Longitudinal Data Analysis

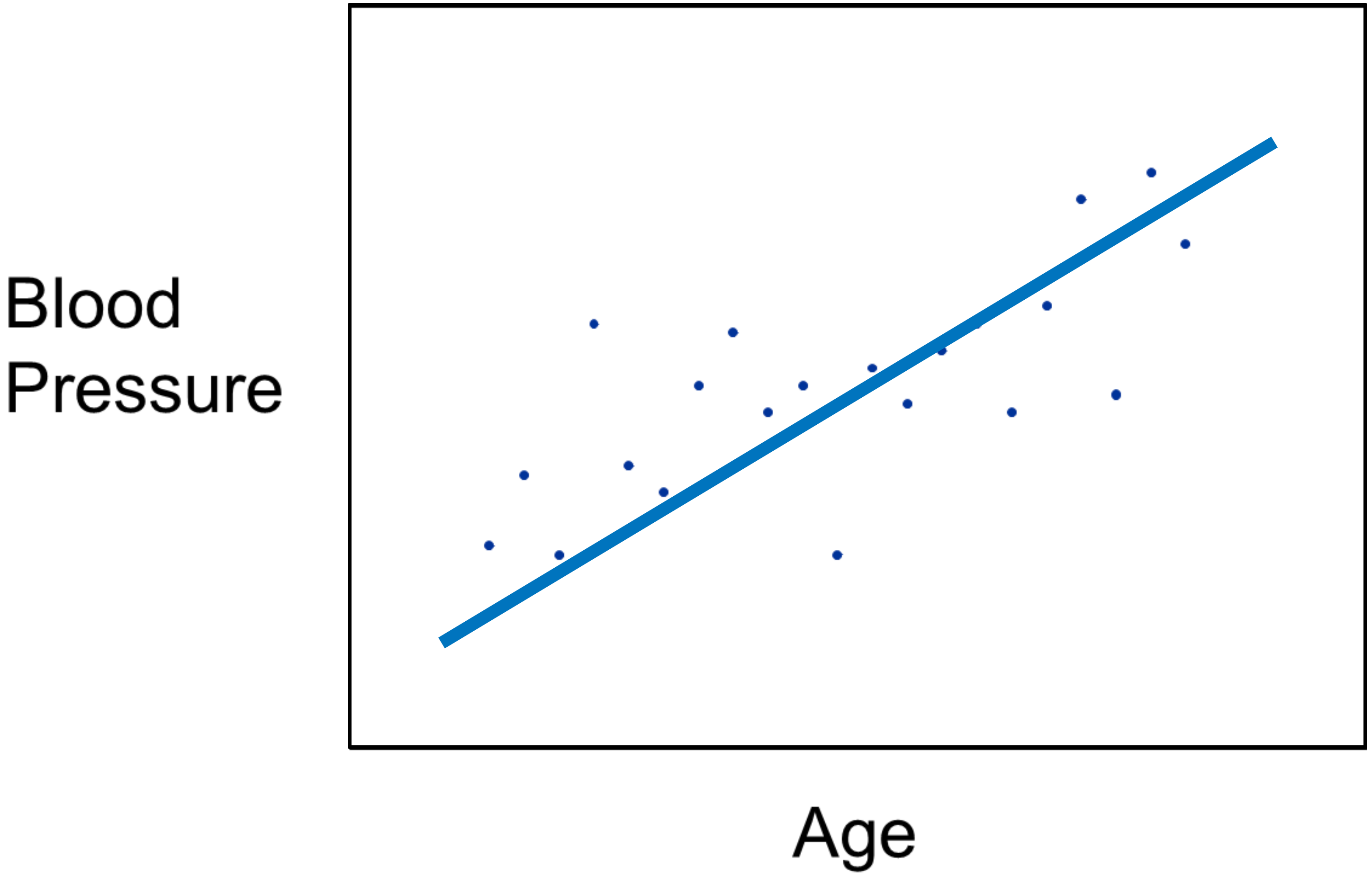
Western Kentucky University
June 22, 2023

Jackie Johnson
Jacqueline.Johnson@sas.com

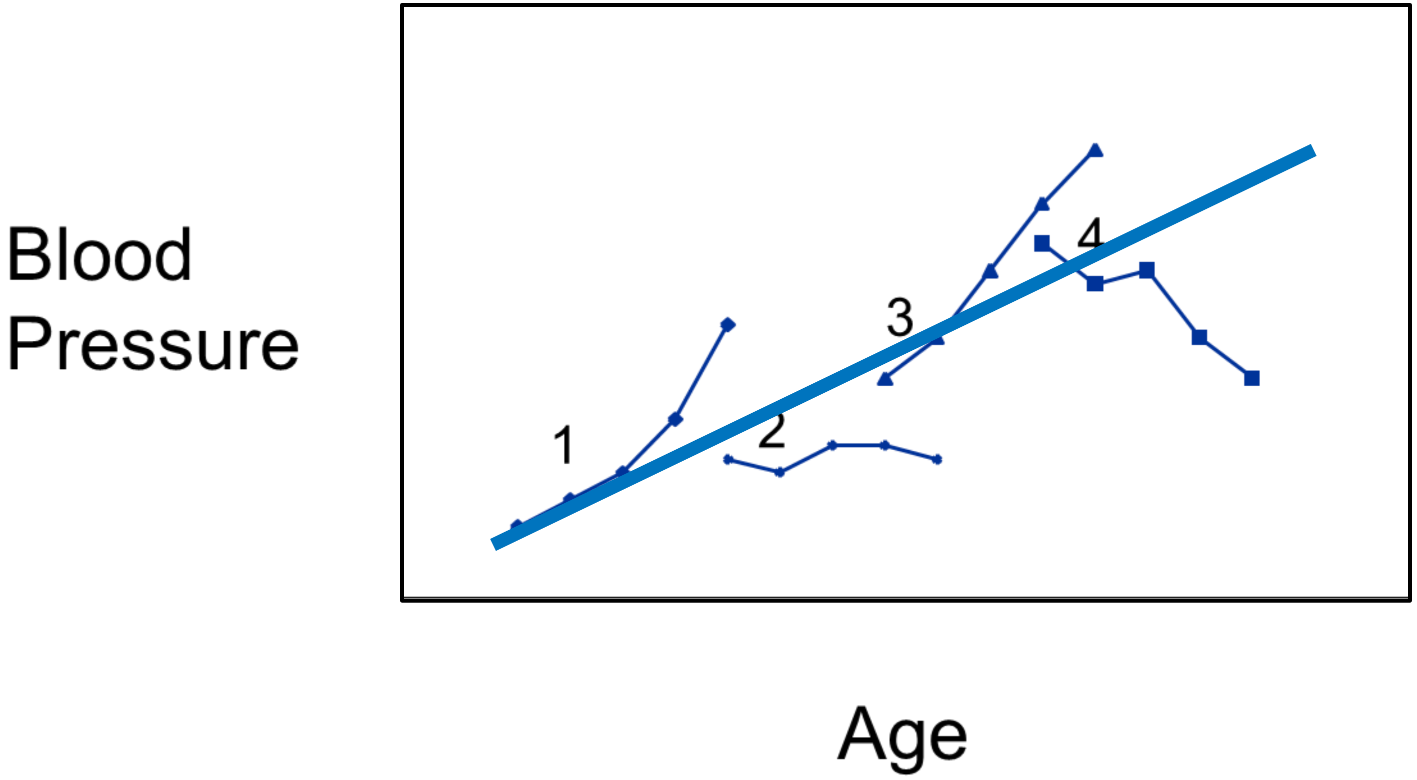
Longitudinal Data Analysis

- The defining feature is that repeated measurements are taken on a subject through time.
- The models can distinguish changes over time within subjects from differences between subjects at their baseline levels and can also study changes over time between groups.
- The models can estimate individual-level (subject-specific) regression parameters and population-level regression parameters.

Cross-Sectional Analysis



Longitudinal Analysis



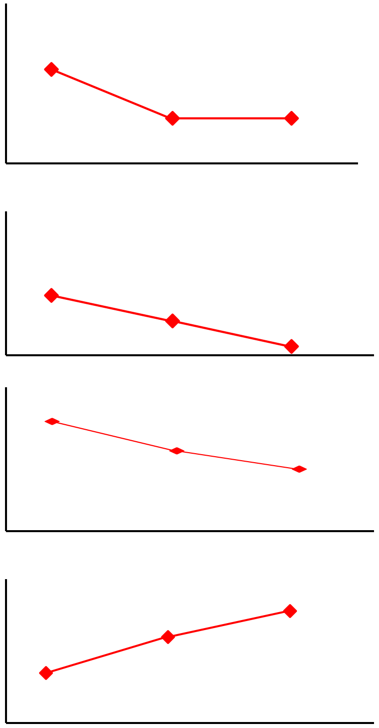
Variance-Covariance Matrix for OLS Regression

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

σ^2			**0**
	σ^2		
		σ^2	
0			σ^2

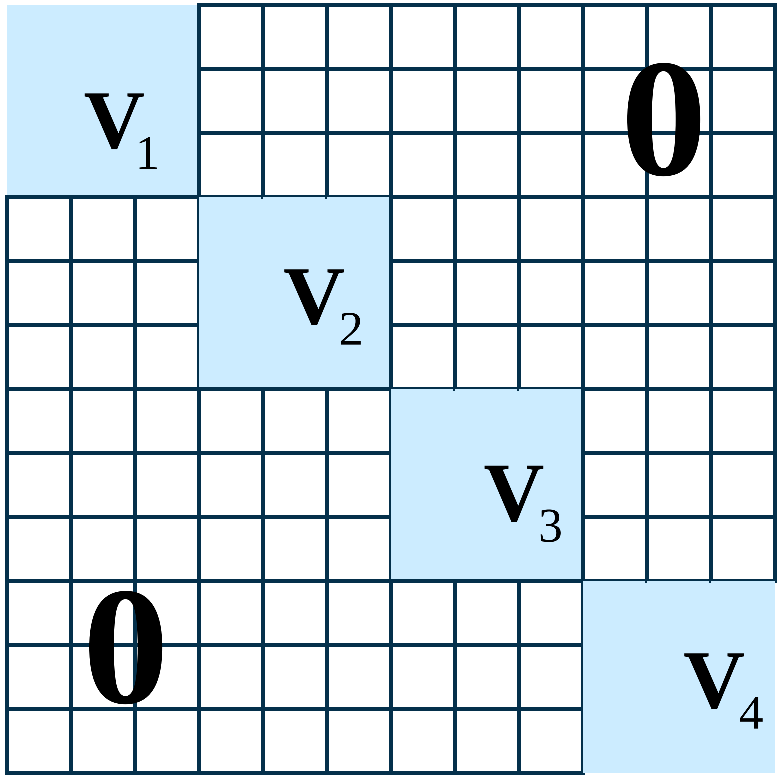
Longitudinal Data

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



General Linear Model

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

where y is the vector of observed responses

X is the design matrix of predictor variables

β is the vector of regression parameters

ε is the vector of random errors.

Assumes: e ~ N(0, sigmasq*I)

General Linear Mixed Model

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

where Z is the design matrix of random variables

γ is the vector of random-effect parameters — Covariance matrix= G

ε is no longer required to be independent and homogeneous. — Covariance matrix= R

Both gamma and epsilon contribute to covariance matrix

MIXED Procedure in SAS

General form of the MIXED procedure:

```
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

Model Assumptions in PROC MIXED

- Random effects and error terms are normally distributed with means of 0.
- Random effects and error terms are independent of each other.
- The relationship between the response variable and predictor variables is linear.

Gamma ~ N (0, G)

Epsilon ~ N (0, R)

y ~ N(XB, V)

V = ZGZ' + R

Gamma and Epsilon independent

Estimation in Mixed Models for Fixed Effects

Estimated generalized least squares (EGLS)

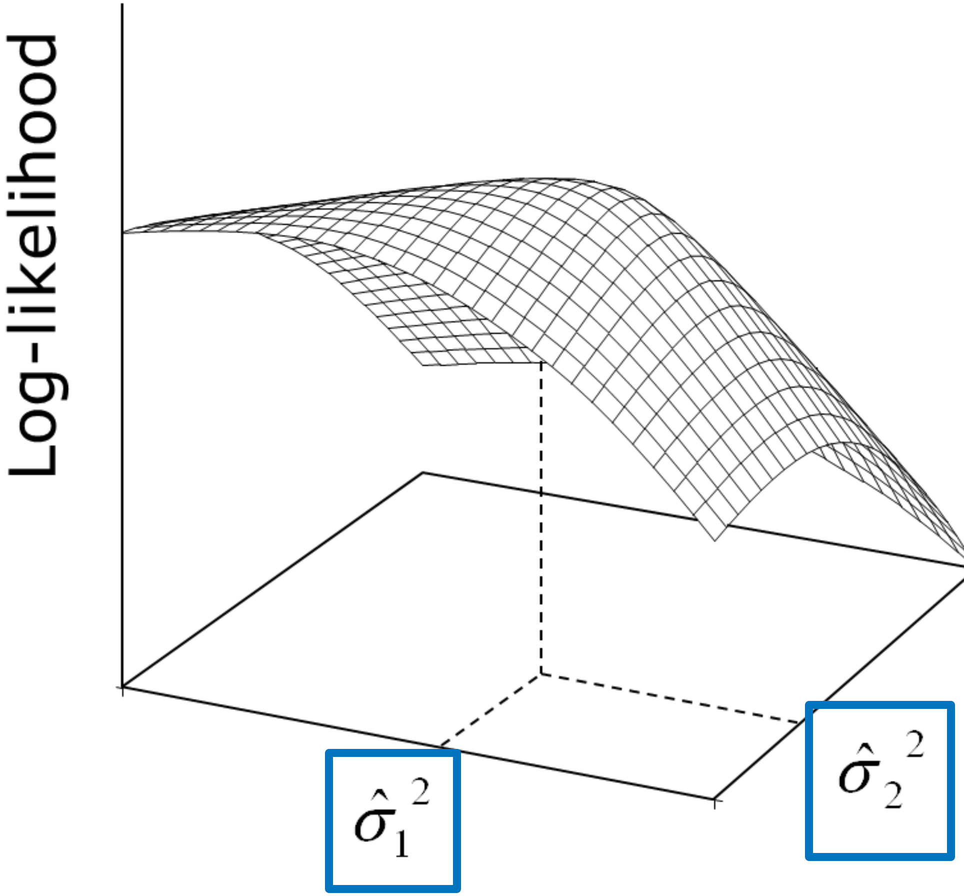
- takes into account the covariance structures **G** and **R**
- requires a reasonable estimate of **G** and **R**
- is the solution for fixed effects.

The formula for EGLS is

$$\hat{\beta} = (\mathrm{X}'\hat{\mathrm{V}}^{-1}\mathrm{X})^{-}\mathrm{X}'\hat{\mathrm{V}}^{-1}Y$$

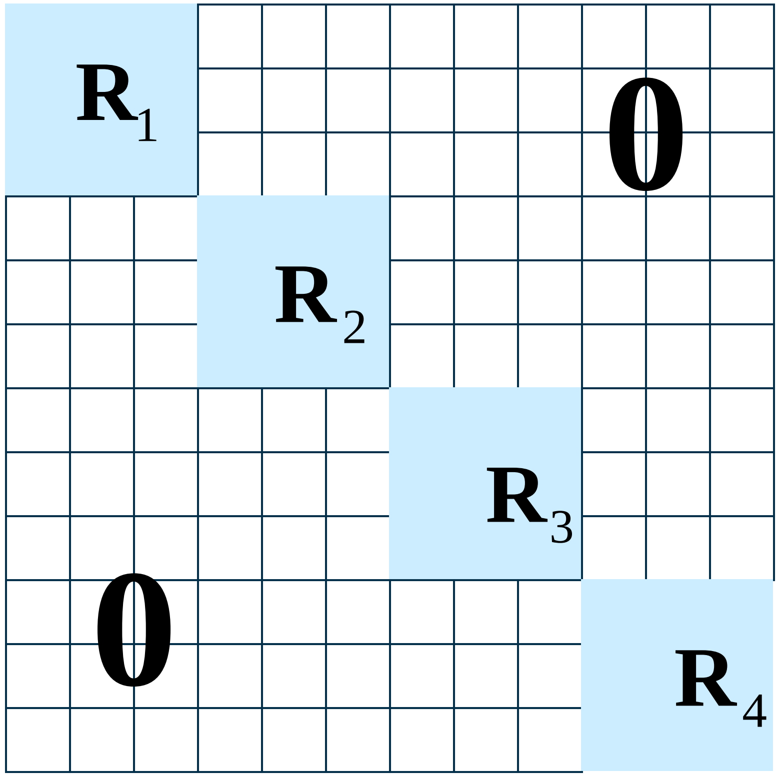
where $\hat{V} = \mathrm{Z}\hat{\mathrm{G}}\mathrm{Z}' + \hat{R}$

Restricted Maximum Likelihood Estimation



Likelihood = probability density function for your data

Block-Diagonal Covariance Matrix



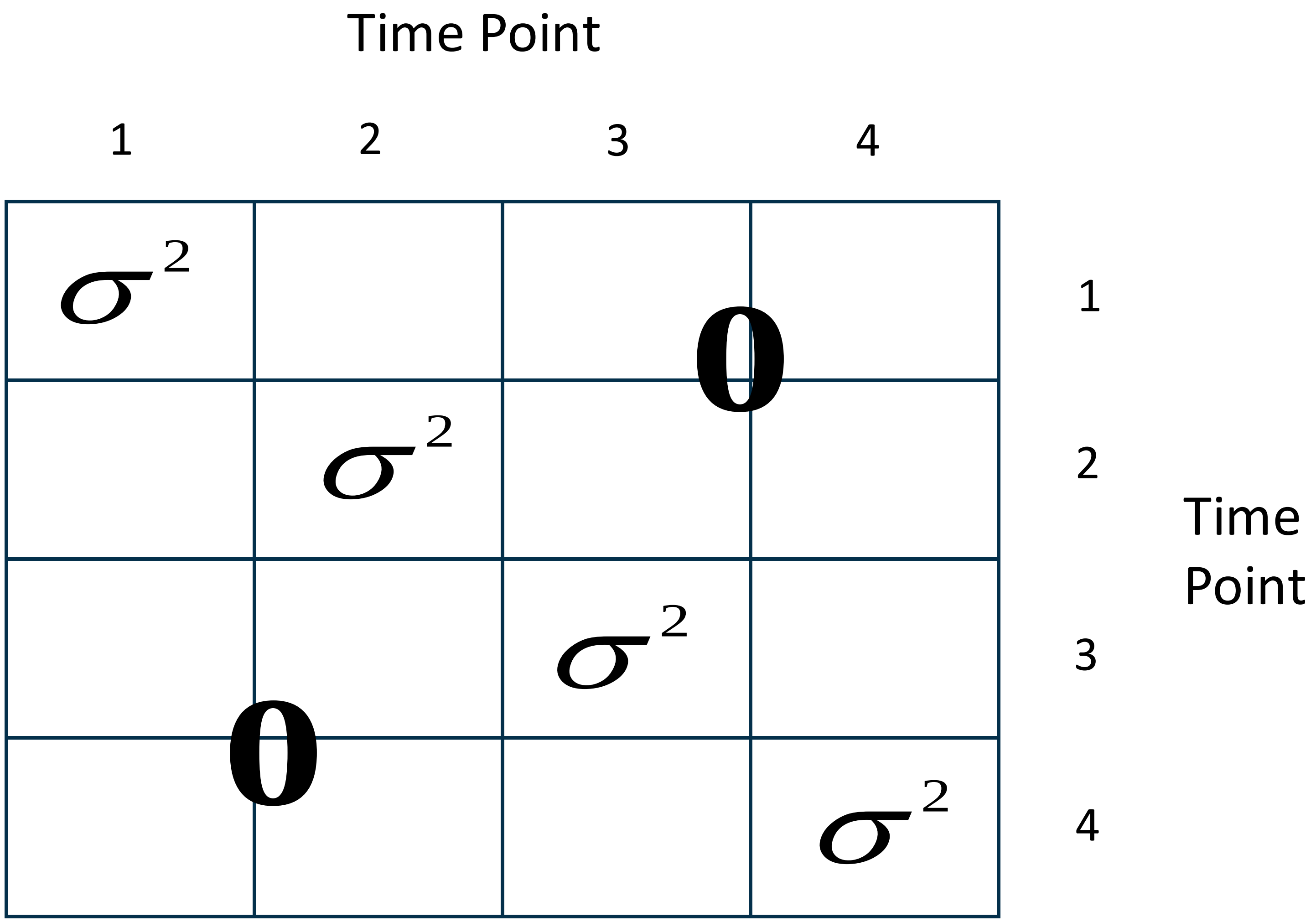
This discussion is about modeling the R matrix, not G

Selecting the Appropriate Covariance Structure

When finding reasonable estimates for **R**,

- if you choose a structure that is too simple, then you risk increasing the Type I error rate
- if you choose a structure that is too complex, then you sacrifice power and efficiency.

Variance Component (VC) or Simple



Default for both RANDOM and REPEATED statements

Compound Symmetry

Equal correlations between all time points

Time Point

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

Time Point

Makes sense for non-longitudinal data

Unstructured Covariance

T*(T+1)/2 parms estimated

Time Point

1	2	3	4	Time Point
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

Time Point

Requires equally spaced and same time points on subjects

Spatial Power

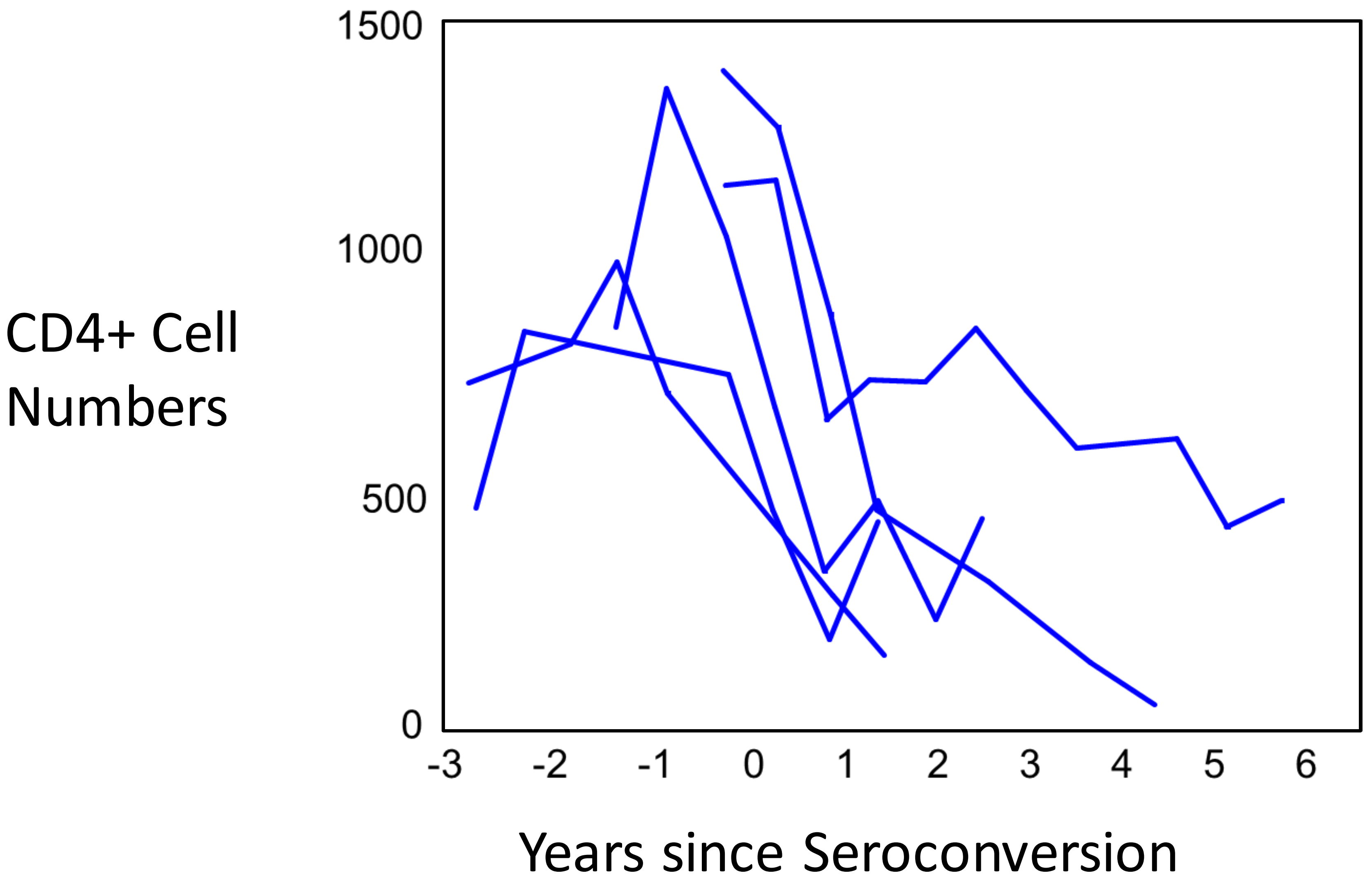
Time Point

σ^2 ×

	1	2	3	4	Time Point
	1.0	$\rho^{\lvert t_1 - t_2 \rvert}$	$\rho^{\lvert t_1 - t_3 \rvert}$	$\rho^{\lvert t_1 - t_4 \rvert}$	1
		1.0	$\rho^{\lvert t_2 - t_3 \rvert}$	$\rho^{\lvert t_2 - t_4 \rvert}$	2
			1.0	$\rho^{\lvert t_3 - t_4 \rvert}$	3
				1.0	4

First to allow unequally spaced and different time points

CD4+ Cell Numbers Data Set



Objectives of CD4+ Cell Numbers Study

- Estimate the average time course of CD4+ cell depletion.
- Estimate the time course for individual men.
- Characterize the degree of heterogeneity across men in the rate of progression.
- Identify factors that predict CD4+ cell changes.

Fitting a Longitudinal Model in PROC MIXED

This demonstration illustrates the concepts discussed previously.

Models with Only the REPEATED Statement

- No random effects are included in the model.
- Covariance structure for the data is completely determined by the covariance structure for the residual error.
- There is an **R** matrix but no **G** matrix.
- Usually, the model of choice when the longitudinal data are obtained at fixed points in time and when the within-subject correlations are adequately modeled using a specified covariance structure.

$$G=0$$
$$V = ZGZ' + R$$

Random Coefficient Models

- Random effects representing natural heterogeneity between subjects are used to describe the covariance structure of the data.
- The regression coefficients for one or more covariates are assumed to be a random sample from some population of possible coefficients.
- There is an **R** and a **G** matrix.   R = sigmasq * I   V = ZGZ' + R
- These models are useful for highly unbalanced data with many repeated measurements per subject.
- Usually, these are the model of choice when the longitudinal data are not obtained at fixed points in time and the within-subject correlations are not adequately modeled by a specified covariance structure.

Random Coefficient Model

$$y = X\beta + \boxed{Z\gamma} + \varepsilon$$

where β represents:

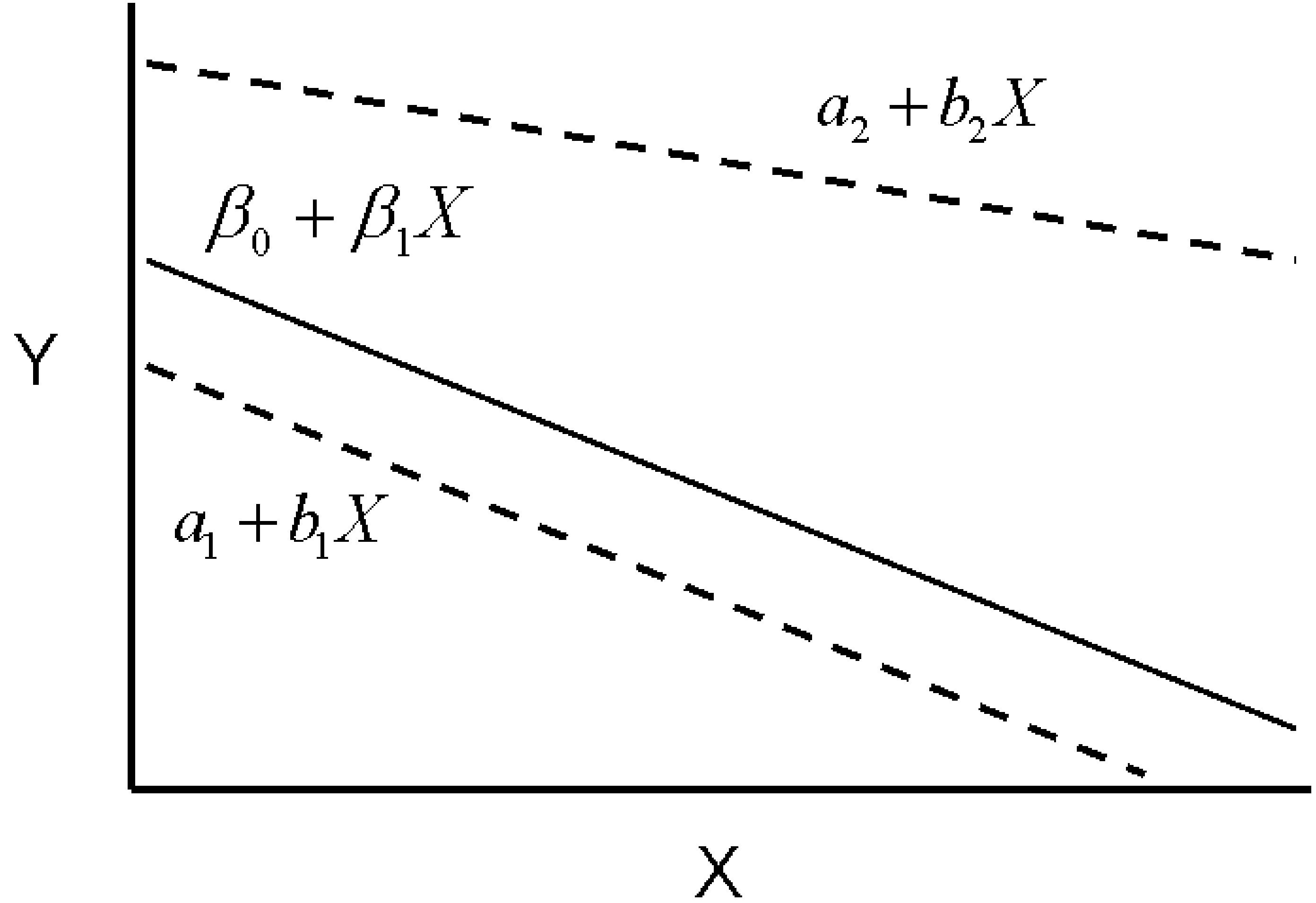
- the population average
- parameters that are assumed to be the same for all subjects

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



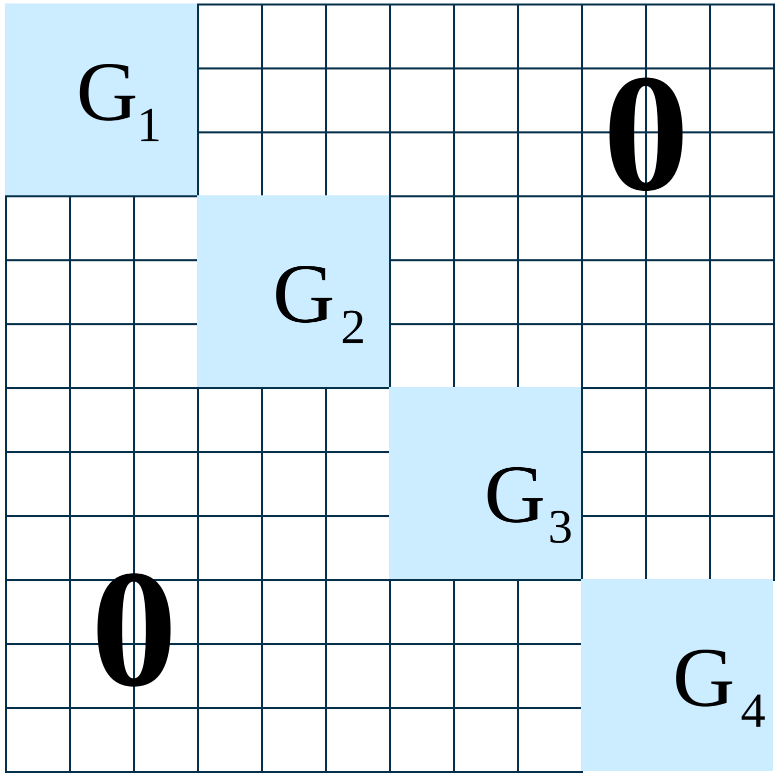
Random Coefficient Model

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

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!

	a_1	b_1	a_2	b_2
a_1	σ^2_a	σ_{ab}	0	
b_1	σ_{ab}	σ^2_b		
a_2	0		σ^2_a	σ_{ab}
b_2			σ_{ab}	σ^2_b

G specified in MIXED using RANDOM statement

Random Coefficient Models

This demonstration illustrates the concepts discussed previously.

Resources for Teaching and Learning

SAS Software for Learners

SAS OnDemand For Academics

- Available for free to educators, students, and independent learners!
- Access SAS 9.4 software through a web site.
- Submit code through the SAS Studio programming environment.
- Use SAS Studio point and click tasks to generate SAS code.
    - Useful for teaching students new to programming!
- No installation is necessary.
- Streamlined setup-- software works the same regardless of operating system!

Register for OnDemand for Academics:
http://www.sas.com/ondemand

SAS Software for Learners

SAS OnDemand For Academics

- This is full SAS, not a limited educational version!
- Access SAS software including these products:
  - SAS/STAT® – State-of-the-art statistical software.
  - Base SAS® – Flexible, extensible fourth-generation programming language for data access, transformation and reporting.
  - SAS/IML® – Interactive matrix programming
  - SAS/OR® – Enhanced operations research methods
  - SAS/QC® – Tools for statistical quality improvement.
  - SAS/ETS® – Model, forecast and simulate processes with econometric and time series analysis.

Getting Started with SAS OnDemand for Academics

https://youtu.be/tmL8fdOd-pI

Educators Come Join Us in Cary This August!

https://tinyurl.com/SASEduConf2023

Internship Opportunities at SAS

Keslie Carrion

Accessibility Designer Intern

“It's been wonderful to be a part of it all. I wouldn't have believed it was possible for a company to be like this if I didn't have that firsthand experience!”

- Technical & Analytic Internships
- Business Internships
- United in STEM: SAS Diversity Internship Program

Check out sas.com/jobs for details!

Outreach
Ongoing

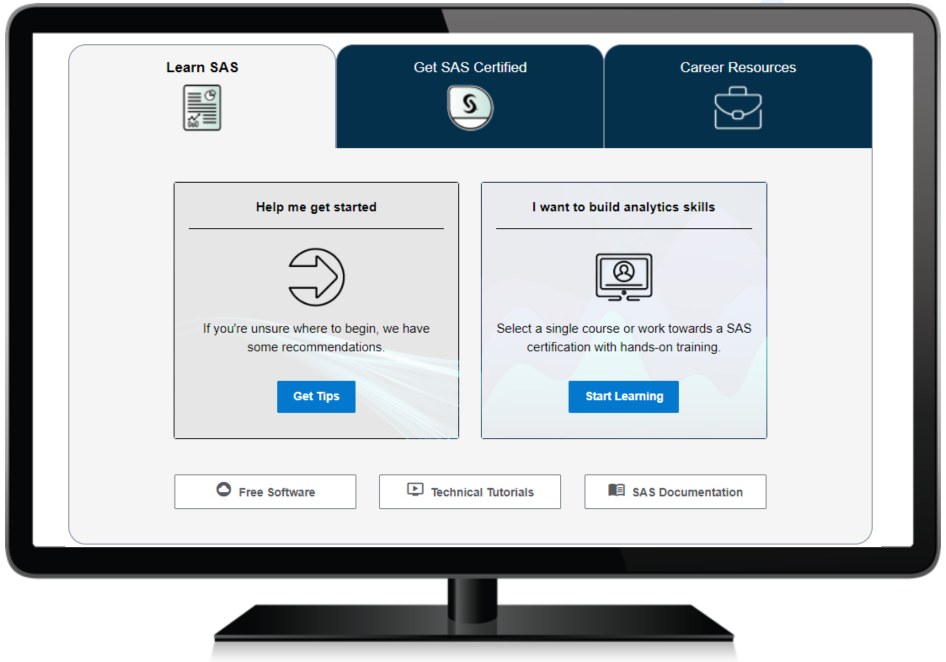
Apply
Open in September

Interview
October - April

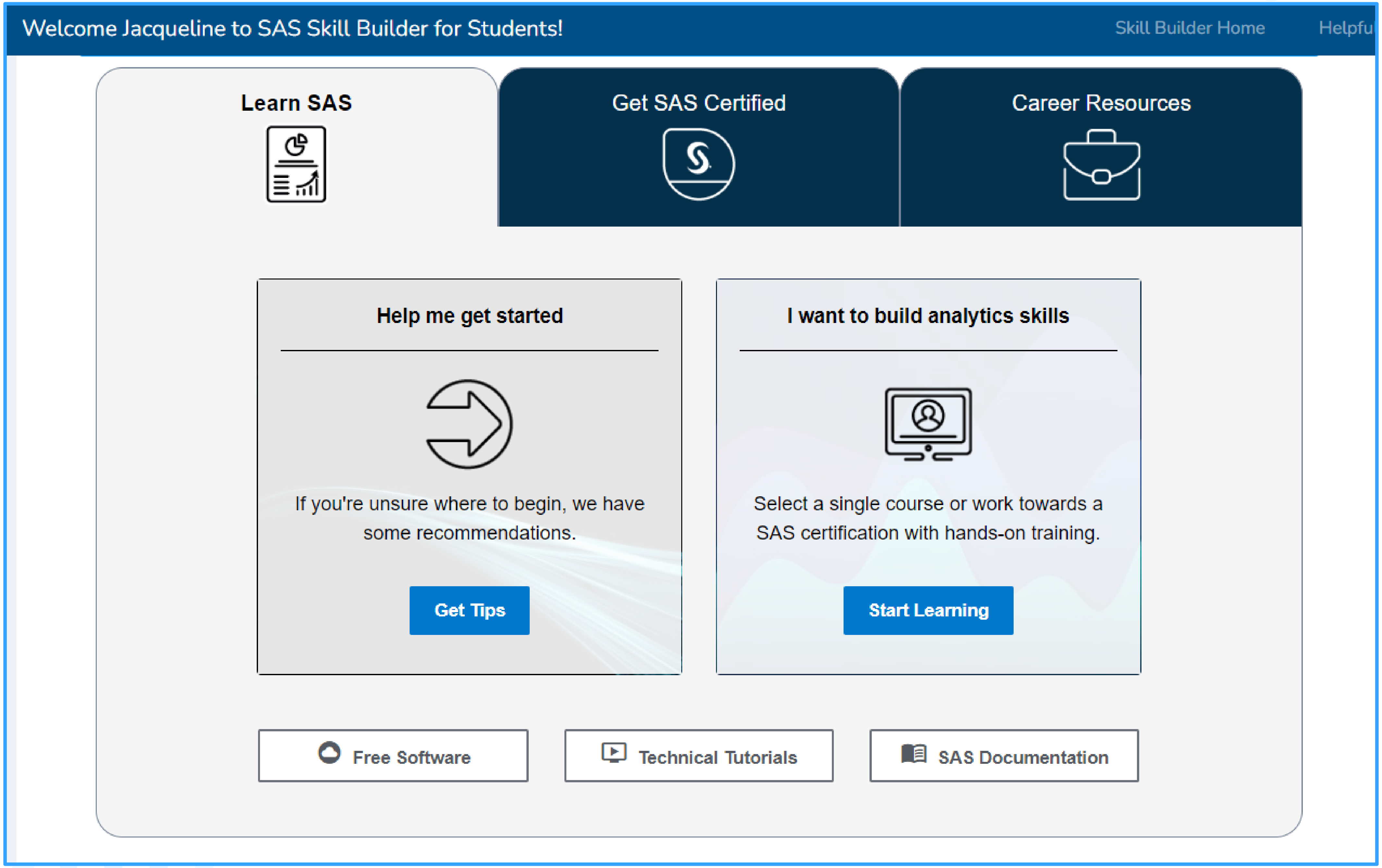
You’re In
May & June

SAS Skill Builder for Students

- Learn data science concepts on your time
  - 20+ free comprehensive e-learning courses
  - Nine certification pathways with prep materials
- Stay motivated with career resources
  - Discover why data analytics is a rewarding career path
  - Learn about what it means to be a data analytics professional



Register for SAS Skill Builder for Students using your academic email address: https://www.sas.com/skillbuilder


Welcome Jacqueline to SAS Skill Builder for Students!
Skill Builder Home
Helpfu
Learn SAS
Get SAS Certified
Career Resources
Help me get started
If you're unsure where to begin, we have some recommendations.
Get Tips
I want to build analytics skills
Select a single course or work towards a SAS certification with hands-on training.
Start Learning
Free Software
Technical Tutorials
SAS Documentation

Welcome Jacqueline to SAS Skill Builder for Students!

Skill Builder Home | Help

E-Learning Courses and Certification Pathways

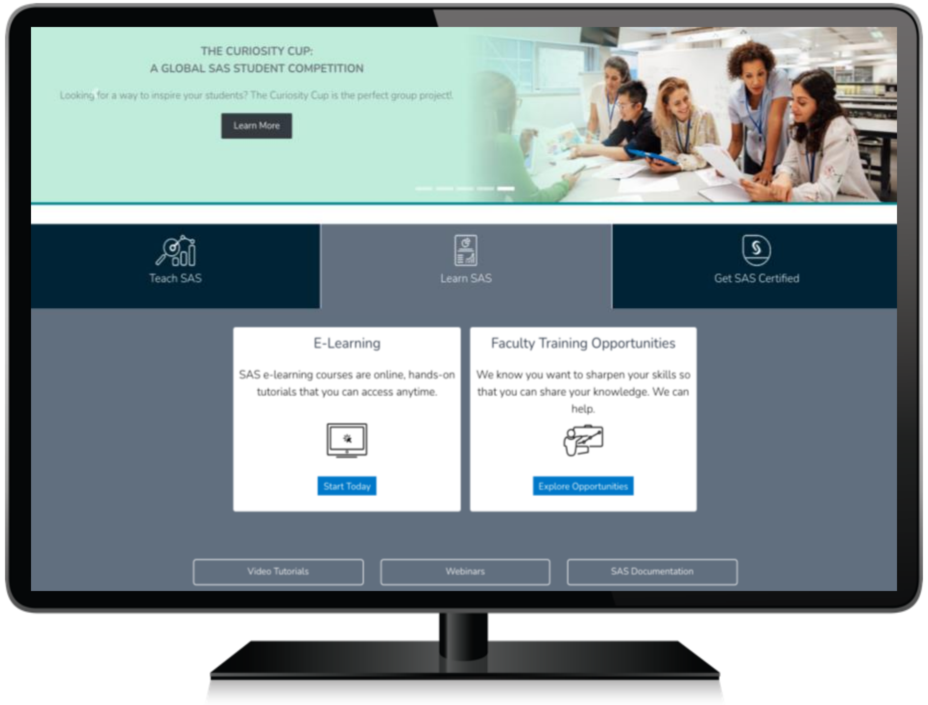
Back to Learn SAS

Expand All

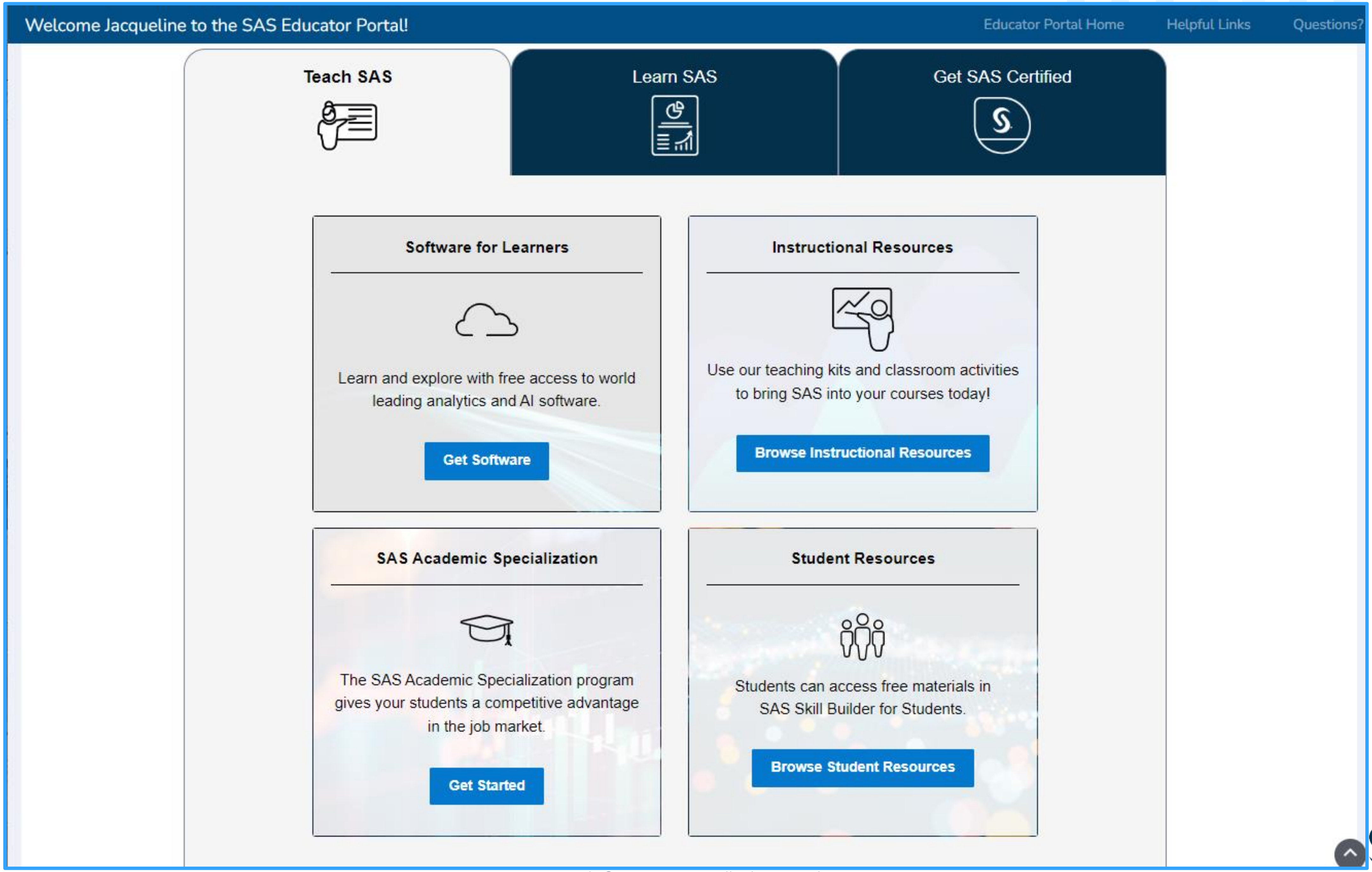
- ⊕ Visual Analytics and Visual Statistics
- ⊕ Programming
- ⊕ Statistical Analysis, Predictive Modeling, and Machine Learning
- ⊕ Foundational Skills
- ⊕ JMP Statistical Analysis

SAS Educator Portal

- Instructional materials for data science, machine learning, statistics, and more!
  - 100+ teaching kits
  - Industry-specific classroom activities
- E-learning and certification paths
- Educator workshops and training
- Discounts and promotions



Register for the SAS Educator Portal using your academic email address:
https://www.sas.com/teach


Welcome Jacqueline to the SAS Educator Portal!
Educator Portal Home
Helpful Links
Questions?
Teach SAS
Learn SAS
Get SAS Certified
Software for Learners
Learn and explore with free access to world leading analytics and AI software.
Get Software
Instructional Resources
Use our teaching kits and classroom activities to bring SAS into your courses today!
Browse Instructional Resources
SAS Academic Specialization
The SAS Academic Specialization program gives your students a competitive advantage in the job market.
Get Started
Student Resources
Students can access free materials in SAS Skill Builder for Students.
Browse Student Resources

SAS Educator Portal

Full course notes are available for a 3-day longitudinal data analysis course

Statistical Analysis

Course Title	OnDemand for Academics	SAS Viya for Learners	University Licensed Software	Software Version
Longitudinal Data Analysis With Discrete & Continuous Responses	✔		✔	SAS/STAT on SAS 9.4
Manipulating Data and Analytics Using SAS University Edition	✔		✔	SAS/SAS on SAS 9.4, SAS Studio
Mixed Models Analyses Using SAS	✔		✔	SAS/STAT on SAS 9.4
Multilevel Modeling of Hierarchical and Longitudinal Data Using SAS	✔		✔	SAS/STAT on SAS 9.4

Questions?

Jacqueline.Johnson@sas.com
Lori.Downen@sas.com