# Practical biostatistics

Department of Clinical Epidemiology, Biostatistics and Bioinformatics

AMC

Repeated measurements

# Day 10 Repeated measurements

ASSOCIATIONS

CROSSSECTIONAL | LONGITUDINAL

BIVARIATE | MULTIVARIATE

| | | |
|---|---|---|
| Pearson's Rho | linear regression | Repeated measures: GLM / mixed |
| Spearman's Rho | (non-linear regression) | |
| Odds-ratio | logistic regression | survival analysis: Kaplan-Meier / proportional hazards |
| | (logistic reg. / discriminant) | |

# Day 10 - Contents

- What are repeated measurements?
- Incorrect analysis of repeated measurements
- Statistically simple solutions
- Linear mixed effects models

# Measurements

- In general measurements are carried out on **independently** sampled units
  - patients
  - healthy volunteers
  - rats, bats or cats
  - petri-dishes of cells
  - sampling unit

# Single or multiple measurements?

- Statistical analysis is often easier if only one measurement is made on each unit

- However, there are often very good clinical or contextual reasons for measuring each unit more than once

# What are repeated measurements?

- Repeated measurements occur when each unit is measured more than once, even if the unit is examined under different conditions
  - increase accuracy
  - increase efficiency
  - compare different conditions
  - examine changes over time

# Example

- Randomised double-blind controlled trial of N-Acetylcysteine in patients with amyotrophic lateral sclerosis (ALS)
  - 110 patients
  - assessed at monthly intervals for 12 months
  - decline in pulmonary function

# Contents

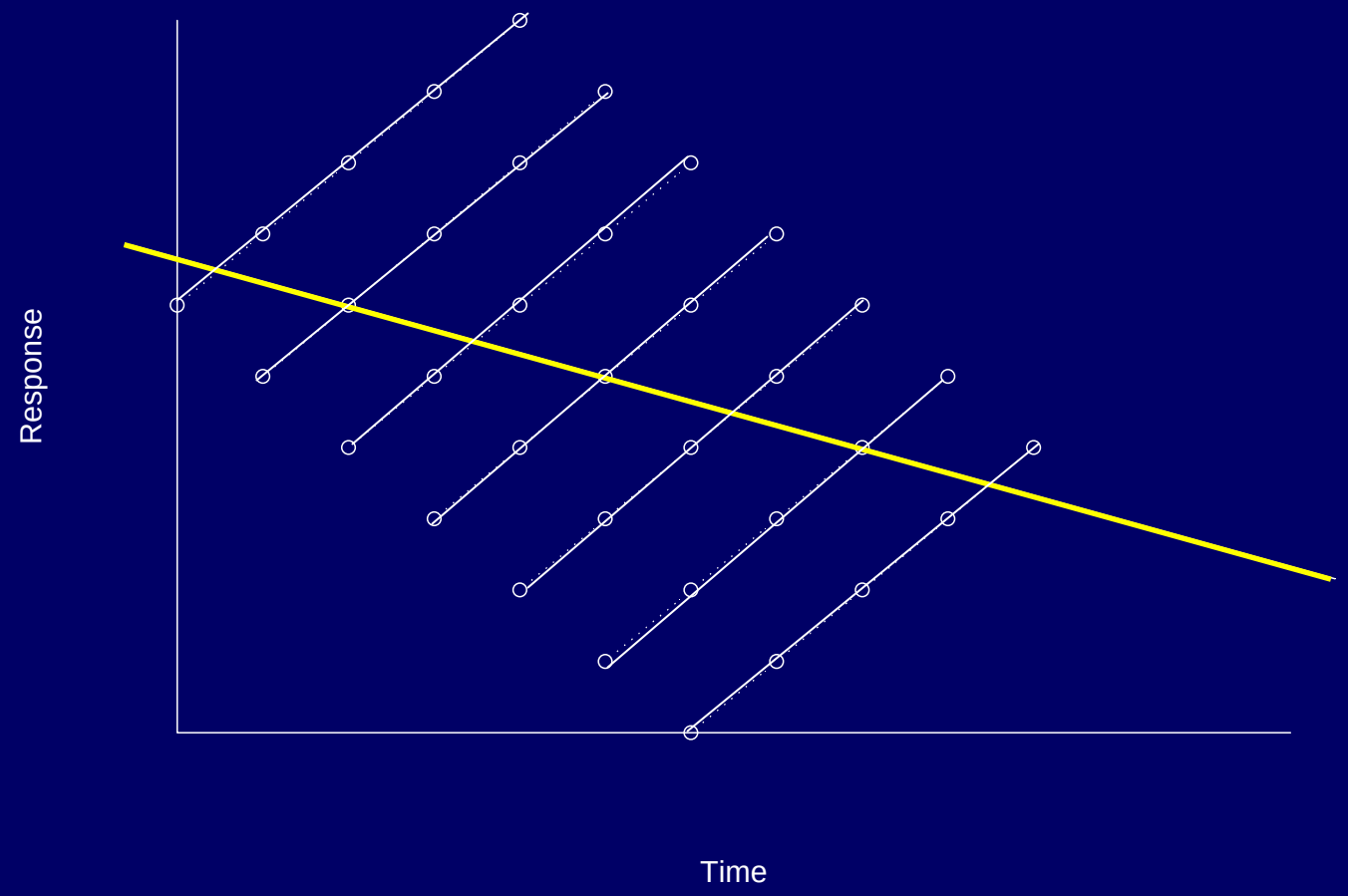
- What are repeated measurements?
- Incorrect analysis of repeated measurements
- Statistically simple solutions
- Linear mixed effects models

# Incorrect analysis of repeated measurements

- It is incorrect to treat the measurements as if each was made on a separate unit
  - ALS data: 110 patients each measured 12 times resulting in 1320 ‘measurements’

- Results, which are not really significant, can appear so

# Incorrect analysis of repeated measurements



# Incorrect analysis of repeated measurements

- Multiple *t*-tests at each time point
- 1 *t*-test
  - P(null hypothesis rejected if actually true) = 0.05
- 5 *t*-tests
  - P(...) $= 1 - (1 - 0.05)^5 = 0.23$
- 10 *t*-tests
  - P(...) $= 1 - (1 - 0.05)^{10} = 0.40$

# Contents

- What are repeated measurements?
- Incorrect analysis of repeated measurements
- Statistically simple solutions
- Linear mixed effects models

# Statistically simple solutions

- There are a number of statistically simple solutions to the problem of analysing repeated measurements
  - easily implemented in SPSS
  - novel uses of techniques discussed in introductory courses

# Using one measurement

- Define one particular measurement as ‘most important’ or ‘clinically most interesting’
  - one ‘number’ per unit
  - analysis very straightforward (often *t*-test)
  - a lot of information is lost

  - e.g. ALS data at 12 months (endpoint)

# Difference between two measurements

- Examine the within-unit difference between two measurements
  - ‘corrected’ for, say, baseline
  - analysis straightforward (paired *t*-test)
  - a lot of information is lost
  - e.g. in the ALS data, the difference between baseline and 12 months

# Mean of each patient

- If multiple measurements have been carried out in identical circumstances, then the mean of each patient can be used in standard techniques
  - increases accuracy
  - loss of information
  - multiple blood tests under identical circumstances

# Individual regression model

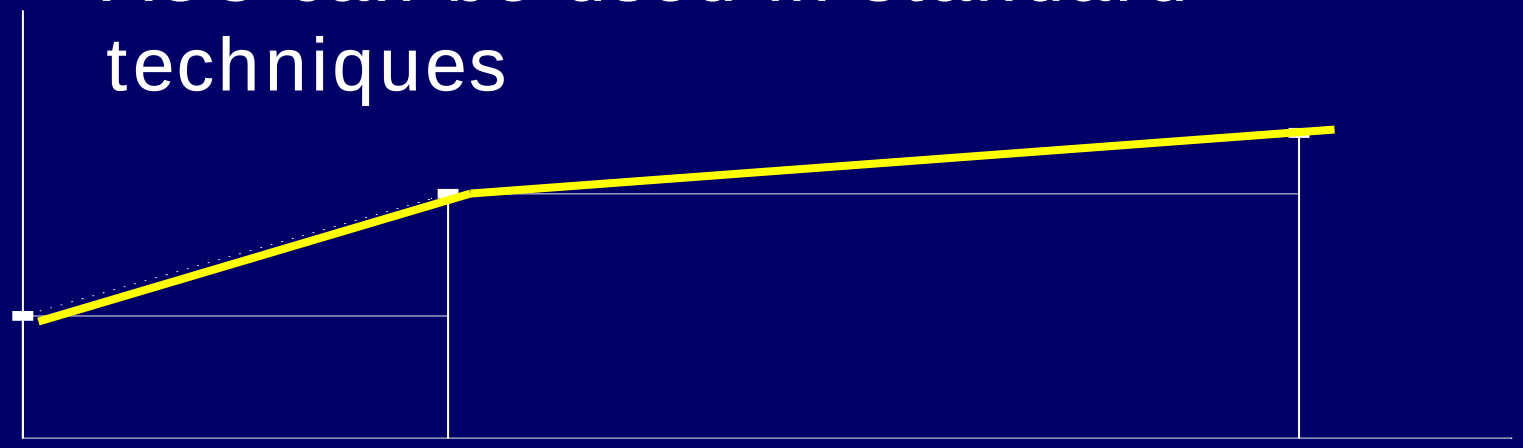
- If interest is in individuals, then a regression (or other) model can be built for each patient
  - useful for data exploration
  - as many models as patients
  - difficult to compare different models

# Time to a particular level

- In a longitudinal study, it can be useful to examine the time taken for a patient to fall below or rise above a certain level
  - survival, or Cox regression, analysis
  - e.g. time taken for ALS patients to experience a 50% reduction in lung function

# Area Under the Curve (AUC)

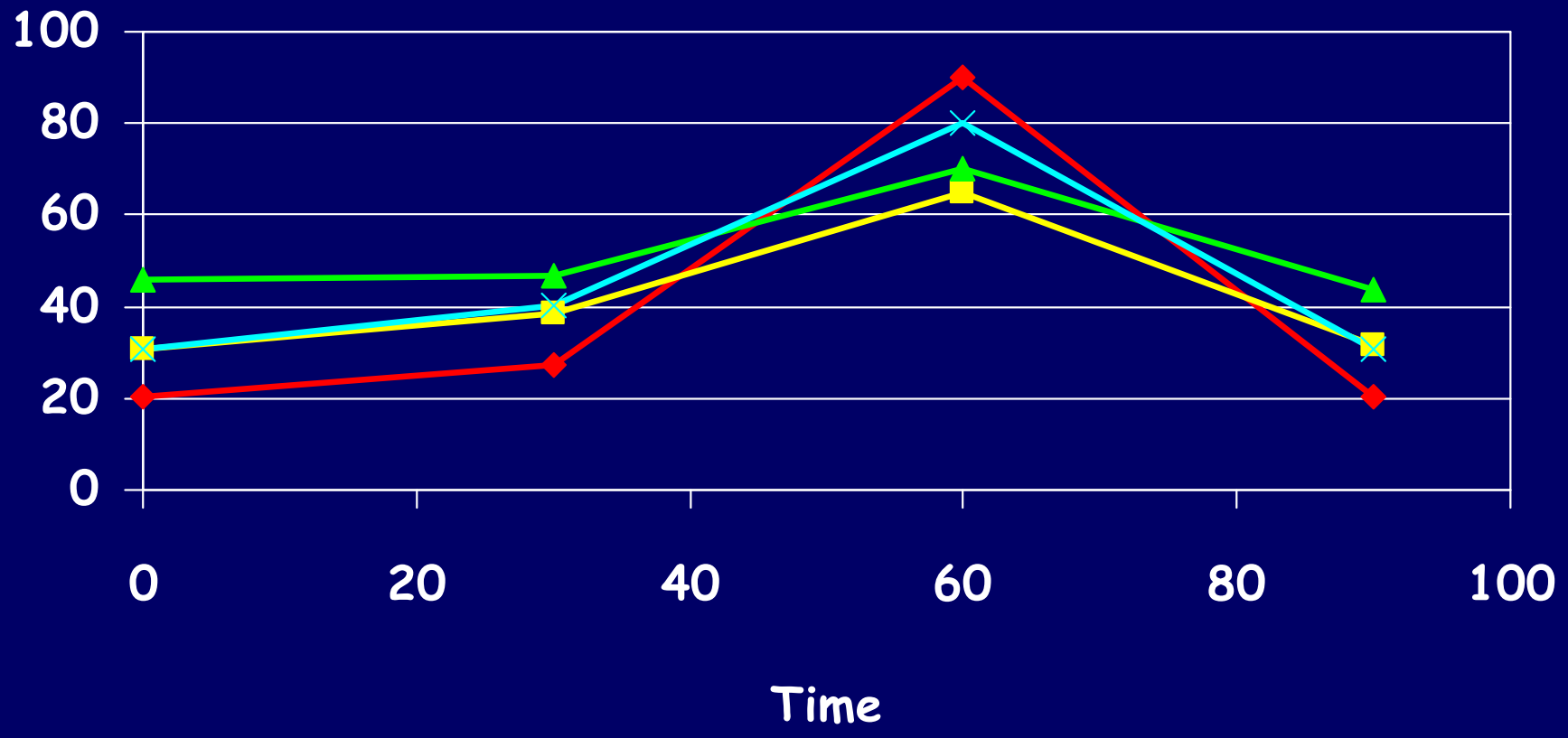
- In a longitudinal study the area under the curve can be used as an indication of the ‘total’ in a given period
  - AUC can be used in standard techniques



# Contents

- What are repeated measurements?
- Incorrect analysis of repeated measurements
- Statistically simple solutions
- Linear mixed effects models

# Example

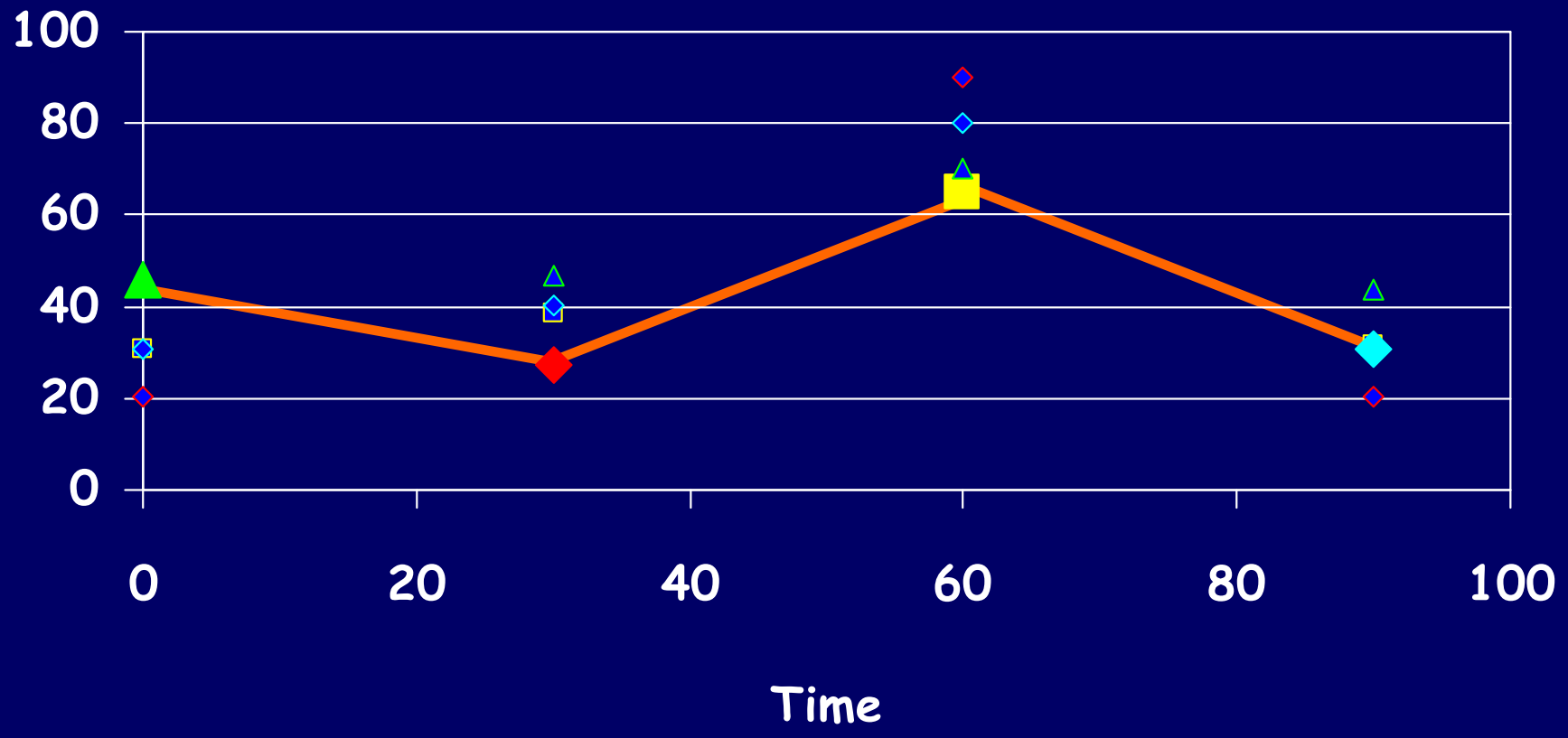


# Linear regression - single measurement

- Single measurement of *Y* and single measurement of *X* on each of *n* patients
  - Y lung function
  - X time point in ALS trial

$$y_i = a + bx_i + e_i$$

# Linear regression - single measurement

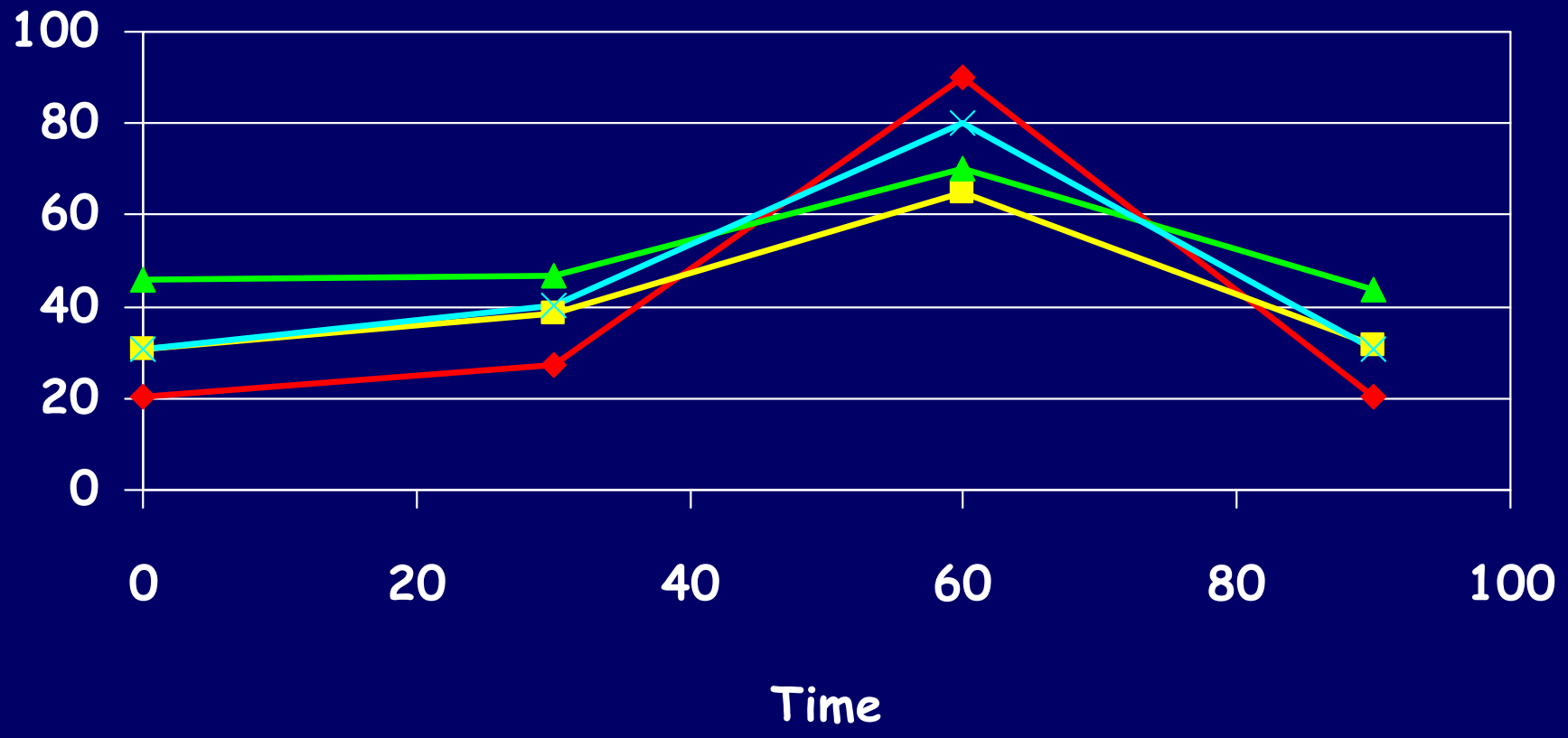


# Linear regression - single patient

- *J* measurements of *Y* and of *X* on a single patient
  - *Y* lung function
  - *X* time point in ALS trial

$$y_{1j} = a_1 + b_1 x_j + e_{1j}$$

# Example



# Problem and solution?

- We need to find a way of linking these two situations together

- Linear mixed-effects models

# Linear mixed-effects models

- **Step 1: structure within a person**
  - Variance – covariance structure

- **Step 2: structure between measurements**
  - Applying variance – covariance structure from step 1

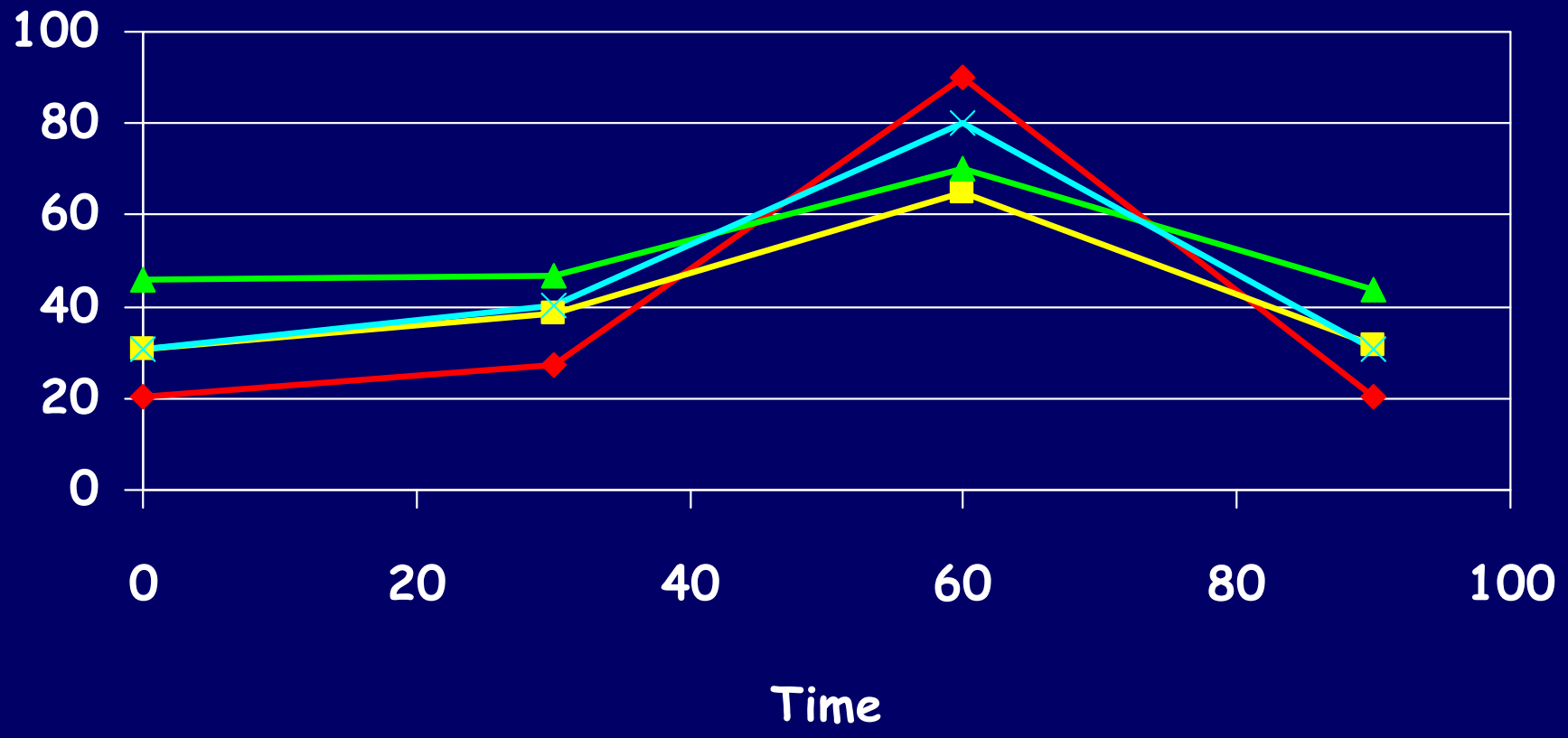
# Linear mixed-effects models

- **Step 1: structure within a person**
  - Variance – covariance structure

- **4 measurements per patient**

$$\frac{\sum_{i=n}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$$

# Example



# Covariance matrix between 4 time points

$$
\begin{array}{cc}
 & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\
R = & \begin{pmatrix} \sigma_1^2 & \lambda_{12} & \lambda_{13} & \lambda_{14} \\ \lambda_{12} & \sigma_2^2 & \lambda_{23} & \lambda_{24} \\ \lambda_{13} & \lambda_{23} & \sigma_3^2 & \lambda_{34} \\ \lambda_{14} & \lambda_{24} & \lambda_{34} & \sigma_4^2 \end{pmatrix} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array}
\end{array}
$$

# Estimation of parameters

- When using mixed effects models the parameters are estimated using maximum likelihood methods

- The LRTS and AIC are important indicators of model fit

# Fit statistics

- Likelihood ratio test statistic (LRTS)
  - To examine two hierarchical models
  - Produces a *p*-value
- Akaike's Information Criterion (AIC)
  - To examine two models
  - Does NOT produce a *p*-value
  - In SPSS 'smaller is better'

# Step 1: Unstructured

Model Dimension

| | | Number of Levels | Covariance Structure | Number Parameters |
|---|---|---|---|---|
| Fixed Effects | Intercept | 1 | | 1 |
| | time | 11 | | 10 |
| | code | 2 | | 1 |
| Repeated Effects | time | 11 | Unstructured | 66 |
| Total | | 25 | | 78 |

*Dependent Variable: FVC.*

Information Criteria

| | |
|---|---|
| -2 Restricted Log Likelihood | 1163 |
| Akaike's Information Criterion (AIC) | 1295 |

# Step 1: Compound symmetry

Model Dimension

| | | | Number of Levels | Covariance Structure | Number Parameters |
|---|---|---|---|---|---|
| Fixed Effects | Intercept | | 1 | | 1 |
| | time | | 11 | | 10 |
| | code | | 2 | | 1 |
| Repeated Effects | | time | 11 | Compound symmetry | 2 |
| Total | | | 25 | | 14 |

Dependent Variable: FVC.

Information Criteria

| | |
|---|---|
| -2 Restricted Log Likelihood | 1786 *(Un: 1163)* |
| Akaike's Information Criterion (AIC) | 1790 *(Un: 1295)* |

# Step 1: AR(1)

Model Dimension

| | | Number of Levels | Covariance Structure | Number Parameters |
|---|---|---|---|---|
| Fixed Effects | Intercept | 1 | | 1 |
| | time | 11 | | 10 |
| | code | 2 | | 1 |
| Repeated Effects | time | 11 | First order autoregresive | 2 |
| Total | | 25 | | 14 |

Dependent Variable: FVC.

Information Criteria

| | |
|---|---|
| -2 Restricted Log Likelihood | 1329 *(Un: 1163)* |
| Akaike's Information Criterion (AIC) | 1333 *(Un: 1295)* |

# Linear mixed-effects models

- Step 1: structure within a person
  - Variance – covariance structure
- Step 2: structure between measurements
  - Applying variance – covariance structure from step 1
  - “mixed effects models”

# Step 2

Estimates of Fixed Effects(b)

| Parameter | Estimate | Std. Error | df | t | Sig. |
|---|---|---|---|---|---|
| Intercept | 3.11 | .18 | 122 | 17.2 | .000 |
| [code=0] | -.12 | .24 | 106 | -.49 | .627 |
| [code=1] | 0(a) | 0 | . | . | . |
| time | -.10 | .01 | 145 | -14.1 | .000 |

a This parameter is set to zero because it is redundant.

b Dependent Variable: FVC.

# What are mixed effects models?

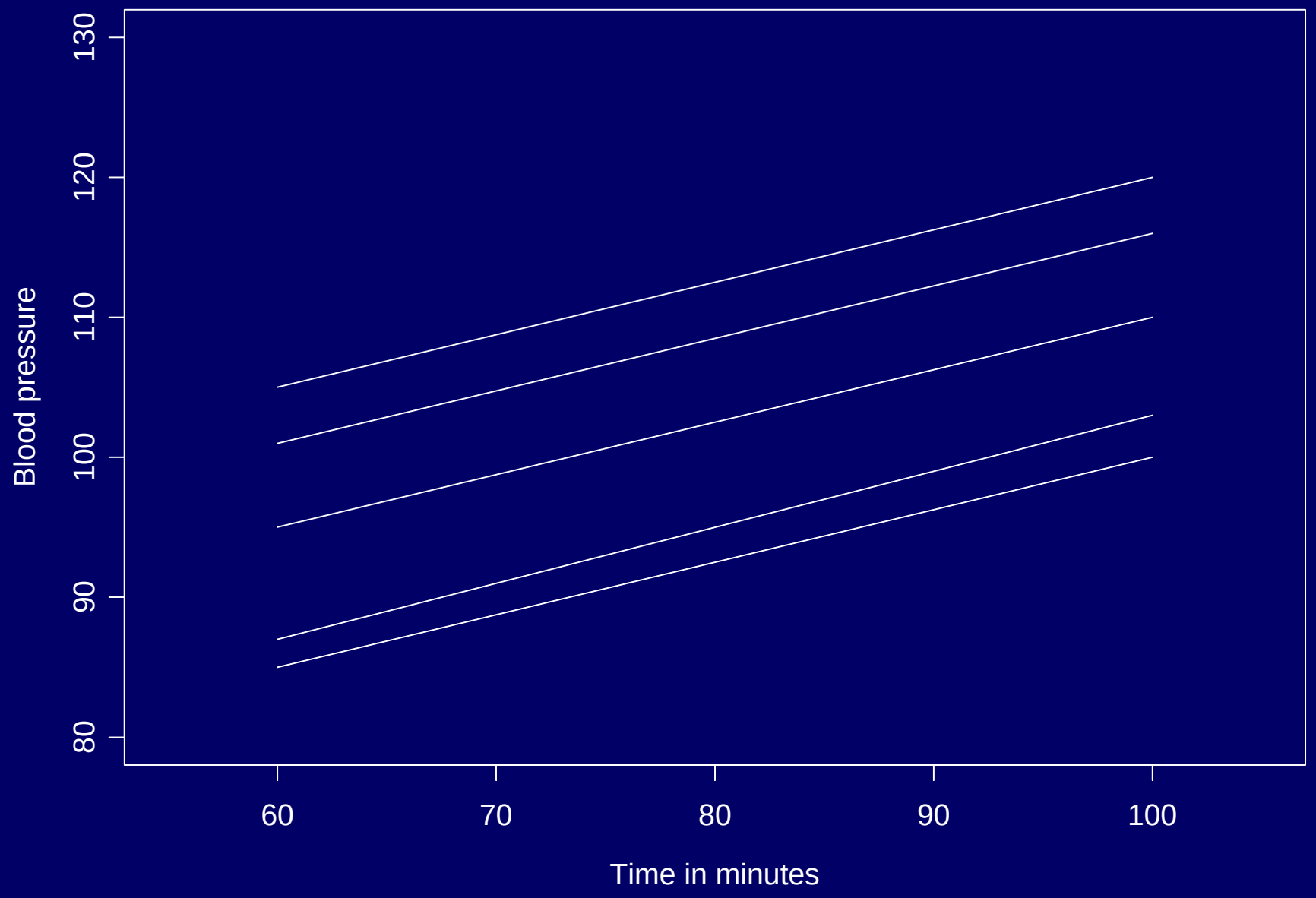
- Extension of regression techniques
- Include both ‘random’ and ‘fixed’ effects
  - Fixed effects are ‘interesting’
  - Random effects are an essential part of the model

# Random intercept

- Assume that the ‘slope’ is the same for all patients

$$y_{ij} = a_i + bx_{ij} + e_{ij}$$

Figure 5: random intercept

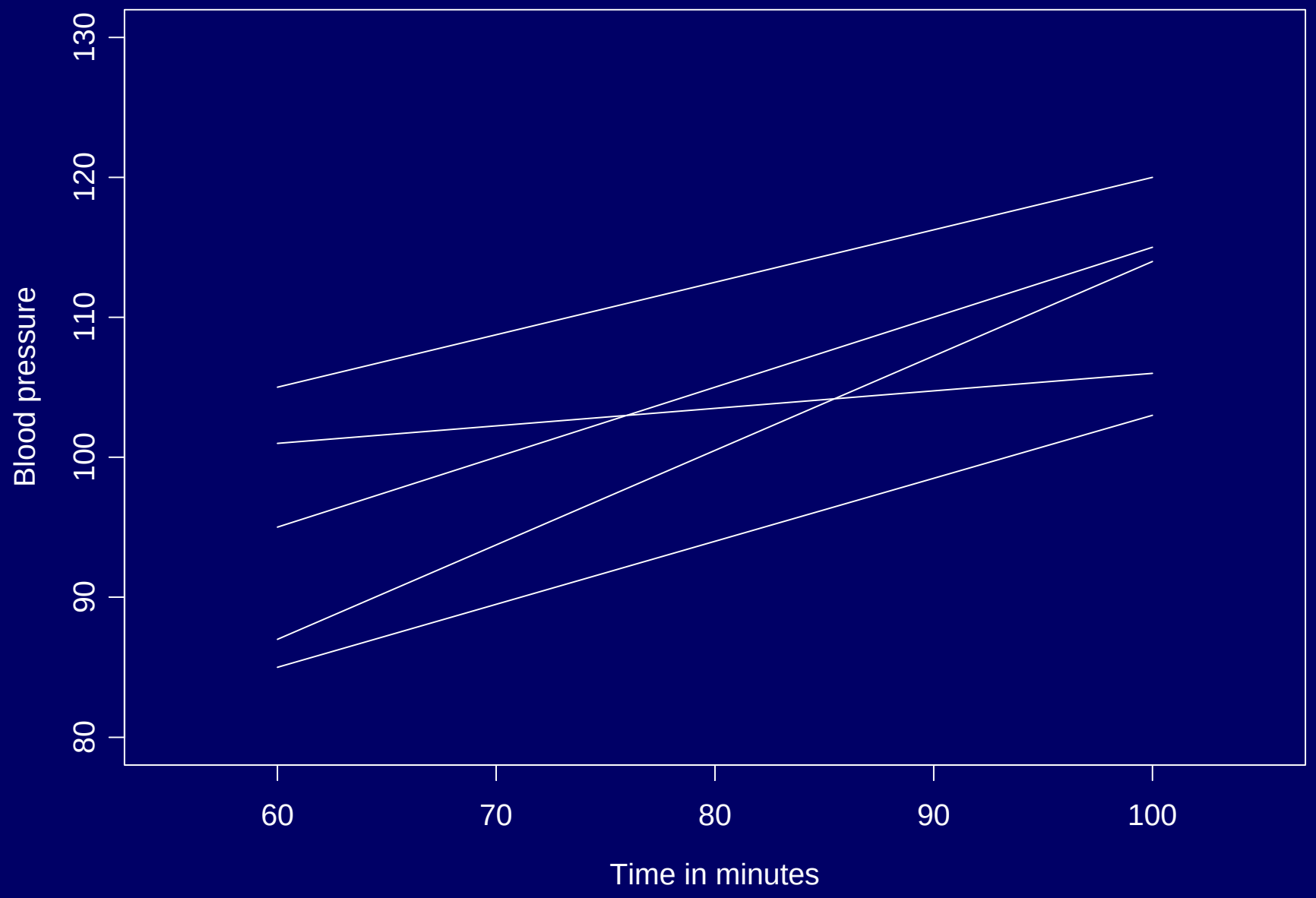


# Random intercept and slope

- Assume that neither the ‘slope’ nor the intercept are the same for all patients

$$y_{ij} = a_i + b_i x_{ij} + e_{ij}$$

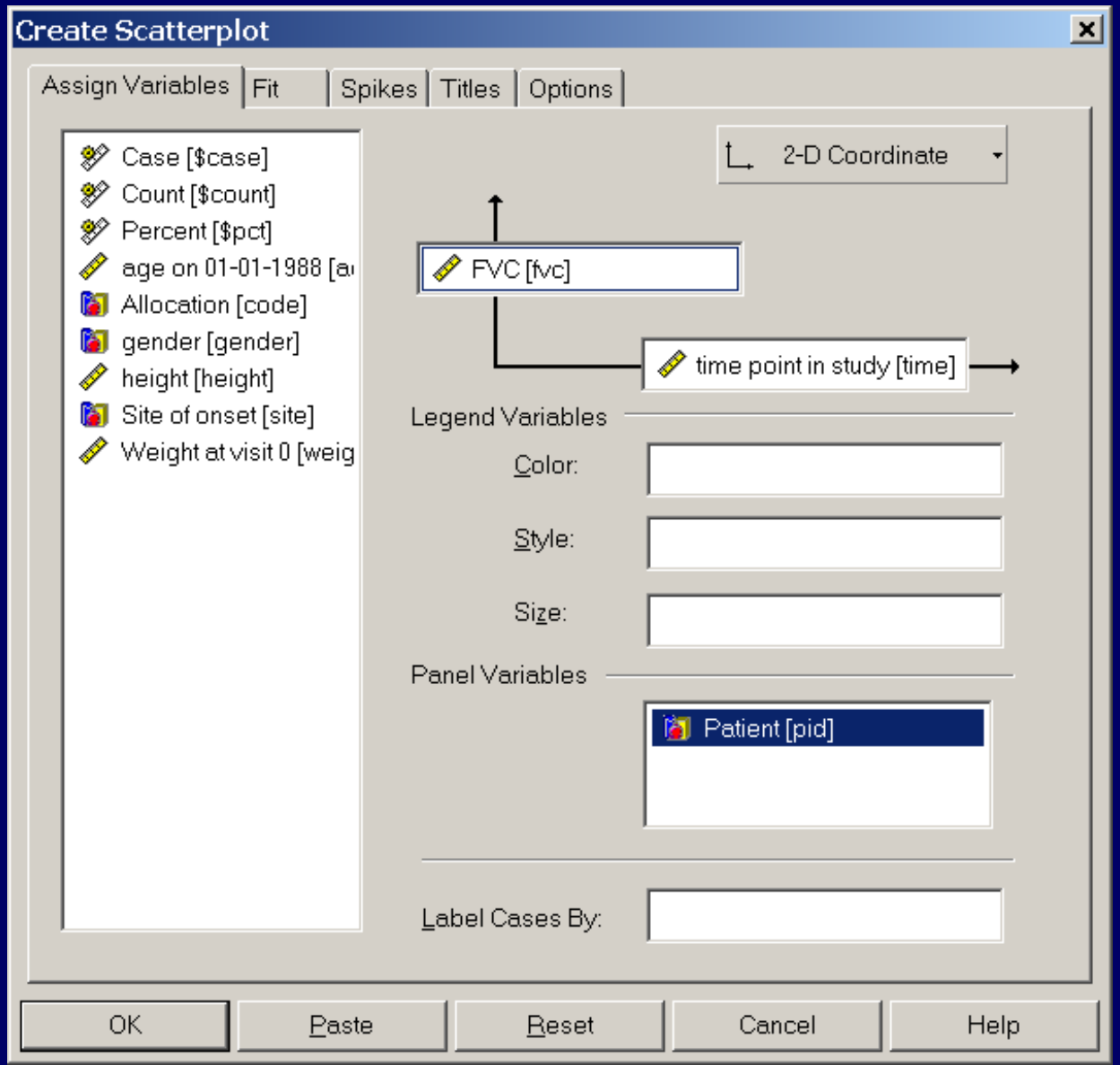
Figure 6: random intercept and random slope



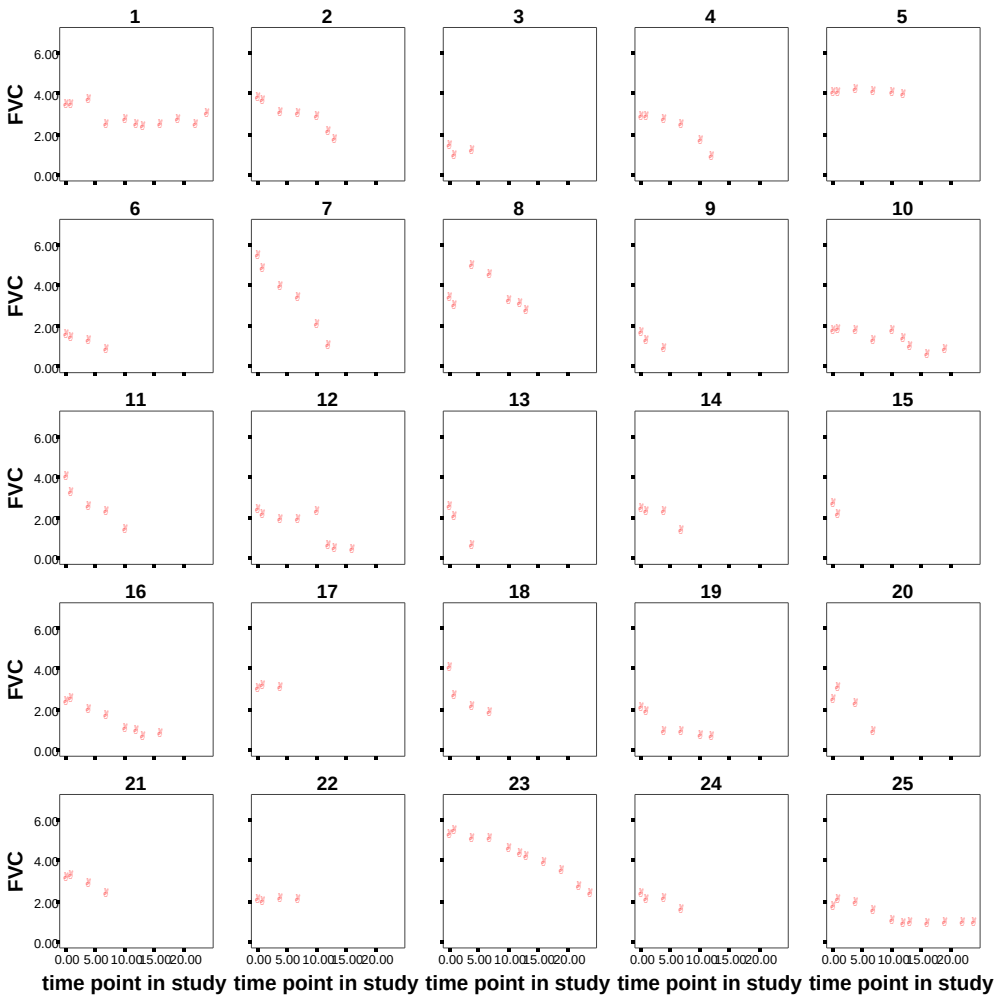
# SPSS

- Data format
  - In order to estimate a linear mixed model in SPSS the data need to be ‘under each other’
- Preliminary analysis
  - A useful feature of SPSS 11 is
  - **Graphs** → **Interactive** → **Scatter…**
- This produces, amongst other things, panel graphs (with horrendous syntax)

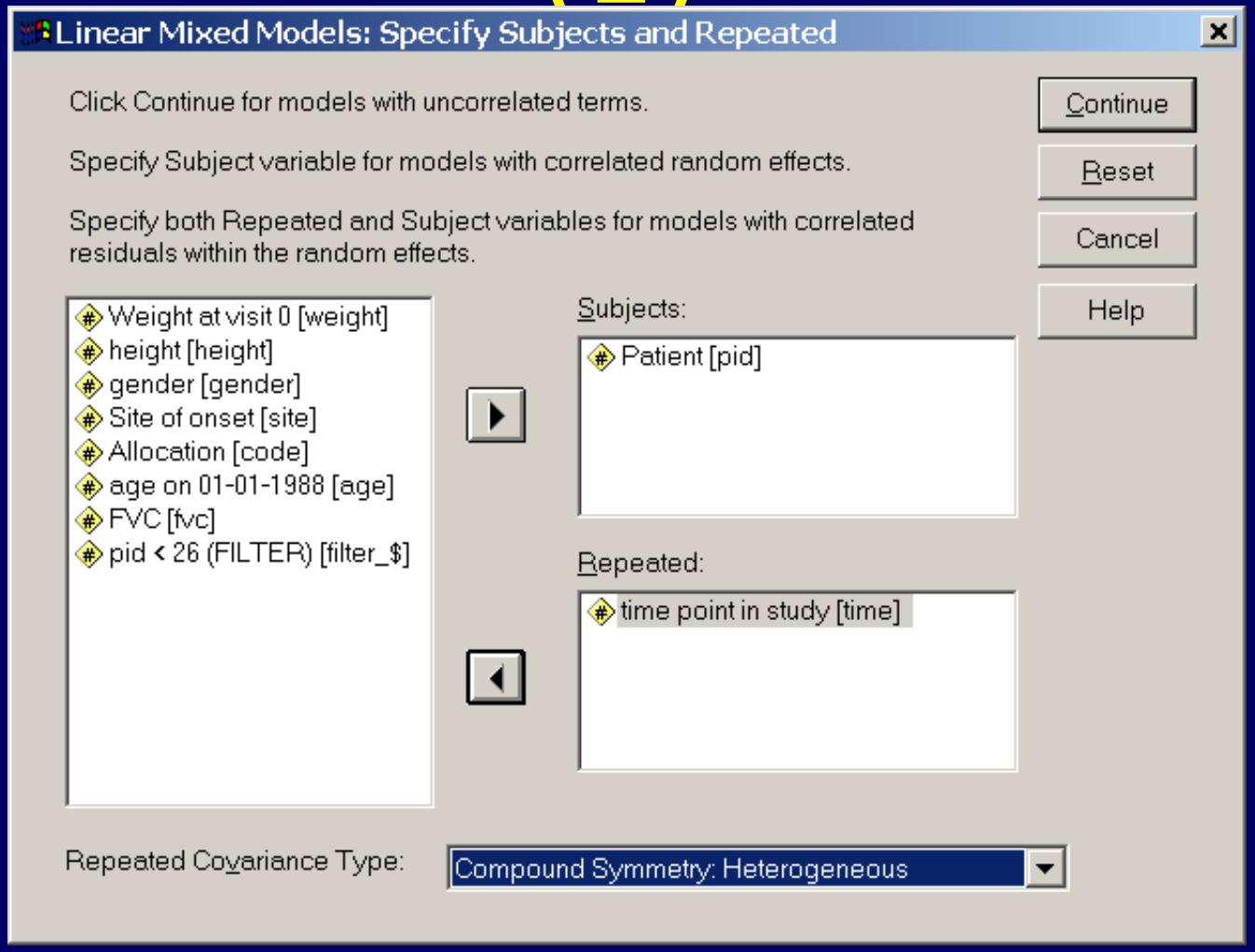
# Preliminary: SPSS menu



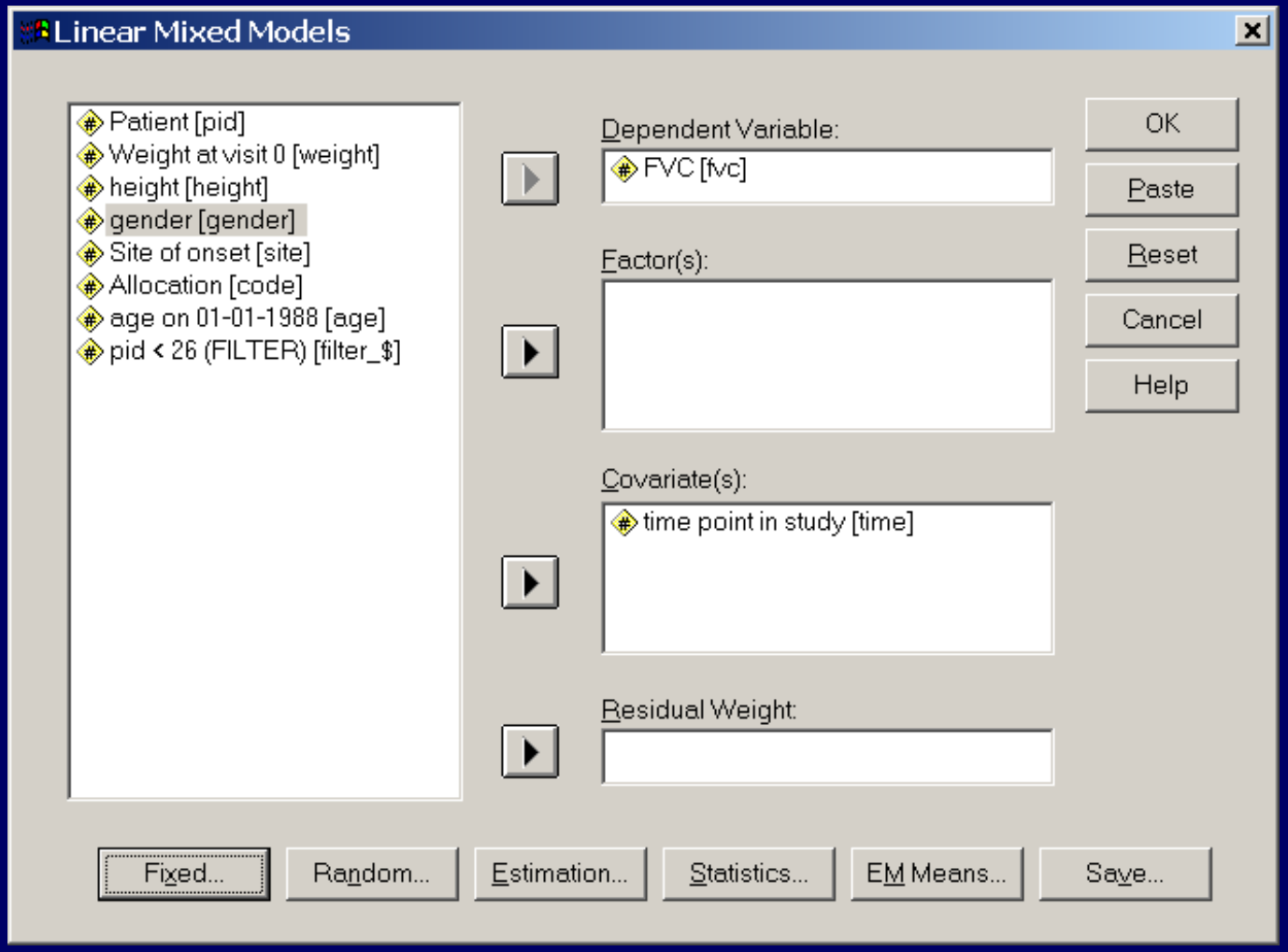
# Preliminary: SPSS output



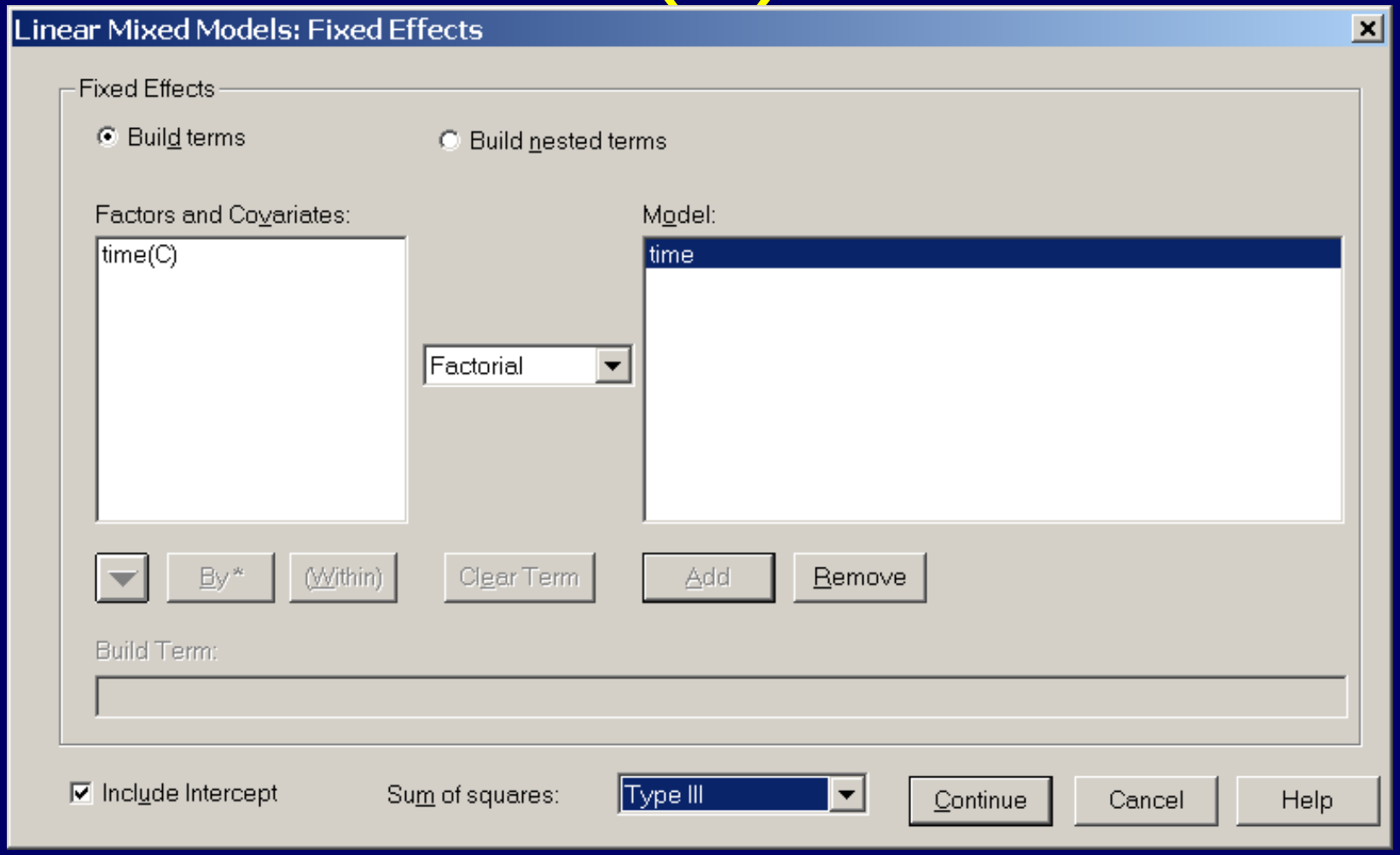
# Mixed models: SPSS menu (1)



# Mixed models: SPSS menu (2)



# Mixed models: SPSS menu (3)



# Mixed models: SPSS output

SPSS Pivot Table - table1

File Edit View Insert Pivot Format Help

**Type III Tests of Fixed Effects[a]**

| Source | Numerator df | Denominator df | F | Sig. |
|---|---|---|---|---|
| Intercept | 1 | 21.035 | 257.447 | .000 |
| gender | 1 | 63.914 | 39.042 | .000 |
| site | 2 | 59.082 | 2.754 | .072 |
| time | 1 | 72.370 | 154.976 | .000 |

a. Dependent Variable: FVC.

SPSS Pivot Table - table2

File Edit View Insert Pivot Format Help

**Estimates of Fixed Effects[b]**

| Parameter | Estimate | Std. Error | df | t | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|---|
| | | | | | | Lower Bound | Upper Bound |
| Intercept | 4.012084 | .2670330 | 47.873 | 15.025 | .000 | 3.4751419 | 4.5490270 |
| [gender=0] | -1.84939 | .2959792 | 63.914 | -6.248 | .000 | -2.4406876 | -1.2580842 |
| [gender=1] | 0[a] | 0 | . | . | . | . | . |
| [site=1] | -.4971426 | .3995606 | 56.683 | -1.244 | .219 | -1.2973459 | .3030607 |
| [site=2] | -.7196616 | .3170634 | 67.782 | -2.270 | .026 | -1.3523885 | -.0869346 |
| [site=4] | 0[a] | 0 | . | . | . | . | . |
| time | -.0956064 | .0076799 | 72.370 | -12.449 | .000 | -.1109146 | -.0802981 |

a. This parameter is set to zero because it is redundant.

b. Dependent Variable: FVC.

# Summary

- Repeated measurements: observations not independent -> standard techniques not appropriate
- Often: simple solutions satisfactory and sufficient
- Sometimes: more complex approach is desired:
  - Specific effects (trends, interactions)
  - Efficient use of data for small effects