- the size of difference we wish to detect decreases

The sample size required will decrease with more follow-up measurements

Example: Comparing differences in average SF-36 subscale scores over time in relation to alcohol consumption among older women.

The 8 SF-36 dimension scores (or subscales) of interest are bodily pain (BP), general health (GH), mental health (MH), physical functioning (PF), role emotional (RE), role physical (RP), social functioning (SF), vitality (VT). Each of these subscales has values between 0 and 100, with higher scores representing better health. In a longitudinal study we may seek to determine whether or not there are significant differences in the average subscale scores for women consuming varying amounts of alcohol.

For each scale, the standard deviation of scores was calculated at each survey (see Table 5.2)

Table 5.2: Standard deviations of SF-36 subscales for Older cohort

Scale	Survey 1	Survey 2	Survey 3	Average Stdev.
Bodily Pain (BP)	26.87	27.17	26.78	27
General Health (GH)	22.09	21.21	20.55	21
Physical Functioning (PF)	26.36	25.94	28.16	26
Role Physical (RP)	42.52	41.68	41.08	42
Mental Health (MH)	17.04	16.20	15.81	16
Role emotional (RE)	37.37	34.93	35.74	36
Social Functioning (SF)	25.59	25.16	27.04	26
Vitality (VT)	20.99	21.05	20.63	21

The sample size formula requires a single estimate for standard deviation of the response, not individual estimates at each time point. As the standard deviations vary little across time, using either the maximum standard deviation, or an average of the 3 standard deviations, should be adequate for sample size calculations.

The final column in Table 5.2 above gives the values of standard deviation used (averages) in the final calculation (and in the XL spreadsheet). Note the values are rounded to the nearest integer for simplicity in entering the data into the spreadsheet.

We also need to calculate the correlation between repeated measurements from the survey data. Some example correlation matrices are given below, in Table 5.3, for the BP and RE subscales.

Table 5.3 Correlation matrices for BP and RE subscales between three times.

	O1BP	O2BP	O3BP		O1RE	O2RE	O3RE
O1BP	1	0.57	0.54	O1RE	1	0.36	0.34
O2BP		1	0.58	O2RE		1	0.37
O3BP			1	O3RE			1

As for the standard deviations, the correlation values varied little across time for these two subscales. This is also true for the remaining 6 subscales but for simplicity, the remaining correlation matrices are not shown in this document.

The spreadsheet has options to change the values of α, β, T and ρ , and to vary the magnitude of the detectable difference (ν).

Assuming we want to test for a difference in the average response between groups, not the rate of change of the response over time, we need to apply Equation 4 above.

For the bodily pain (BP) scale, in order to detect an average difference of 2 units with 80% power at the 5% significance level, the calculation for sample size would be as follows:

$$N = \frac{(1.96 + 0.84)^2 \times 27^2 \times (1 + 1) \times [1 + (2 - 1) \times 0.57]}{2^2 \times 1 \times 2}$$

$$= 2246$$

If the difference to be detected increases to 3, a sample size of 998 per group is required, and for a difference of 5 units, the sample size required reduces to $N = 359$ per group at each time point.

*** Hypothesis: Comparing the average response for 2 groups

		Response variables							
		BP	GH	MH	PF	RE	RP	SF	VT
stdev of reponse	σ	27	21	16	26	36	42	26	21
Correlation	ρ	0.57	0.67	0.58	0.72	0.35	0.46	0.43	0.63
Diff detected	v	2	2	2	2	2	2	2	2
	N	2246	1445	794	2282	3433	5054	1897	1411
Diff detected	v	3	3	3	3	3	3	3	3
	N	998	642	353	1014	1526	2246	843	627
Diff detected	v	4	4	4	4	4	4	4	4
	N	561	361	198	570	858	1263	474	353
Diff detected	v	5	5	5	5	5	5	5	5
	N	359	231	127	365	549	809	303	226

Figure 5.1: Example of spreadsheet for calculating sample size when comparing the averages of 2 groups.

PART 6: References

1. SAS. (1999). SAS/STAT Users Guide. SAS Inst. Cary, NC.
2. <http://www.umanitoba.ca/centres/mchp/concept/dict/Statistics/random_effects/random_effects.html> ©2004 Manitoba Centre for Health Policy, accessed 16May 2005.
3. Ware, JE., Sherbourne, CD. (1992). The MOS 36-item Short-Form health survey (SF36): I Conceptual framework and item selection. *Med Care* 30: 473-483.
4. Dobson, AJ. (1984). Calculating Sample Size. Transactions of the Menzies Foundation 1984 Volume 7 pp75-79
5. Catellier, D. (2003). Office of Behavioural and Social Sciences Research: Summer Institute on Design and conduct of Randomized Clinical trials Involving Behavioural Interventions.
6. Diggle, PJ., Heagerty, P., Liang, K-Y., Zeger, S.L. (2002). Analysis of Longitudinal Data. Oxford University Press
7. Twisk, JWR. (2003). Applied Longitudinal Data Analysis for Epidemiology. Cambridge Press.
8. Schumm, P. (2004). Sample Size Calculation for Longitudinal Studies. North American Stats Users' Group Meeting 2004.
9. Hedeker, D., Gibbons, R.D. (1999). Sample size Estimation for Longitudinal Designs with Attrition: Comparing Time-related Contrasts Between Two Groups. *Journal of Educational and Behavioural Statistics* 24:70-93

The following books and papers provide further information on longitudinal analysis.

Barnett AG., van der Pols JC., Dobson AJ. Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology* 2005; 34: 215-220.

Dobson AJ. *An introduction to Generalized Linear Models* (second edition). Chapman & Hall/CRC, 2001 London and Boca Raton.

Diggle, PJ., Heagerty, P., Liang, KY., Zeger, SL. *Analysis of Longitudinal Data* (second edition). Oxford University Press, 2002.

Littell, RC., Henry, PR., Ammerman, CB. (1998). Statistical analysis of repeated measures data using SAS procedures. *Journal of Animal Science* 1998; 76: 1216-1231.

