Linear Mixed Models

PGRM 15

Outline

- Linear regression
- Correlated measurements (*eg* repeated)
- Random effects leading to different components of variance & correlated measurements
- Different Correlation Structures
- Simple Analysis of Clustered Data
- Split Plot Analysis
- Repeated Measures Analysis

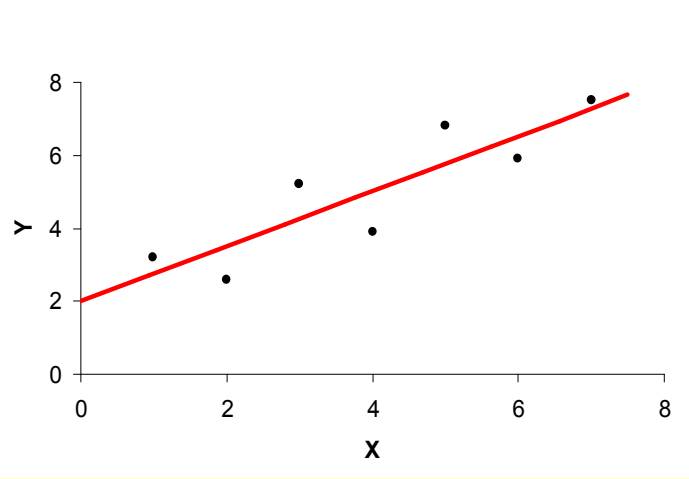
Relationship between variables

- Yield and fertilizer level
- Temperature and bacterial growth
- Light and respiration rate
- Age and alcohol consumption

Can the relationship be represented by a simple curve?

Is there necessarily a causal relationship?

Simple linear regression



Simple linear regression model

Data: (x_i, y_i)

Probabilistic model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon$$

Random term with variance σ^2

Population Characteristic	Sample Statistic
β_0	$\hat{\beta}_0$
β_1	$\hat{\beta}_1$
σ^2	s^2

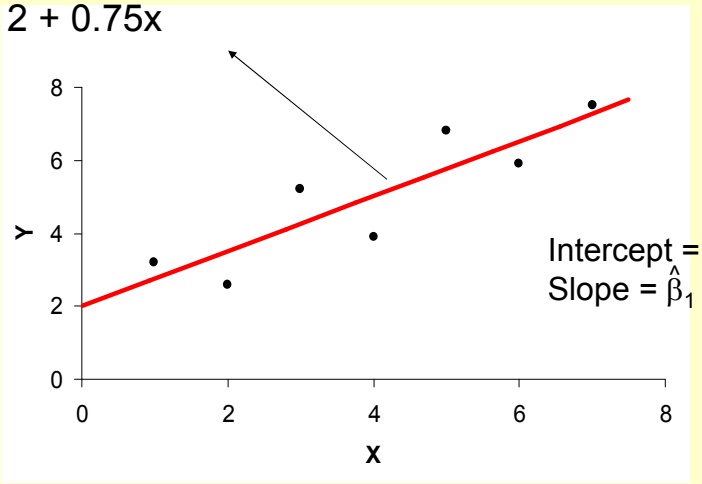
To predict from the model:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Simple linear regression

Estimated equation of the line:

$\hat{y} = 2 + 0.75x$



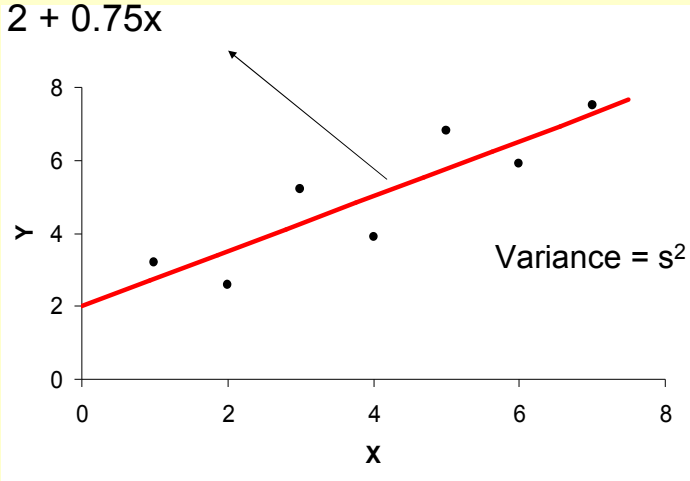
Intercept = $\hat{\beta}_0 = 2$

Slope = $\hat{\beta}_1 = 0.75$

Simple linear regression

Estimated equation of the line:

$\hat{y} = 2 + 0.75x$

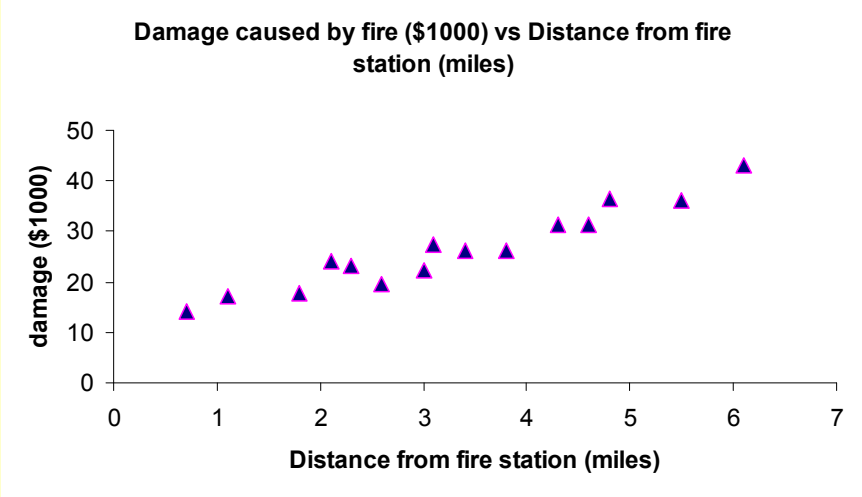


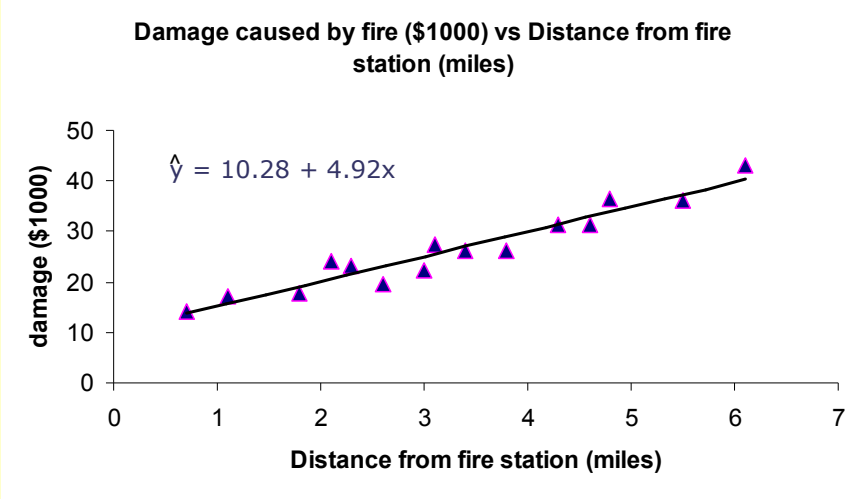
Assumptions for a Simple Linear Regression model

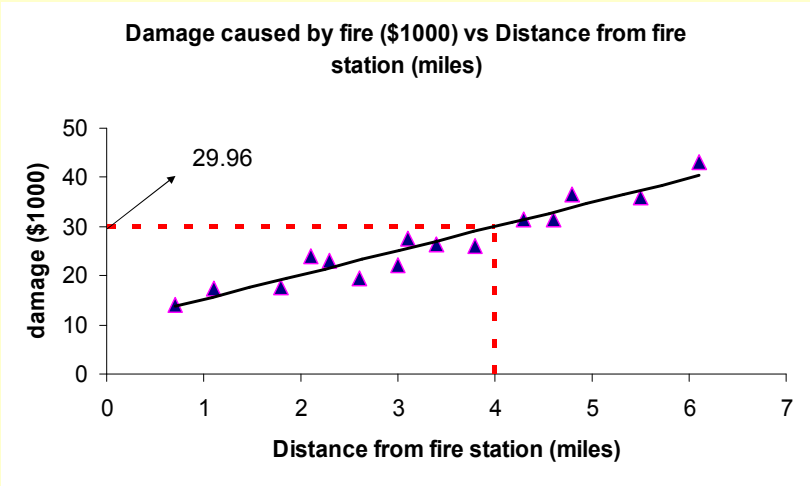
Note: If you are fitting a simple linear regression model to your own data, there are assumptions that must be satisfied. Find details of how to test the assumptions for your fitted model in any basic statistics text book.

Regression example

Distance from fire station x (miles)	Fire damage y ($1000)
3.4	26.2
1.8	17.8
4.6	31.3
2.3	23.1
3.1	27.5
.	.
.	.




Damage caused by fire ($1000) vs Distance from fire station (miles)
ŷ = 10.28 + 4.92x
damage ($1000)
50
40
30
20
10
0
0
1
2
3
4
5
6
7
Distance from fire station (miles)


Damage caused by fire ($1000) vs Distance from fire station (miles)
29.96
damage ($1000)
50
40
30
20
10
0
0
1
2
3
4
5
6
7
Distance from fire station (miles)

Correlated Measurements

- Two values are correlated if knowledge of one provides information about the other
  eg if knowing one value is high (relative to the appropriate mean) means the other is also likely to be high (+ve correlation), or low (-ve correlation)
- Correlation can be visualized by trends in plots
- Independent values are uncorrelated
- Positive correlation often arises between values resulting from shared influences
- When modelling, shared influences must be taken into account (*eg* by including random effects, or specifying a correlation structure)

Pearson's correlation coefficient ρ (rho)

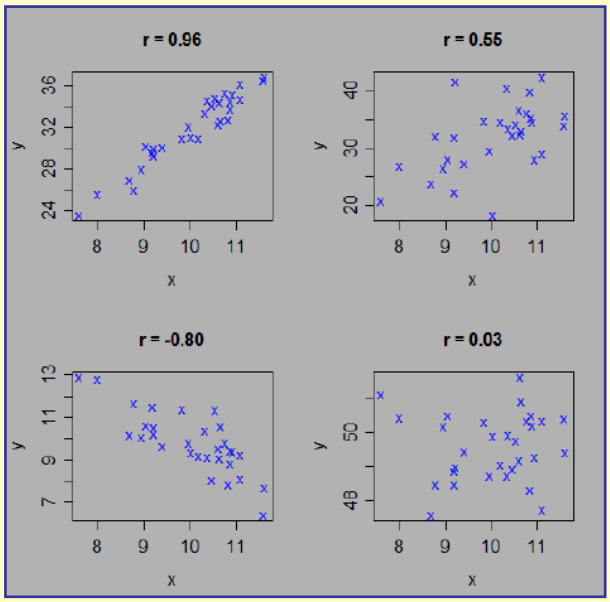
- Measures the strength and direction (+ve or –ve) of linear relationship between two variables measured on the same experimental unit
- Does not depend on the units of measurement
- Lies between –1 and +1
- In fitting a regression line y = a + bx the estimate of ρ, r, gives:

  $100r^2$ = % of variation in y
  explained by the linear regression

  Hence r = –0.5 indicates a negative relationship that explains 25% of the variation in y.
  r = 0.8 explains approximately 2/3 of the variation.

Interpreting r

A high value of r corresponds to points closely scattered around a line



Examples of correlated responses

- Clustered data
  - Measurements on piglets from several litters
  - responses from piglets within a litter are related, sharing parental and environmental influences
  - Experiment at several sites – responses within a site are related
- Split-plot designs (with main and sub-plots – *see later*)
  - Responses from the same main plot will be related, sharing the main plot characteristics
- Repeated Measures from experimental units
  - responses from the same unit are related, sharing the influential characteristics of the unit
  - if repeated over time, responses from the same unit close in time may be more closely related than responses further apart

Identifying Correlation
&
Different Sources of Variation

Components of variation & correlation

- So far background noise has been accounted for by a single source of variation – the variation between experimental units sharing the same values of explanatory variables.
- With clustered data, some of the variation will be due to differences between clusters (*eg* litters, sites). Individuals within a litter are related genetically. Plots within a site are related as they share common environmental elements.
- Clusters may form a source of random variation (any particular litter could have been allocated to a different treatment group but all piglets in the litter would then have the same treatment and would not be independent.
- Variation between sites may contribute to the variation of the estimate of a treatment effect based on a number of sites. The treatment effect may also vary over sites.
- Observations with the same value of a random effect will be positively correlated

More general correlation structures

- Introducing a random effect for clusters of measurements assumes a very restricted structure of correlation
- More generally we
  - identify clusters whose members are correlated
  - specify the correlation structure

Example of clustered data analysis

27 sites with **30** plots per site = **810** plots

Four species – 2 grasses (G) & 2 legumes (L)
Simplex design at two densities

Location of Sites

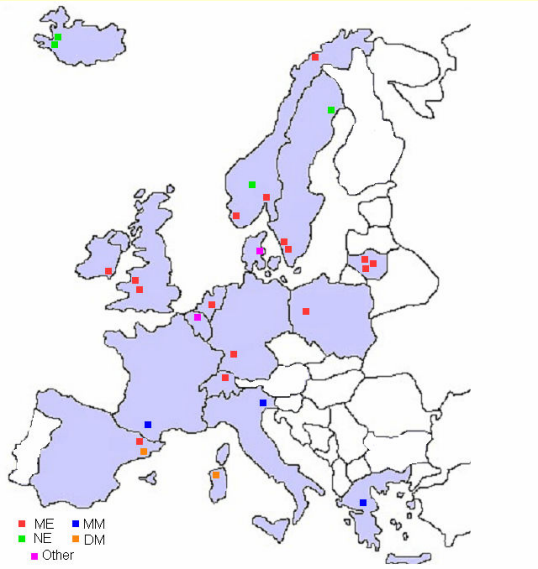
16 Mid European (ME)

4 Northern European (NE)

3 Moist Mediterranean (MM)

2 Dry Mediterranean (DM)

2 Other



A typical ME site

Species

- G_1 *Lolium perenne*
- G_2 *Dactylis glomerata*
- L_1 *Trifolium pratense*
- L_2 *Trifolium repens*

First complete year's data was used.

Diversity effect is positive and significant at most sites

Country	Type	Av mono (t ha^{-1})	$\hat{\delta}$ (t ha^{-1})
Germany	ME	13.6	4.84
Ireland	ME	13.6	3.75
Lithuania	ME	5.5	0.34
Lithuania	ME	8.7	2.44
Lithuania	ME	9.1	2.17
Netherlands	ME	9.1	3.03
Norway	ME	7.8	6.85
Norway	ME	7.1	4.18
Norway	ME	9.9	2.13
Poland	ME	6.7	1.77
Spain	ME	6.8	2.01
Sweden	ME	5.6	5.75
Sweden	ME	8.8	2.17
Switzerland	ME	10.4	5.64
Wales	ME	6.7	4.37
Wales	ME	6.8	4.68
Average		8.5	3.5

Site fixed vs site random

Site Fixed

```
Proc glm data=yield;
  Class site monomix density;
  Model tot_yld=site density
          monomix ;
  Lsmeans monomix/stderr;
Run;
```

Diff 2.39 SE 0.133

LSmeans

	Mean	SE
Mono	7.8	0.114
Mix	10.2	0.069

Site Random

```
Proc mixed data=yield;
  Class site monomix density;
  Model tot_yld= density monomix;
  Lsmeans monomix/stderr;
  Random site
  Run;
```

Diff 2.39 SE 0.133

LSmeans

	Mean	SE
Mono	7.8	0.735
Mix	10.2	0.730

Monomix tested against random site*monomix

- Estimate of diversity effect is the same (2.39)
- SE of diversity effect is now greater (0.254)

Loss of precision is compensated for by a wider range of inference about the diversity effect

Any new site – predict a diversity effect of 2.39 but use the se = 0.254 for setting bounds for the prediction.

Unbalanced mixed model analysis

- The example data was balanced – each treatment (combination of a level of V with a level of N) appeared the same number of times (once!) in each block
- For unbalanced data every comparison has a different SE.
- It's important
  - to use the
    `ddfm = kenwardroger`
    option on the `model` statement
  - to drop interaction terms if non-significant

See PGRM 15.2.1.2

Cost data revisited - meta analysis

- **Consistency of treatment effects across multiple sites (meta analysis)**
- If an experiment is repeated across a number of sites the question arises as to how well the average treatment effect estimated across all sites represents the difference between treatments.
- Is there variation in the Monomix effect across sites?
- If sites are regarded as random (drawn at random from some population of sites), how well does the average Monomix effect over sites represent the 'typical' monomix effect at a site drawn from the same population..
- To use the average monomix effect in this way its should contain variation due to how the effect varies from site to site (i.e. information on the site x monomix interaction. This is achieved by including in the SAS program the line

random site site*monomix;

Statistics in Science

Diversity effect is positive and significant at most sites – Meta analysis

Country	Type	Av mono (t ha^{-1})	$\hat{\delta}$ (t ha^{-1})
Germany	ME	13.6	4.84
Ireland	ME	13.6	3.75
Lithuania	ME	5.5	0.34
Lithuania	ME	8.7	2.44
Lithuania	ME	9.1	2.17
Netherlands	ME	9.1	3.03
Norway	ME	7.8	6.85
Norway	ME	7.1	4.18
Norway	ME	9.9	2.13
Poland	ME	6.7	1.77
Spain	ME	6.8	2.01
Sweden	ME	5.6	5.75
Sweden	ME	8.8	2.17
Switzerland	ME	10.4	5.64
Wales	ME	6.7	4.37
Wales	ME	6.8	4.68
Average		8.5	3.5

Statistics in Science

SAS programs - Site random

```
proc mixed data=yield;
   class site monomix density;
   model tot_yld = density monomix
   /ddfm=kenwardroger ;
   random  site;
   lsmeans monomix;
   estimate 'diversity effect' monomix -1 1;
run;
```

SAS programs - Site random

```
proc mixed data=yield;
   class site monomix density;
   model tot_yld = density monomix
   /ddfm=kenwardroger ;
   random  site;
   random  site*monomix;
   lsmeans monomix;
   estimate 'diversity effect' monomix -1 1;
run;
```

Site fixed vs site random and Site*monomix included

Site Fixed

```
Proc glm data=yield;
  Class site monomix density;
  Model tot_yld=site density
        monomix site*monomix;
  Lsmeans monomix/stderr;
Run;
```

Diff 2.39 SE 0.128

LSmeans

	Mean	SE
Mono	7.8	0.109
Mix	10.2	0.066

Site Random

```
Proc mixed data=yield;
  Class site monomix density;
  Model tot_yld= density monomix;
  Lsmeans monomix/stderr;
  Random site
  Random site*monomix
  Run;
```

Diff 2.39 SE 0.254

LSmeans

	Mean	SE
Mono	7.8	0.714
Mix	10.2	0.709

Comparison of models

Random effects	SITE	SITE SITE*MONOMIX
Variance Estimates		
Site	15.3	13.7
Site * Monomix		0.70
Residual	3.02	2.75
-2 log Likelihood	3572.6	3531.7
Mono	7.81	7.81
SE	0.735	0.714
Mixture	10.20	10.20
SE	0.730	0.709
Diversity effect	2.39	2.39
SE	0.133	0.254
P value	<0.001	<0.001

Split Plot Experimental Design

Split Plot: effect of fertiliser on 3 varieties of oats

In each of 6 homogeneous blocks: Varieties allocated to main plots at random

In each sub plot varying levels of fertiliser N (kg/ha) are allocated at random

Block 1				
Mainplot 1	Victory 25	Victory 0	Victory 75	Victory 50
Mainplot 2	Marvellous 0	Marvellous 50	Marvellous 25	Marvellous 75
Mainplot 3	Golden Rain 75	Golden Rain 25	Golden Rain 50	Golden Rain 0
Block 2				
Mainplot 1	Marvellous 50	Marvellous 0	Marvellous 75	Marvellous 25
Mainplot 2	Victory 75	Victory 50	Victory 0	Victory 25
Mainplot 3	Golden Rain 25	Golden Rain 75	Golden Rain 50	Golden Rain 0
⋮			⋮	
Block 6				
Mainplot 1	Marvellous 75	Marvellous 0	Marvellous 25	Marvellous 50
Mainplot 2	Golden Rain 25	Golden Rain 50	Golden Rain 0	Golden Rain 75
Mainplot 3	Victory 50	Victory 75	Victory 25	Victory 0

Analysis of split plot design

- Interest is on
  - effect of V (variety) and N (nitrogen level) on Y (yield)
  - possible sources of background variation
- Initially regard N level as categorical
- Model terms would include B (block), V, N and the interaction V*N
- In addition model must reflect the split plot structure: yields from the same main plot will share similar (randomly allocated) characteristics, and consequently be correlated
- Main plots are identified by specifying B and V, *ie* B*V.
- SAS/MIXED procedure caters for the random effects of B*V.

SAS/MIXED code

```
proc mixed data = oats;
     class b v n;
     model y = b n v n*v / htype = 1;
     random b*v;
     estimate 'Variety 1v2' v 1 -1 0;
     estimate 'N 1v2' n 1 -1 0 0;
run;
```

b*v

each level of this corresponds to a main plot which contributes its characteristics to the response

SAS/MIXED OUTPUT

Type 1 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
b	5	10	5.28	0.0124
n	3	45	37.69	<.0001
v	2	10	1.49	0.2724
v*n	6	45	0.30	0.9322

1. V*N interaction is far from significant (so the model could be re-run without this term)
2. N has a strong effect
3. Varieties do not appear to differ

SAS/MIXED OUTPUT

Covariance Parameter Estimates	
Cov Parm	Estimate
b*v	106.06
Residual	177.08

There are <u>2 sources of background variation</u>:

1. within main plots (component of variance: 177.08)
2. between main plots (component of variance: 106.06)

Different comparisons will use appropriate combinations of these

LSmeans & SEDs

- Code
  `lsmeans v n v*n / diff;`
- Main effect means (*ie* for v (SED = 7.079), n (SED = 4.436) and SEDs are appropriate if v*n is non-significant
- Comparison of the 12 v*n means involves 2 different SEDs
  - When comparing 2 levels of N within a particular variety (SED = 7.683)
  - When making comparisons where the variety changes (SED = 9.715)

lsmeans / diff; OUTPUT

fferences of Least Squares Means

Effect	Variety	Nitrogen	Variety	Nitrogen	Estimate	Standard Error	DF	t Value	Pr > |t|
v	Golden Rain		Marvellous		-5.2917	7.0789	10	-0.75	0.4720
v	Golden Rain		Victory		6.8750	7.0789	10	0.97	0.3544
v	Marvellous		Victory		12.1667	7.0789	10	1.72	0.1164
n		0		25	-19.5000	4.4358	45	-4.40	<.0001
n		0		50	-34.8333	4.4358	45	-7.85	<.0001
n		0		75	-44.0000	4.4358	45	-9.92	<.0001
n		25		50	-15.3333	4.4358	45	-3.46	0.0012
n		25		75	-24.5000	4.4358	45	-5.52	<.0001
n		50		75	-9.1667	4.4358	45	-2.07	0.0446
v*n	Golden Rain	0	Golden Rain	25	-18.5000	7.6830	45	-2.41	0.0202
v*n	Golden Rain	0	Golden Rain	50	-34.6667	7.6830	45	-4.51	<.0001
v*n	Golden Rain	0	Golden Rain	75	-44.8333	7.6830	45	-5.84	<.0001
v*n	Golden Rain	0	Marvellous	0	-6.6667	9.7150	45	-0.69	0.4961
v*n	Golden Rain	0	Marvellous	25	-28.5000	9.7150	45	-2.93	0.0053

Is there a simple description of the N effect?

Using SAS/MEANS procedure calculate the mean yield for each treatment (V, N combination)

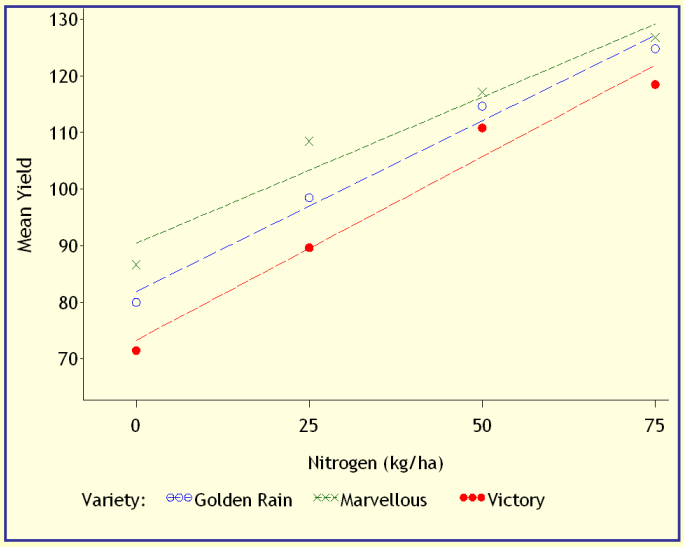
Plot

Mean Yield v Nitrogen Level

for each variety

Fit a line for each variety.

Split plot design



SAS/MIXED: Split Plot Analysis

- N seems to have equal effect on yield for the 3 varieties
- Plot suggests effect of N is linear
- Add variable `nval = n` to data set `oats`
- Fit & test for separate slopes and LOF:

```
proc mixed data = oats;
     class b v n;
     model y = b nval n v nval*v n*v
          / htype = 1;
     random b*v;
run;
```

`n` `n*v` terms measure LOF
`nval*v` checks equal slopes

SAS OUTPUT

Type 1 Tests of Fixed Effects				
Effect	*Num DF*	*Den DF*	*F Value*	*Pr > F*
b	5	10	5.28	0.0124
nval	1	45	110.32	<.0001
n	2	45	1.37	0.2653
v	2	10	1.49	0.2724
*nval*v*	2	45	0.48	0.6248
*v*n*	4	45	0.22	0.9279

1. LOF terms not significant (p = .27, .93)
2. Linear effect of N same for the 3 varieties (p = .63)
3. Linear effect of N highly significant (p < .0001)
4. Refit model without LOF terms and interaction `nval*v` to get coefficient of `nval`: 0.59 (SE: 0.0543)

Practical exercise

SAS/MIXED for Split Plot Analysis

Lab Session 6 Exercise 6.2 (a) – (e)

Simple Analyses of Correlated Data

A simple analysis – combine correlated values

- Seedlings are planted 20 in each of 10 containers and the dry mass of each plant measured after 60 days.
  - Responses from plants in the same container will be correlated.
  - Use as response variable the total dry mass of all 20 plants in the same container. This gives 10 independent response values

- Responses from piglets in the same litter could be combined, though adjustments would have to be made for varying litter size. Averages from large litters could give more precise information than from small, but may also be lower.

A simple analysis – combine correlated values

- Measuring plant height every 14 days for 10 weeks gives 6 measurements for each plant (after 0, 2, 4, 6, 8, and 10 weeks)

- If daily growth is of interest a single value from each plant could be calculated by

  (final height – initial height)/140

  the slope of the fitted line
  to the data from the plant

Repeated Measures Analysis

- Mutiple measures on the same units at the same time
- Repeated measurements on the same units over time

Statistics in Science

Example

The weights of 27 male and female children were measured at ages 8, 10, 12 and 14 years.

Child	Gender	8	10	12	14
1	F	21.0	20.0	21.5	23.0
2	F	21.0	21.5	24.0	25.5
3	F	20.5	24.0	24.5	26.0
4	F	23.5	24.5	25.0	26.5
5	F	21.5	23.0	22.5	23.5
6	F	20.0	21.0	21.0	22.5
7	F	21.5	22.5	23.0	25.0
8	F	23.0	23.0	23.5	24.0
9	F	20.0	21.0	22.0	21.5
10	F	16.5	19.0	19.0	19.5
11	F	24.5	25.0	28.0	28.0
12	M	26.0	25.0	29.0	31.0
13	M	21.5	22.5	23.0	26.5
14	M	23.0	22.5	24.0	27.5
15	M	25.5	27.5	26.5	27.0
16	M	20.0	23.5	22.5	26.0
17	M	24.5	25.5	27.0	28.5
18	M	22.0	22.0	24.5	26.5
19	M	24.0	21.5	24.5	25.5
20	M	23.0	20.5	31.0	26.0
21	M	27.5	28.0	31.0	31.5
22	M	23.0	23.0	23.5	25.0
23	M	21.5	23.5	24.0	28.0
24	M	17.0	24.5	26.0	29.5
25	M	22.5	25.5	25.5	26.0
26	M	23.0	24.5	26.0	30.0
27	M	22.0	21.5	23.5	25.0

To keep SAS happy replace 8, 10, etc with w_age8, w_age10, etc

Statistics in Science

Issues in Repeated Measures Analysis

1. What questions are to be addressed by the analysis?

   Can these be answered by reducing the measurements to a single value?

   eg

   - final weight
   - weight gain over a particular period (from 8 to 10 *or* 10 to 14 yr)
   - the slope of a straight line

Issues in Repeated Measures Analysis

2. Are the measurements taken at equally spaced time points?

   When times are not equally spaced, and even more so when patterns of timing differ from unit to unit, the analysis is more difficult

Issues in Repeated Measures Analysis

3. What is a reasonable error/correlation structure for the data?

- Are measurements equally related, irrespective of whether they are close together in time or otherwise?
- Does the correlation decline as time points are farther apart?
- Can this decline be described simply?

Child Weights Example

simple analysis using 2 possible single values

```
data c;
  set childWeights;
  wgain1 = y4 - y1;
run;

proc glm data = c;
  class gender;
  model wgain1 = gender;
  lsmeans gender;
  estimate 'diff M-F' gender -1 1;
quit;
```

SAS OUTPUT

Gain 8 –10 yr

Parameter	Estimate	Standard Error	t Value	Pr > |t|
diff M - F	-0.10795	0.79949	-0.14	0.8937

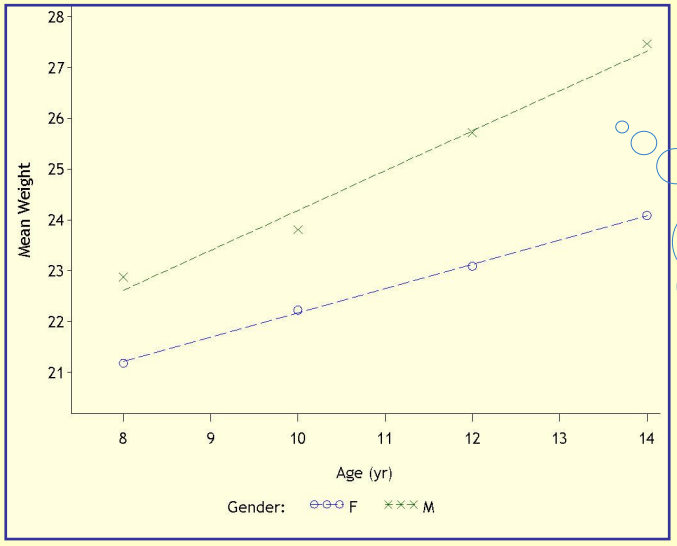
Gain 10 –14 yr

Parameter	Estimate	Standard Error	t Value	Pr > |t|
diff M - F	1.79261	0.62376	2.87	0.0082

Differences between genders depend on age
ie there is a
gender × year interaction
The complete data set needs to be analysed

Statistics in Science

Getting the picture!



Are the slopes equal?

Statistics in Science

Repeated Measures Analysis of Weight

Restructure data for SAS

Obs	Child	Gender	weight	age
1	1	F	21.0	8
2	1	F	20.0	10
3	1	F	21.5	12
4	1	F	23.0	14
5	2	F	21.0	8
6	2	F	21.5	10

Centre age using

```
cage = age - 11
```

This means that gender comparisons will be made at age 11 yr.

Repeated Measures Analysis of Weight

Observations from the same child will be correlated

A simple way to include this is to introduce a random effect for the child

<u>However</u>: this assumes that any two observations on the same child are equally correlated, independent of whether far apart or not (COMPOUND SYMMETRY)

```
proc mixed data = cw;
  class gender cage child;
  model weight = gender cage cage*gender
     / ddfm = kenwardroger htype = 1;
  random child;
run;
```

<u>Note</u>: it's best to use `model` option: `ddfm = kenwardroger`

MIXED – selected OUTPUT

Least Squares Means						
Effect	Gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Gender	F	22.6477	0.5861	25	38.64	<.0001
Gender	M	24.9688	0.4860	25	51.38	<.0001

Type 1 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Gender	1	25	9.29	0.0054
cage	1	79	122.45	<.0001
cage*Gender	1	79	6.30	0.0141

Covariance Parameter Estimates	
Cov Parm	Estimate
Child	3.2986
Residual	1.9221

The common correlation between observations on the same subject (child) is

$$3.3986/(1.9221 + 3.3986) = 0.63$$

MIXED – selected OUTPUT

Gender Difference

Differences of Least Squares Means							
Effect	Gender	Gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Gender	F	M	-2.3210	0.7614	25	-3.05	0.0054

- LSmeans and difference refer to cage = 0

MIXED – selected OUTPUT

Age Effect

Solution for Fixed Effects						
Effect	Gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		24.9688	0.4860	25	51.38	<.0001
Gender	F	-2.3210	0.7614	25	-3.05	0.0054
Gender	M	0	.	.	.	.
cage		0.7844	0.07750	79	10.12	<.0001
cage*Gender	F	-0.3048	0.1214	79	-2.51	0.0141
cage*Gender	M	0	.	.	.	.

Slopes F: 0.7844 – 0.3048 = 0.4796
M: 0.7844

These differ significantly (p = 0.0141)

MIXED – selected OUTPUT

rcorr option gives correlation between the 4 measurements on each subject (child)

Estimated R Correlation Matrix for Child 1				
Row	Col1	Col2	Col3	Col4
1	1.0000	0.6318	0.6318	0.6318
2	0.6318	1.0000	0.6318	0.6318
3	0.6318	0.6318	1.0000	0.6318
4	0.6318	0.6318	0.6318	1.0000

Note the CS structure and value of the correlation

Was compound symmetry a good choice for Child Weights example?

What are the alternatives?

Alternatives to Compound Symmetry

Alternatively we can specify that measurements on the same subject are correlated with compound symmetry or another structure as follows:

MIXED syntax:

replace

```
random child;
```

with

```
repeated / type = cs subject = child rcorr;
```

Choice of correlation structure:
`cs`
`ar(1)`
`unstructured`

With balanced data these are equivalent, but `repeated` permits more complex structures:

- AR(1) correlation gets less for measurement father apart being ρ, ρ^2, ρ^3, … for measurements 1, 2, 3, … time units apart (*eg* .5, .25, .125, …)
- UNSTRUCTURED no particular structure assumed

Getting the correlation structure correct

Unstructured correlation

Estimated R Correlation Matrix for Child 1				
Row	Col1	Col2	Col3	Col4
1	1.0000	0.5682	0.6589	0.5220
2	0.5682	1.0000	0.5806	0.7249
3	0.6589	0.5806	1.0000	0.7396
4	0.5220	0.7249	0.7396	1.0000

1. Not very different from CS
2. Other results similar
3. Correlation does not decrease with separation – don't use AR(1) correlation

Getting the correlation structure correct

Tests for correlation structure

- Chi-squared tests can compare CS and AR(1) with unstructured (UN) (they are included in UN)
- If both CS and AR(1) are not rejected when compared to UN there isn't a formal test; be guided by appropriateness and the principle of maximising likelihood
- For the formal test we need to note the number of estimated parameters in the structures being compared – the difference is the df for the relevant chi-squared test

Chi-squared test

- Need to re-run MIXED with different structures.
  The default method used by MIXED is REML which is best for determining random effects
- MIXED OUTPUT (using REML) includes a table of Fit Statistics including
  -2 Res Log Likelihood
- The number of covariance parameters is OUTPUT in the Dimensions
- Testing whether a simplified covariance structure will do the chi-squared df is the difference in covariance parameters

```
p = 1 - cdf('chisq', d, df);
```

where

d = change in (-2 Res Log Likelihood)

Child Weights Example

Structure	df	-2 Res Log Likelihood	Change from UN	p
UN	10	424.5		
CS	2	433.8	9.3 (8 df)	0.32
AR(1)	2	444.6	20.1 (8 df)	0.01

```
data _null_;
   p = 1-cdf('chisq', 9.3, 8);
   put p = ;
run;
```

Result - in SAS LOG:

```
p=0.317623866
```

- CS ok
- AR(1) rejected

More correlation structures

See SAS Help!

For example:

suppose some subjects are more variable than others, but there is still a common correlation between different measurements on the same subject.

Use

Heterogeneous CS:

```
type = csh
```

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Var(1)	Child	5.6959
Var(2)	Child	4.2362
Var(3)	Child	6.0896
Var(4)	Child	4.8253
CSH	Child	0.6353

Structure	df	-2 Res Log Likelihood	Change from CSH	p
CSH	5	432.0		
CS	2	433.8	1.8 (3 df)	0.61

CS not rejected when compared to CSH

SAS/MIXED code – The Idea Summarised

- Values that are correlated must be identified - each set of correlated values being a level of some factor
- Introduce a variable identifying correlated values
- Include it in the `class` statement
- Use a `repeated` statement & specify the correlation structure
- Alternatively: use a `random` statement if the correlation is simply due to units sharing a common influences

Practical exercise

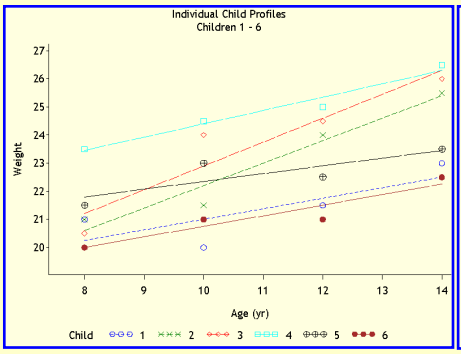
SAS/MIXED for Repeated Measures Analysis

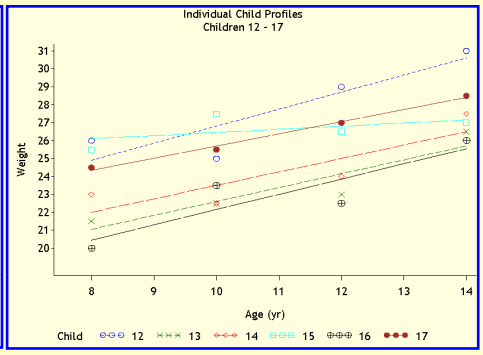
Lab Session 6 Exercise 6.3 (a) – (d)

Random coefficients model

- Previous Model assumes a single regression line, with subjects (children) deviating from this each year with correlated deviations.
- More generally we could allow each subject to have its own regression line;
  this has the advantage in being possible when subjects are measured at different times
- Since subjects are randomly allocated to treatments, the individual intercepts and slopes are random variables
- We fit a fixed line (intercept and slope being the estimated mean for all subjects)
- Each subject has its own line resulting from adding/subtracting from the mean intercept and slope.

Random coefficients model





6 (of 11) females

6 (of 16) males

Plots show a line fit to each child's data
(for the first 6 males and 6 females)

SAS/MIXED: random coefficients model

SAS/MIXED code

```
proc mixed data = cw;
  class gender child;
  model weight = gender cage gender*cage /
     ddfm = kenwardroger htype = 1 solution;
  random intercept cage /
     type = un subject = child gcorr;
run;
```

MIXED: selected OUTPUT

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
UN(1,1)	Child	5.7864
UN(2,1)	Child	-0.2896
UN(2,2)	Child	0.03252
Residual		1.7162

Estimated G Correlation Matrix				
Row	Effect	Child	Col1	Col2
1	Intercept	1	1.0000	-0.6676
2	age	1	-0.6676	1.0000

Type 1 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Gender	1	25	9.29	0.0054
cage	1	25	99.45	<.0001
cage*Gender	1	25	5.12	0.0326

1. Correlation between intercept and slope: -0.67
2. Variances: intercept 5.79, slope 0.033, residual 1.72
3. Conclusions: the pattern of growth differs between males and females (p = 0.0326), as with earlier analysis. Average slopes unchanged.

Statistics in Science

The End

but hopefully just the beginning!

Statistics in Science