Longitudinal Data Analysis Using SAS®

Nebraska SAS Users Group Conference

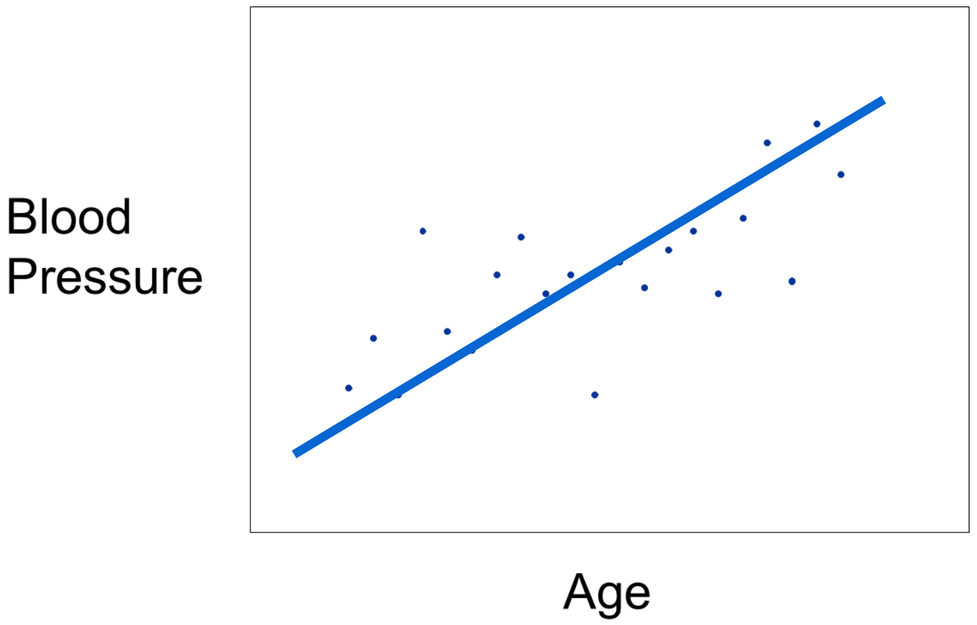
May 14, 2024

Jacqueline Johnson, DrPH, SAS Global Academic Programs

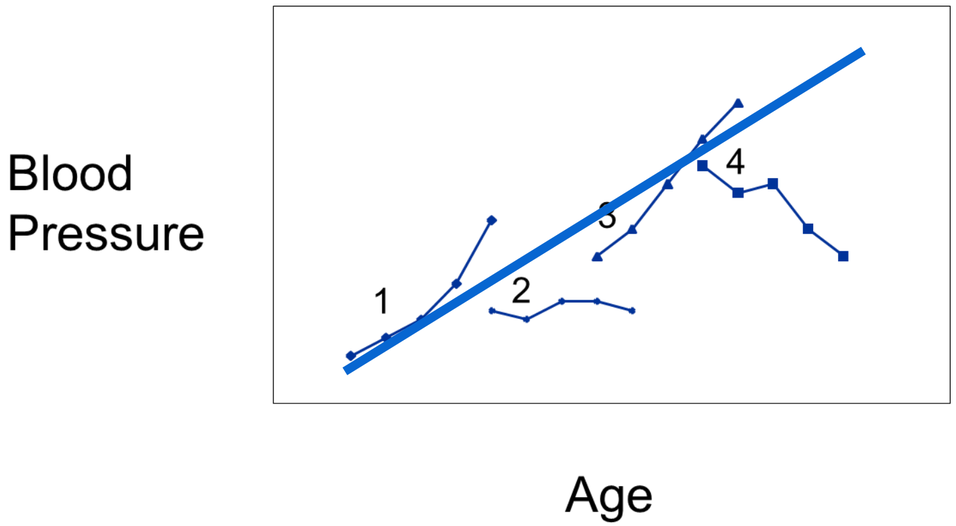
Longitudinal Data Analysis

The defining feature is that repeated measurements are taken on the same subject through time.

Cross-Sectional Data

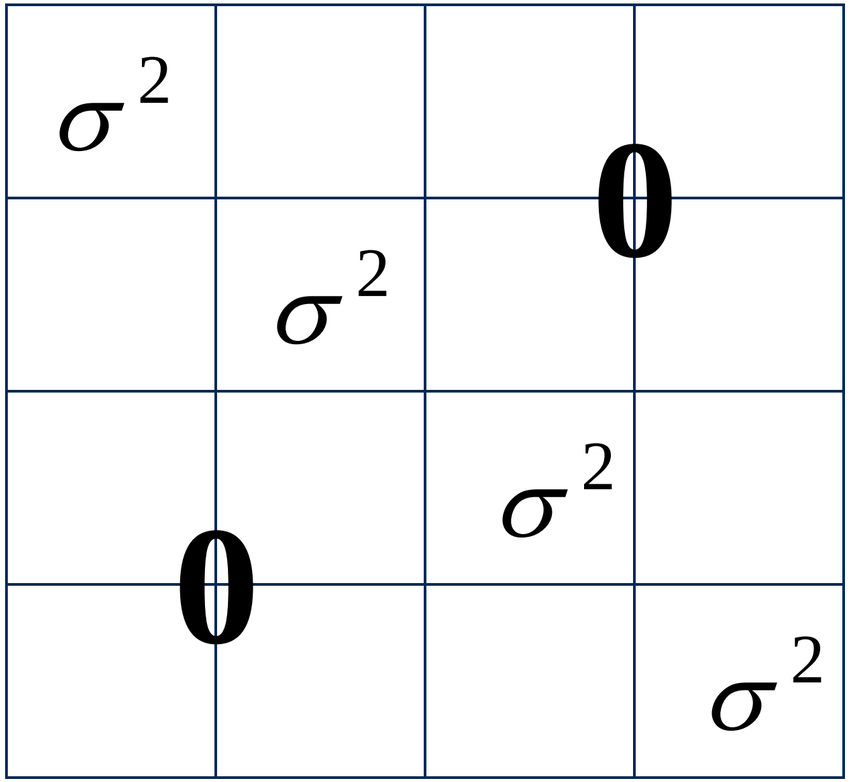


Longitudinal Data



Variance-Covariance Matrix for OLS Regression

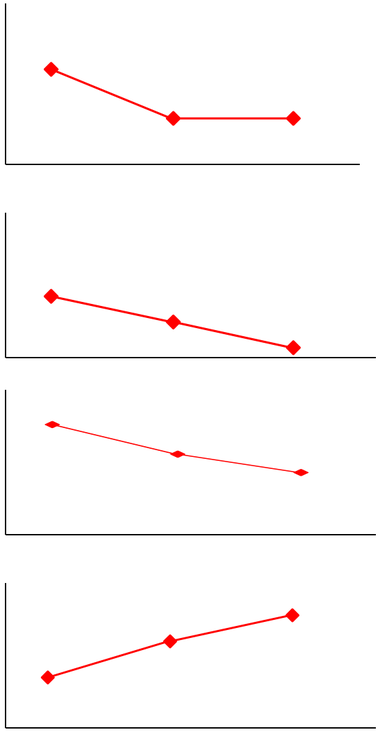
Subject	X	Y
1	4	10
2	2	7
3	6	12
4	8	11



Assumes all observations independent. Not appropriate for longitudinal data.

Longitudinal Data

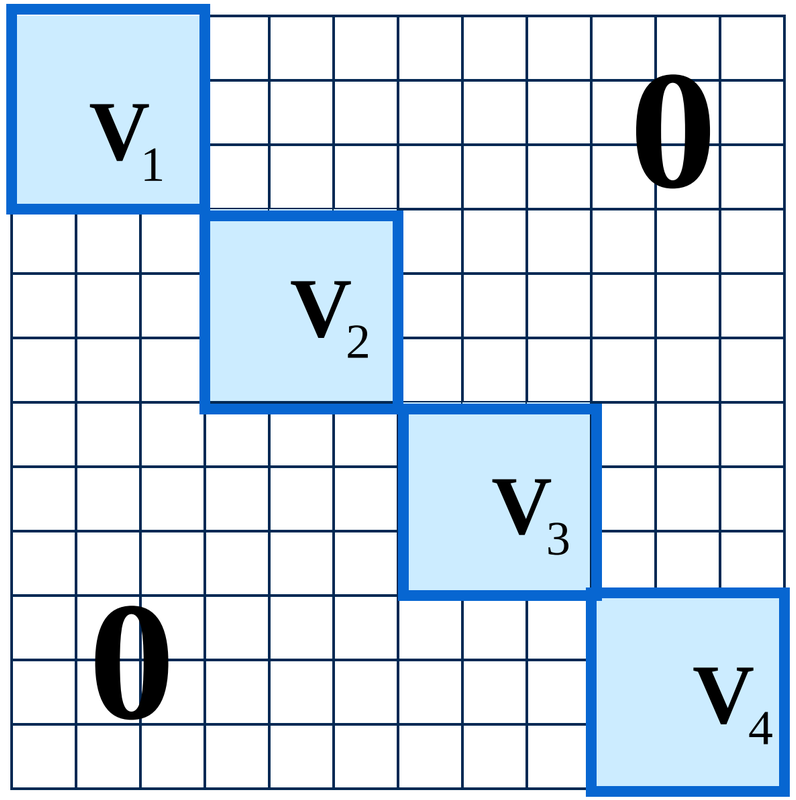
Subject	X	$Y_{t=1}$	$Y_{t=2}$	$Y_{t=3}$
1	4	10	6	6
2	2	7	5	3
3	6	12	9	8
4	8	11	14	16



Variance-Covariance Matrix for Longitudinal Data

Subject	Time	X	Y
1	1	4	10
1	2	4	6
1	3	4	6
2	1	2	7
2	2	2	5
2	3	2	3
3	1	6	12
3	2	6	9
3	3	6	8
4	1	8	11
4	2	8	14
4	3	8	16

Size of V determined by number of observations for each subject.



Same type of covariance structure for all subjects.

Model Notation: General Linear Model

$$y = \mathbf{X}\beta + \varepsilon$$

where y is the vector of observed responses

$\mathbf{X}$ is the design matrix of predictor variables

β is the vector of regression parameters

ε is the vector of random errors.

Not appropriate. Assumes ε independent.

Model Notation: General Linear Mixed Model

$$y = X\beta + Z\gamma + \varepsilon$$

Var(y)= V = ZGZ' + R

where Z is the design matrix of random variables

γ is the vector of random-effect parameters

Var(gamma) = G

ε is no longer required to be independent and homogeneous.

Var(epsilon) = R

MIXED Procedure

General form of the MIXED procedure:

$$Var(y) = V = ZGZ' + R$$

```
PROC MIXED DATA=SAS-data-set <options>;
        CLASS variables;
        MODEL response=<fixed effects></ options>;
        RANDOM random effects </ options>;
        REPEATED <repeated effect> </ options>;
RUN;
```

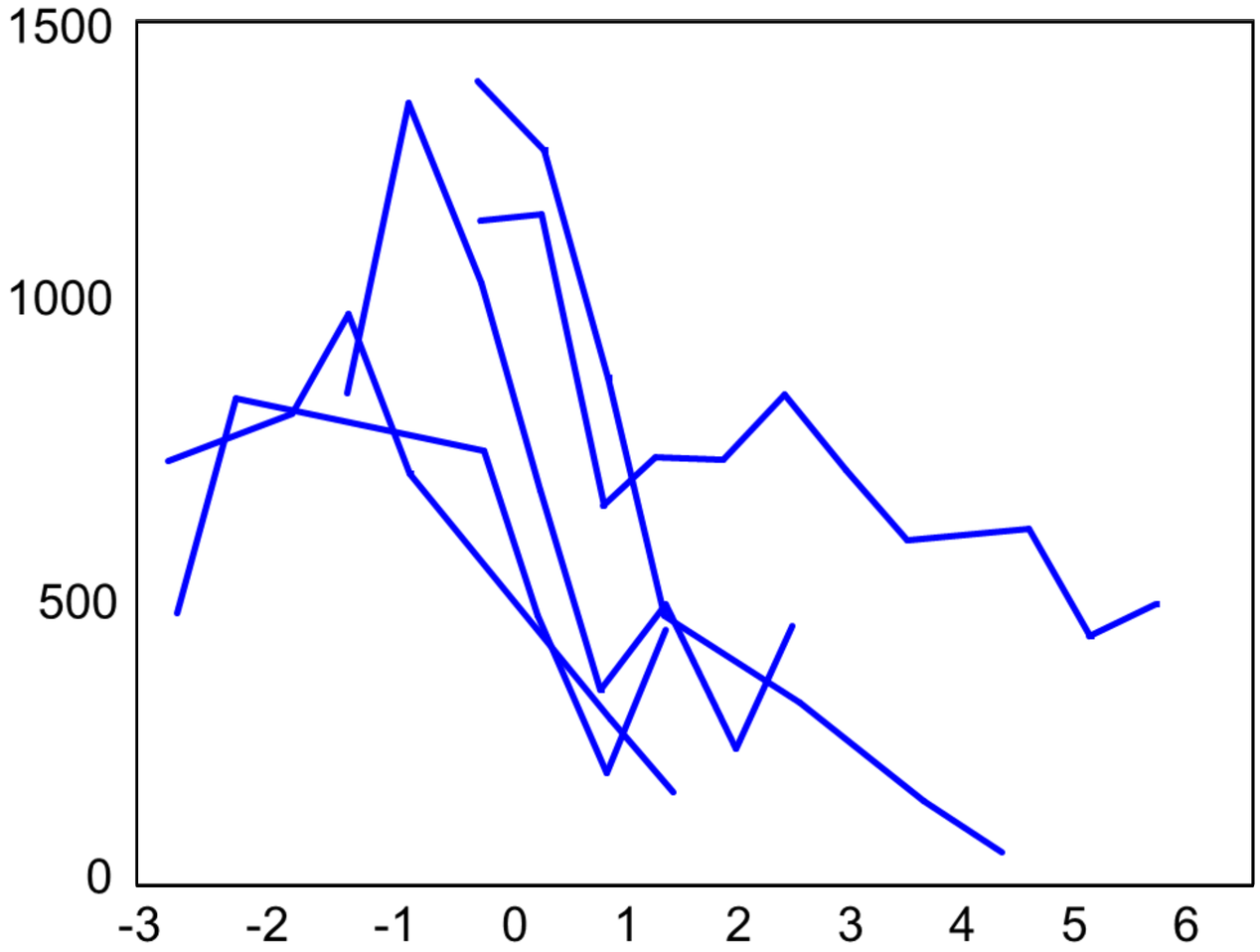
RANDOM Models G

REPEATED Models R

The RANDOM and REPEATED statements together model the variance and covariance properties of the data

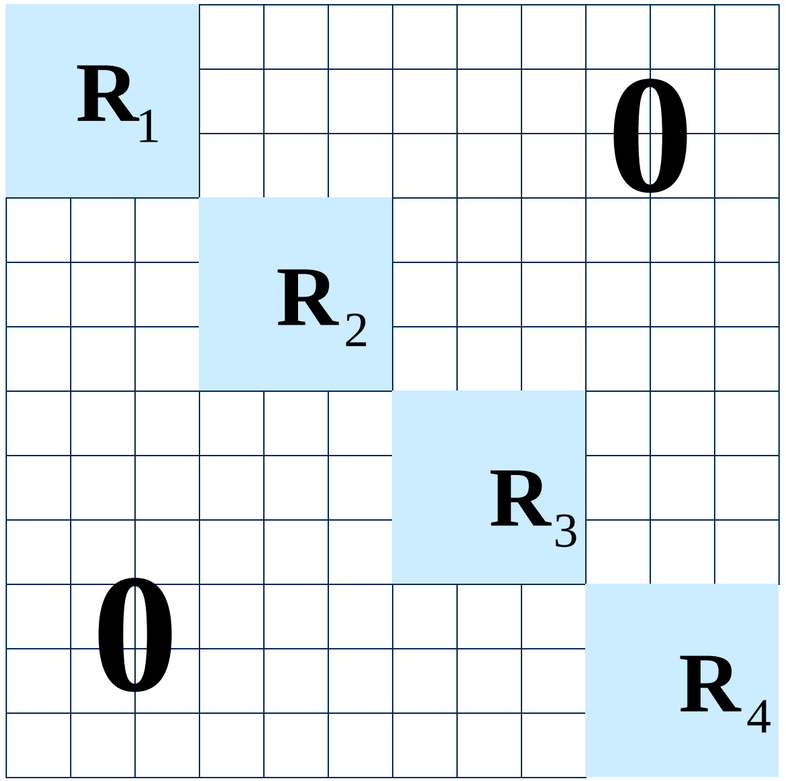
Example: CD4+ Cell Numbers Data Set

CD4+ Cell Numbers



Years since Seroconversion

Repeated Only Model – Has R, No G



$$\text{Var}(y) = V = ZGZ' + R$$
$$= 0 + R$$
$$= R$$

Choice of Covariance Structure for R

Variance Components

Time Point

1	2	3	4	
σ^2		0		1
	σ^2			2
		σ^2		3
0			σ^2	4

Default for both RANDOM and REPEATED statements

Compound Symmetry

Time Point

σ^2

1	2	3	4	
1.0	ρ	ρ	ρ	1
	1.0	ρ	ρ	2
		1.0	ρ	3
			1.0	4

Makes sense for non-longitudinal data

Choice of Covariance Structure for R

Unstructured

Time Point

1	2	3	4	
σ_1^2	σ_{12}	σ_{13}	σ_{14}	1
	σ_2^2	σ_{23}	σ_{24}	2
		σ_3^2	σ_{34}	3
			σ_4^2	4

Requires same time points on all subjects

First-Order Autoregressive AR(1)

Time Point

σ^2

1	2	3	4	
1.0	ρ	ρ^2	ρ^3	1
	1.0	ρ	ρ^2	2
		1.0	ρ	3
			1.0	4

Requires equally spaced and same time points on subjects

Choice of Covariance Structure for R

Spatial Power

Time Point

$$\sigma^2 \begin{array}{|c|c|c|c|c|} \hline 1 & 2 & 3 & 4 & \\ \hline 1.0 & \rho^{|t_1-t_2|} & \rho^{|t_1-t_3|} & \rho^{|t_1-t_4|} & 1 \\ \hline & 1.0 & \rho^{|t_2-t_3|} & \rho^{|t_2-t_4|} & 2 \\ \hline & & 1.0 & \rho^{|t_3-t_4|} & 3 \\ \hline & & & 1.0 & 4 \\ \hline \end{array}$$

Time Point

First to allow unequally spaced and different time points

Repeated Only Model – Has R, No G

```
proc mixed data=aids;
   model cd4_scale=time age cigarettes drug partners depression
         time*age time*depression time*partners time*drug
         time*cigarettes time*time time*time*time
         / solution ddfm=kr(firstorder);
   repeated / type=sp(pow)(time) local subject=id r=13 rcorr=13;
run;
```

Repeated Only Model – Has R, No G

Iteration History			
Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	12668.04910184	
1	3	11883.08815296	0.32992483
2	1	11881.79852820	0.00348677
3	2	11864.84042331	0.10490545
4	2	11801.90993395	2.88713335
5	2	11734.85393060	0.00204795
6	2	11731.57580732	0.00054912
7	1	11729.33587289	0.00001849
8	1	11729.26578521	0.00000003
9	1	11729.26567357	0.00000000

Convergence criteria met.

Dimensions	
Covariance Parameters	3
Columns in X	14
Columns in Z	0
Subjects	369
Max Obs per Subject	12

Repeated Only Model – Has R, No G

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Variance	id	7.8554
SP(POW)	id	0.8554
Residual		4.3300

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	8.0939	0.2434	1100	33.25	<.0001
time	-1.1385	0.1007	991	-11.30	<.0001
age	0.01736	0.01918	385	0.90	0.3661
cigarettes	0.4203	0.07447	1297	5.64	<.0001
drug	0.1522	0.2034	2331	0.75	0.4544
partners	0.04586	0.02291	2245	2.00	0.0454
depression	-0.02620	0.008670	2338	-3.02	0.0025
time*age	-0.01451	0.006072	617	-2.39	0.0172
time*depression	0.001513	0.003823	1644	0.40	0.6924
time*partners	-0.01312	0.01060	1790	-1.24	0.2161
time*drug	0.01618	0.08757	1616	0.18	0.8535
time*cigarettes	-0.1383	0.02984	1032	-4.63	<.0001
time*time	-0.1753	0.02758	966	-6.35	<.0001
time*time*time	0.06103	0.006930	1114	8.81	<.0001

Repeated Only Model – Has R, No G

Estimated R Matrix for Subject 13												
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10	Col11	Col12
1	12.1853	7.2673	6.7060	6.2226	5.7297	5.2850	4.8914	4.5252	4.1739	3.8615	3.5831	3.3050
2	7.2673	12.1853	7.2487	6.7261	6.1934	5.7126	5.2872	4.8914	4.5117	4.1739	3.8730	3.5724
3	6.7060	7.2487	12.1853	7.2891	6.7118	6.1908	5.7297	5.3008	4.8893	4.5233	4.1972	3.8714
4	6.2226	6.7261	7.2891	12.1853	7.2332	6.6717	6.1749	5.7126	5.2692	4.8747	4.5233	4.1722
5	5.7297	6.1934	6.7118	7.2332	12.1853	7.2456	6.7060	6.2040	5.7224	5.2940	4.9124	4.5310
6	5.2850	5.7126	6.1908	6.6717	7.2456	12.1853	7.2704	6.7261	6.2040	5.7396	5.3258	4.9124
7	4.8914	5.2872	5.7297	6.1749	6.7060	7.2704	12.1853	7.2673	6.7032	6.2013	5.7543	5.3076
8	4.5252	4.8914	5.3008	5.7126	6.2040	6.7261	7.2673	12.1853	7.2456	6.7032	6.2199	5.7371
9	4.1739	4.5117	4.8893	5.2692	5.7224	6.2040	6.7032	7.2456	12.1853	7.2673	6.7434	6.2199
10	3.8615	4.1739	4.5233	4.8747	5.2940	5.7396	6.2013	6.7032	7.2673	12.1853	7.2891	6.7233
11	3.5831	3.8730	4.1972	4.5233	4.9124	5.3258	5.7543	6.2199	6.7434	7.2891	12.1853	7.2456
12	3.3050	3.5724	3.8714	4.1722	4.5310	4.9124	5.3076	5.7371	6.2199	6.7233	7.2456	12.1853

Random Coefficient Model

$$y = X\beta + Z\gamma + \varepsilon$$

where β represents:

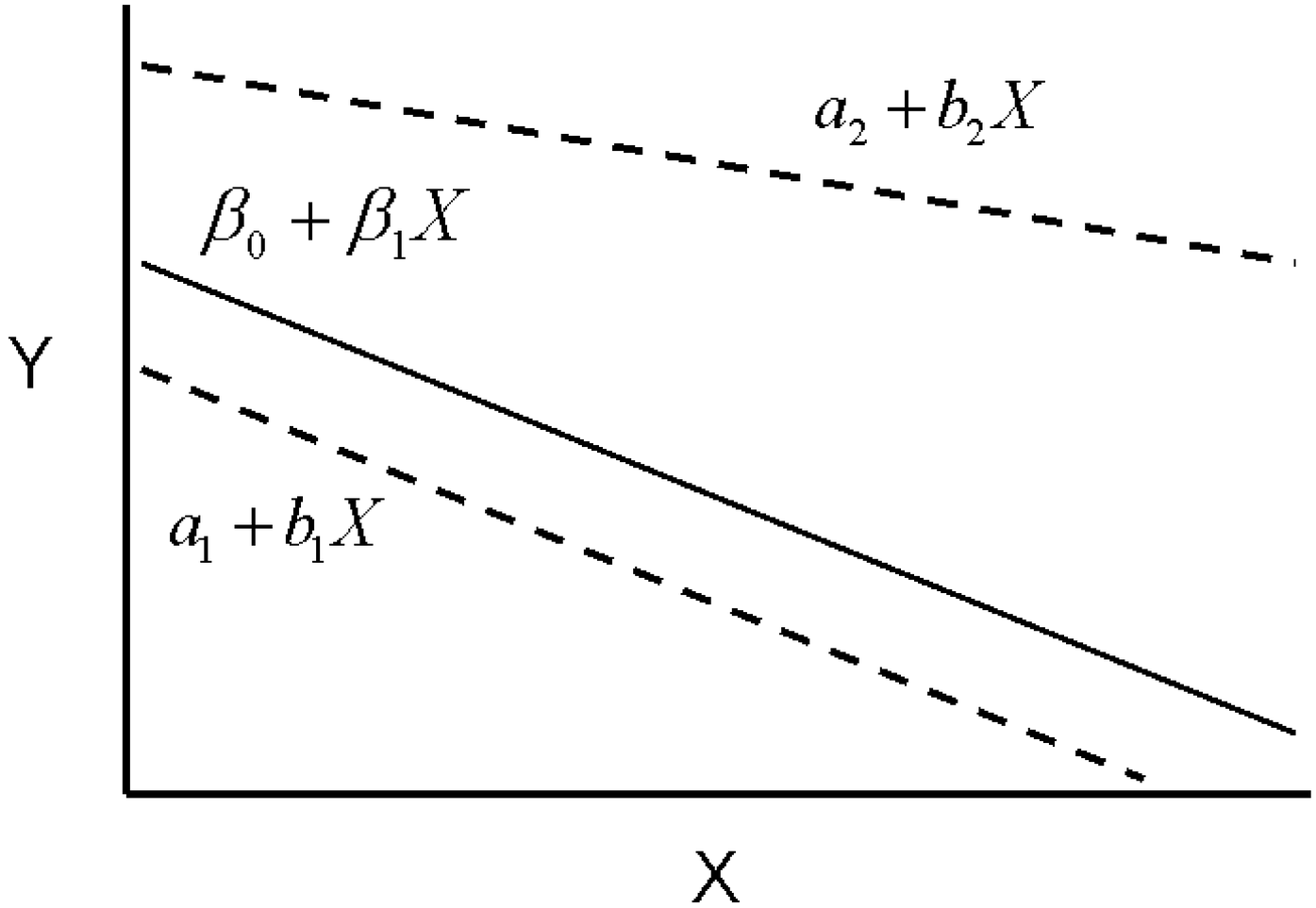
- the population average
- parameters that are assumed to be the same for all subjects

and where γ represents:

- parameters that are allowed to vary over subjects
- subject-specific regression coefficients that reflect the natural heterogeneity in the population

G specified in MIXED using RANDOM statement

Random Coefficient Model



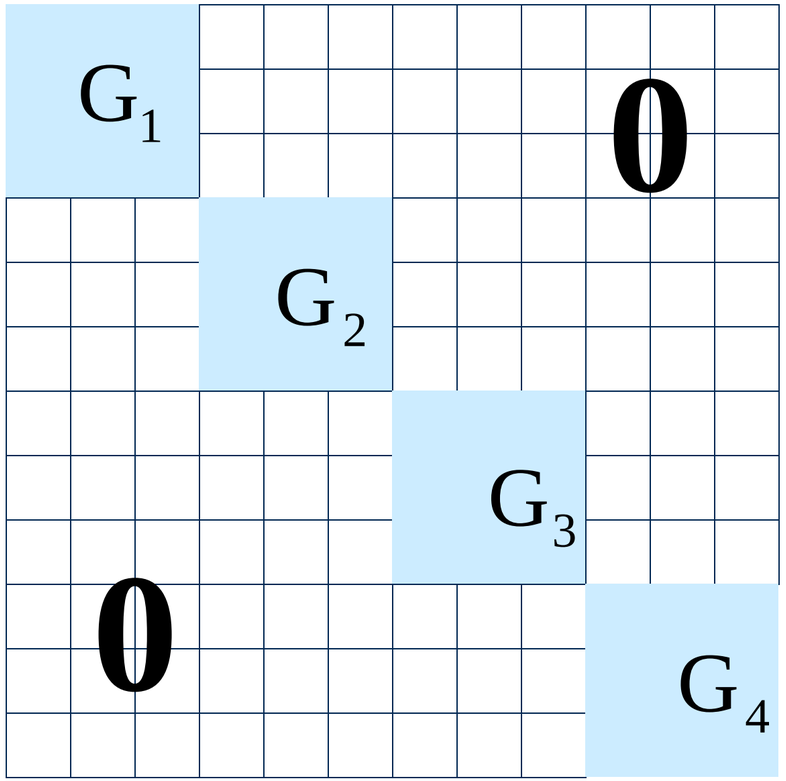
Random Coefficient Model

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + a_i^* + b_i^* x_{ij} + \varepsilon_{ij}$$

Population intercept and slope

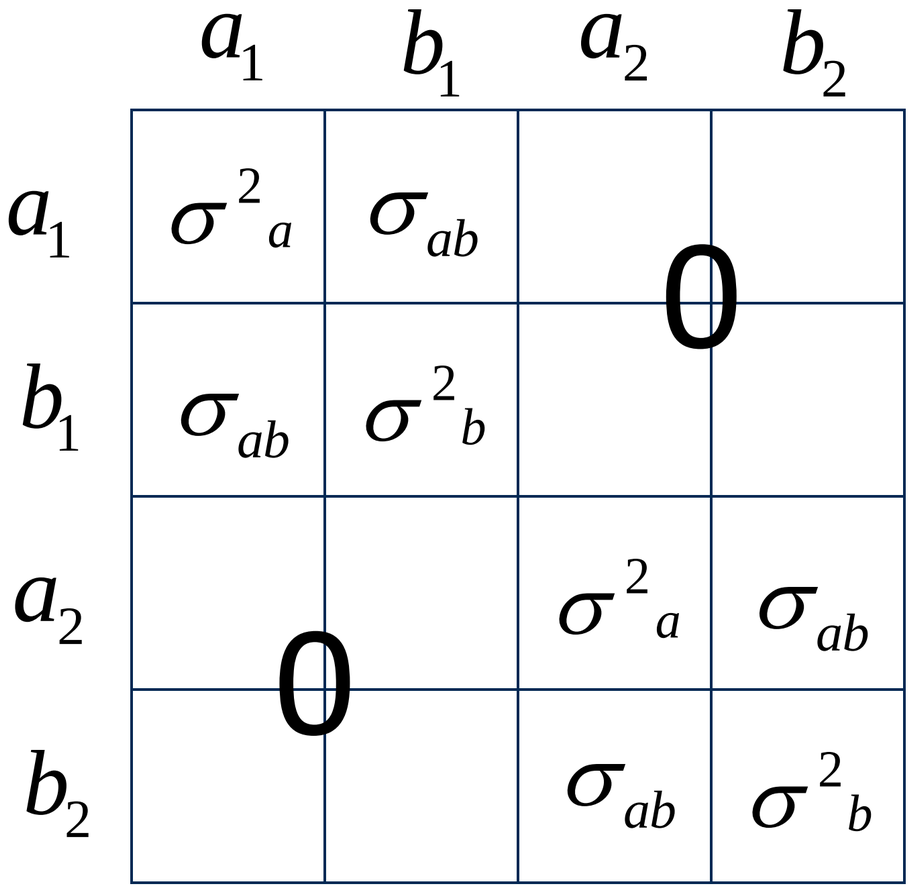
Subject-specific deviation of intercept and slope

Block-Diagonal Covariance Matrix



Unstructured Covariance Matrix with Two Random Effects

Unstructured is the only choice!



G specified in MIXED using RANDOM statement

Random Coefficient Model

```
proc mixed data=aids covtest;
    model cd4_scale=time age cigarettes drug partners depression
          time*age time*depression time*drug time*partners
          time*cigarettes time*time time*time*time / solution
          ddfm=kr;
    random intercept time time*time time*time*time / type=un subject=id
          g gcorr v=13 vcorr=13;
run;
```

Random Coefficient Model

Dimensions	
Covariance Parameters	11
Columns in X	14
Columns in Z per Subject	4
Subjects	369
Max Obs per Subject	12

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	8.1186	0.2469	859	32.88	<.0001
time	-1.1656	0.1072	523	-10.87	<.0001
age	0.01396	0.01951	327	0.72	0.4750
cigarettes	0.3640	0.07538	1001	4.83	<.0001
drug	0.1833	0.2063	1869	0.89	0.3744
partners	0.05915	0.02295	1940	2.58	0.0100
depression	-0.02706	0.008799	2004	-3.08	0.0021
time*age	-0.01401	0.006359	237	-2.20	0.0285
time*depression	0.002205	0.003901	688	0.57	0.5720
time*drug	0.007280	0.08934	699	0.08	0.9351
time*partners	-0.01518	0.01063	715	-1.43	0.1539
time*cigarettes	-0.1182	0.03107	506	-3.80	0.0002
time*time	-0.1770	0.02806	186	-6.31	<.0001
time*time*time	0.06031	0.007156	146	8.43	<.0001

Random Coefficient Model

Estimated G Matrix

Row	Effect	Subject	Col1	Col2	Col3	Col4
1	Intercept	1	7.1562	-1.0342	-0.2397	0.07635
2	time	1	-1.0342	0.8308	0.09795	-0.04079
3	time*time	1	-0.2397	0.09795	0.02921	-0.00973
4	time*time*time	1	0.07635	-0.04079	-0.00973	0.003274

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	id	7.1562	0.6927	10.33	<.0001
UN(2,1)	id	-1.0342	0.2428	-4.26	<.0001
UN(2,2)	id	0.8308	0.1507	5.51	<.0001
UN(3,1)	id	-0.2397	0.09042	-2.65	0.0080
UN(3,2)	id	0.09795	0.03897	2.51	0.0120
UN(3,3)	id	0.02921	0.01680	1.74	0.0411
UN(4,1)	id	0.07635	0.02241	3.41	0.0007
UN(4,2)	id	-0.04079	0.01253	-3.26	0.0011
UN(4,3)	id	-0.00973	0.004414	-2.20	0.0276
UN(4,4)	id	0.003274	0.001367	2.40	0.0083
Residual		4.9781	0.1831	27.18	<.0001

Random Coefficient Model

Estimated V Matrix for Subject 13												
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10	Col11	Col12
1	14.5368	9.1018	8.3506	7.4716	6.3868	5.2946	4.3027	3.4378	2.7642	2.4129	2.4242	2.9024
2	9.1018	13.7358	8.1322	7.3743	6.4232	5.4576	4.5785	3.8147	3.2285	2.9402	2.9887	3.4769
3	8.3506	8.1322	12.6479	7.0842	6.3338	5.5635	4.8591	4.2477	3.7836	3.5668	3.6310	4.0667
4	7.4716	7.3743	7.0842	11.6660	6.1635	5.6156	5.1102	4.6706	4.3393	4.1917	4.2545	4.5989
5	6.3868	6.4232	6.3338	6.1635	10.8896	5.6332	5.3689	5.1361	4.9617	4.8898	4.9382	5.1515
6	5.2946	5.4576	5.5635	5.6156	5.6332	10.6001	5.5967	5.5677	5.5459	5.5441	5.5707	5.6419
7	4.3027	4.5785	4.8591	5.1102	5.3689	5.5967	10.7624	5.9367	6.0494	6.1075	6.1108	6.0498
8	3.4378	3.8147	4.2477	4.6706	5.1361	5.5677	5.9367	11.2207	6.4688	6.5768	6.5592	6.3846
9	2.7642	3.2285	3.7836	4.3393	4.9617	5.5459	6.0494	6.4688	11.7569	6.9241	6.8924	6.6371
10	2.4129	2.9402	3.5668	4.1917	4.8898	5.5441	6.1075	6.5768	6.9241	12.0659	7.0550	6.7739
11	2.4242	2.9887	3.6310	4.2545	4.9382	5.5707	6.1108	6.5592	6.8924	7.0550	12.0148	6.7938
12	2.9024	3.4769	4.0667	4.5989	5.1515	5.6419	6.0498	6.3846	6.6371	6.7739	6.7938	11.6555

Generalized Linear Mixed Models

Generalized linear mixed models have the flexibility to model random effects and correlated errors for non-normal data.

- A linear predictor can contain random effects.
- The random effects are normally distributed.
- The conditional mean relates to the linear predictor through a link function:

$$g(\mathrm{E}(y \mid \gamma)) = \mathrm{X}\beta + \mathrm{Z}\gamma$$

- The conditional distribution (given γ) of the data belongs to the exponential family of distributions.

GLIMMIX Procedure

```
PROC GLIMMIX <options>;
      CLASS variables;
      CONTRAST 'label' contrast-specification </ options>;
      COVTEST <'label'> <test-specification> </ options>;
      EFFECT effect-name = effect-type (var-list < / effect-options >);
      ESTIMATE 'label' contrast-specification </ options>;
      LSMESTIMATE fixed-effect <'label'> values <divisor=n> </ options>;
      MODEL response <(response options)>=<fixed-effects> </ options>;
      NLOPTIONS <options>;
      OUTPUT <OUT=SAS-data-set> <keyword> </ options>;
      PARMS (value-list)...</ options>;
      RANDOM random-effects </ options>;
      WEIGHT variable;
      Programming statements...
RUN;
```

No REPEATED statement!!

GzLMM Formulation and PROC GLIMMIX

$$g(\mathrm{E}(y \mid \gamma)) = \mathbf{X}\beta + \mathbf{Z}\gamma$$

LINK= option

MODEL statement

RANDOM statement

LINK= and
DIST= on
MODEL statement

$Y|\gamma$ ~ exponential family

DIST= option

var(γ) = **G**

Options in the RANDOM statement

var(y| γ)=$A_\mu R A_\mu$

RANDOM _RESIDUAL_ statement

R-side Random Effects

```
proc glimmix data=sasuser.aids;
   model cd4_scale = time;
   random _residual_ / type=sp(pow)(time)
                subject=id;
run;
```

Analogous MIXED code:

```
proc mixed data=sasuser.aids
   model cd4_scale=time;
   repeated / type=sp(pow)(time) subject=id;
run;
```

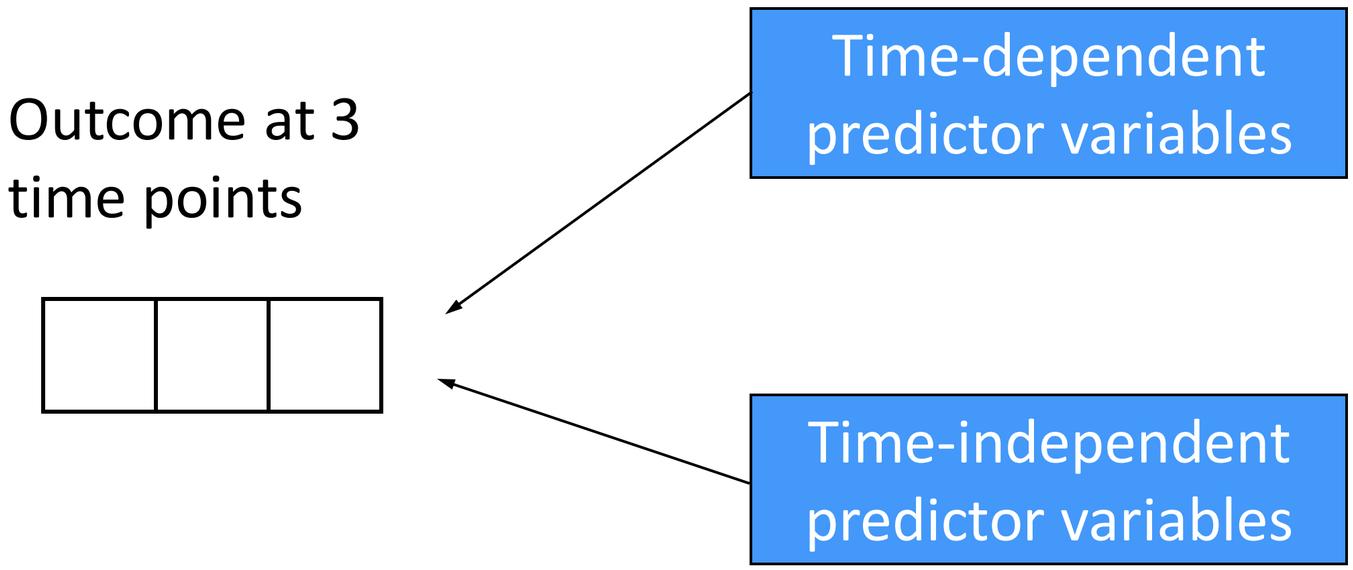
R-side Random Effects

```
data aids;
   set sasuser.aids;
   timec=time;
run;

proc glimmix data=aids noclprint;
   class timec;
   model cd4 scale=time;
   random timec / type=sp(pow)(time)
                  subject=id residual;
run;
```

Use this syntax if need to specify a repeated effect

Radial Keratotomy Study



Outcome:

Continuing effect of the surgery

1=Yes or 0=No

Predictor variables:

- Diameter of clear zone
  smaller means more serious surgery
- Age at baseline
- Gender

R-side Random Effects

```
proc glimmix data=long.keratotomy;
     class patientid gender;
     model unstable(event='1') = age diameter gender visit
              / solution ddfm=kr dist=binary
                or(diff=first
                   at visit diameter =1 4
                   units diameter = -1);
     random _residual_ / subject = patientid type=un;
     nloptions tech=nrridg;
run;
```

R-side Random Effects

The GLIMMIX Procedure

Model Information	
Data Set	LONG.KERATOTOMY
Response Variable	unstable
Response Distribution	Binary
Link Function	Logit
Variance Function	Default
Variance Matrix Blocked By	patientid
Estimation Technique	Residual PL
Degrees of Freedom Method	Kenward-Roger
Fixed Effects SE Adjustment	Kenward-Roger

Dimensions	
R-side Cov. Parameters	6
Columns in X	6
Columns in Z per Subject	0
Subjects (Blocks in V)	356
Max Obs per Subject	3

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
UN(1,1)	patientid	1.2278	0.09278
UN(2,1)	patientid	0.3029	0.05955
UN(2,2)	patientid	0.8653	0.06813
UN(3,1)	patientid	-0.08470	0.06373
UN(3,2)	patientid	0.2806	0.05617
UN(3,3)	patientid	1.1019	0.08384

R-side Random Effects

Solutions for Fixed Effects

Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		1.9131	0.9165	359	2.09	0.0375
age		0.01074	0.01227	362.1	0.88	0.3820
diameter		-1.2162	0.2284	363.4	-5.32	<.0001
gender	Female	-0.5671	0.1819	358.1	-3.12	0.0020
gender	Male	0	.	.	.	.
visit		0.3372	0.02321	372.8	14.53	<.0001

Fit Statistics

-2 Res Log Pseudo-Likelihood	4943.99
Generalized Chi-Square	1041.00
Gener. Chi-Square / DF	1.00

Odds Ratio Estimates

gender	age	diameter	visit	_gender	_age	_diameter	_visit	Estimate	DF	95% Confidence Limits	
	34.964	4	1		33.964	4	1	1.011	362.1	0.987	1.035
	33.964	3	1		33.964	4	1	3.374	363.4	2.153	5.288
	33.964	4	2		33.964	4	1	1.401	372.8	1.339	1.466
Male	33.964	4	1	Female	33.964	4	1	1.763	358.1	1.233	2.521

G-side Random Effects

```
proc glimmix data=long.keratotomy method=quad;
    class patientid gender;
    model unstable(event='1') = age diameter gender visit
          / solution dist=binary ddfm=bw;
    random intercept / subject = patientid;
    nloptions tech=nrridg;
    covtest "H0: No random effects" zerog;
run;
```

G-side Random Effects

Model Information	
Data Set	LONG.KERATOTOMY
Response Variable	unstable
Response Distribution	Binary
Link Function	Logit
Variance Function	Default
Variance Matrix Blocked By	patientid
Estimation Technique	Maximum Likelihood
Likelihood Approximation	Gauss-Hermite Quadrature
Degrees of Freedom Method	Between-Within

Dimensions	
G-side Cov. Parameters	1
Columns in X	6
Columns in Z per Subject	1
Subjects (Blocks in V)	356
Max Obs per Subject	3

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	patientid	1.5405	0.4752

G-side Random Effects

Solutions for Fixed Effects

Effect	gender	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		2.2635	1.0983	352	2.06	0.0400
age		0.01508	0.01475	352	1.02	0.3072
diameter		-1.4759	0.2787	352	-5.29	<.0001
gender	Female	-0.6843	0.2192	352	-3.12	0.0019
gender	Male	0	.	.	.	.
visit		0.4044	0.03232	689	12.51	<.0001

Fit Statistics

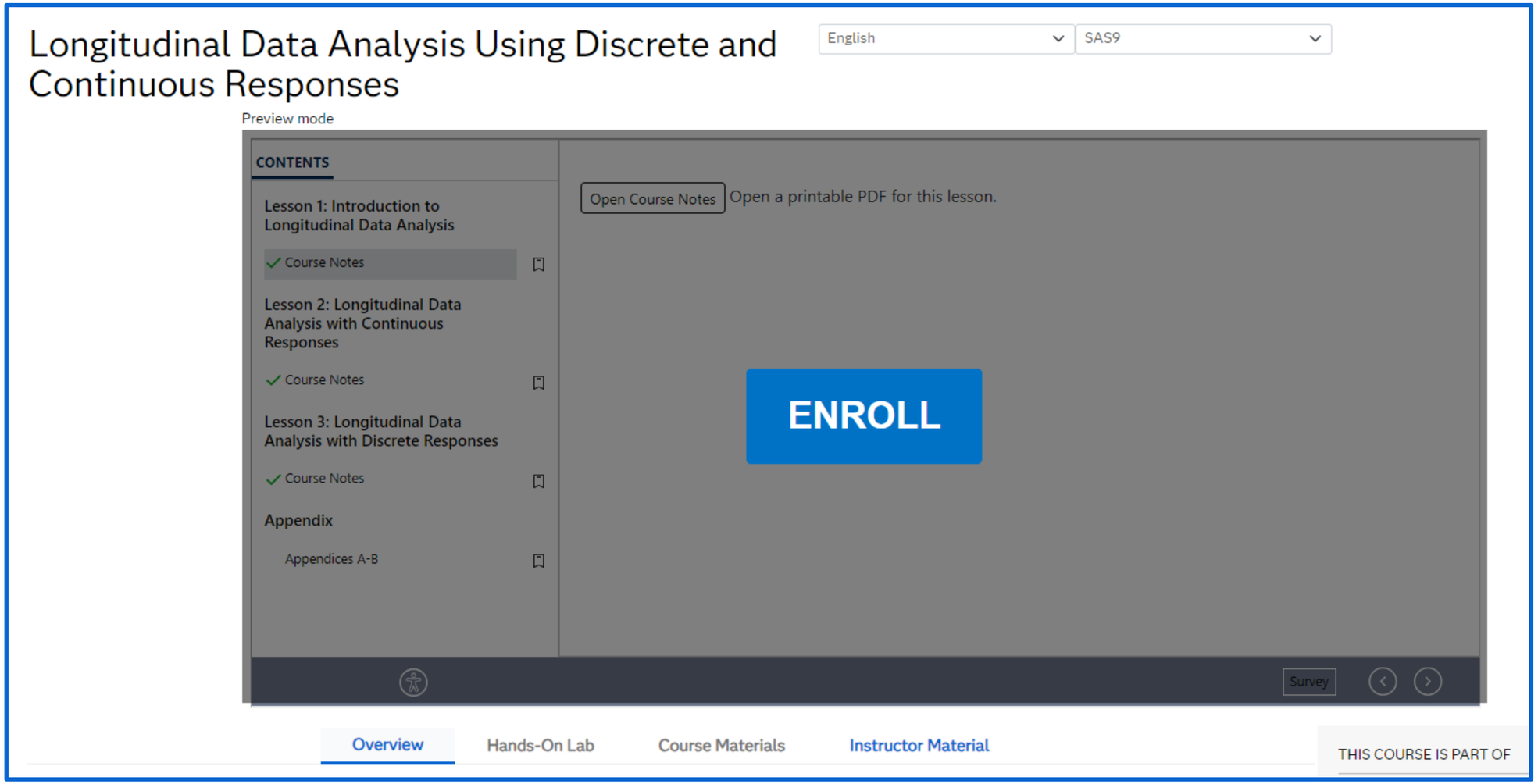
-2 Log Likelihood	1049.37
AIC (smaller is better)	1061.37
AICC (smaller is better)	1061.45
BIC (smaller is better)	1084.62
CAIC (smaller is better)	1090.62
HQIC (smaller is better)	1070.61

Tests of Covariance Parameters Based on the Likelihood

Label	DF	-2 Log Like	ChiSq	Pr > ChiSq	Note
H0: No random effects	1	1074.75	25.38	<.0001	MI

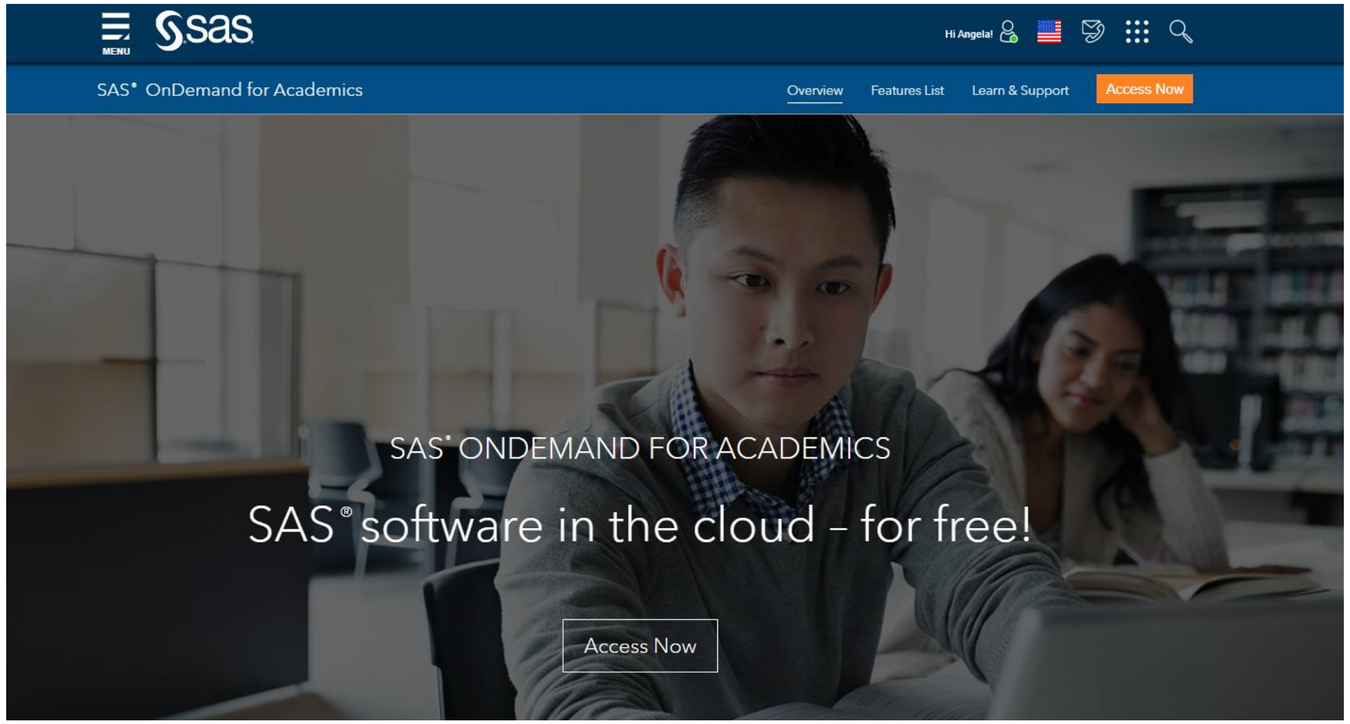
Continue Your Learning with a SAS Course

https://learn.sas.com/course/view.php?id=430 Free 7-day trial!



Continue Your Learning with SAS Programming

SAS OnDemand for Academics



- Free SAS software for students, educators, and independent learners.
- Register at: www.sas.com/ondemand
- Launch at: welcome.oda.sas.com

**Down for system maintenance May 14 – May 15

Questions?

Thank you!

Jacqueline.Johnson@sas.com