An overview of longitudinal data analysis (LDA)

Dr. Mahmoud Danaee

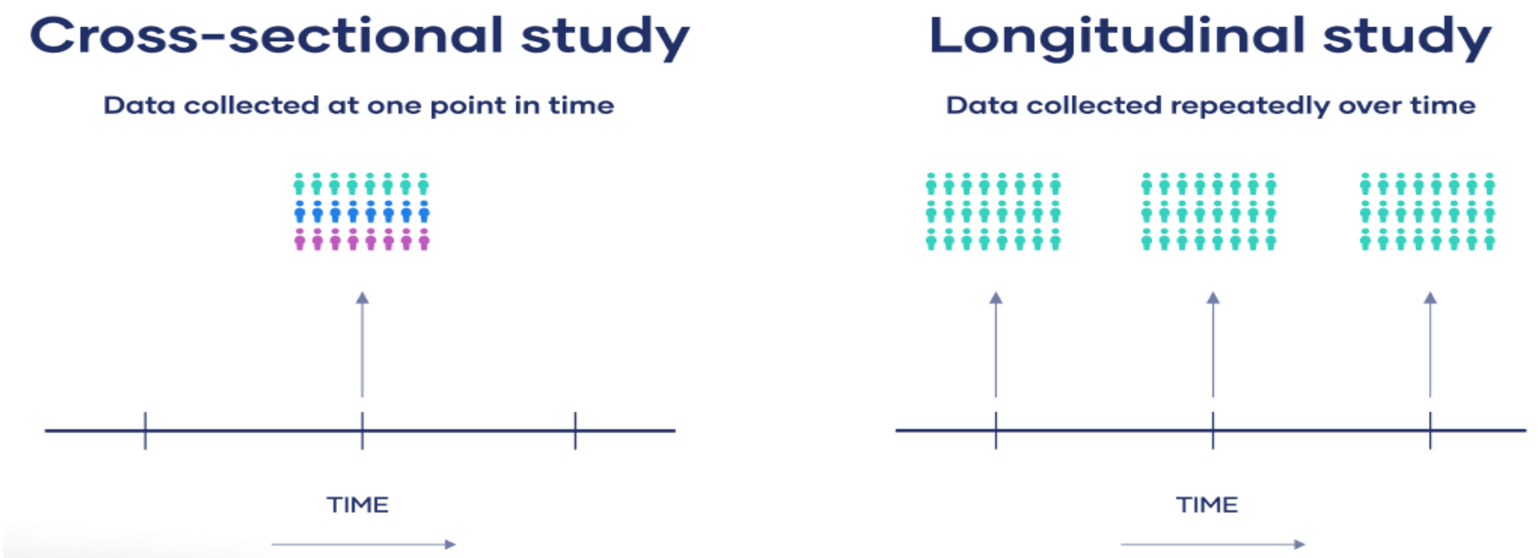
Department of Social and Preventive Medicine,

Longitudinal study

- A longitudinal study is a research strategy in which the same variables (e.g., *individuals*) are observed repeatedly over a short or long period of time (i.e., uses longitudinal data).

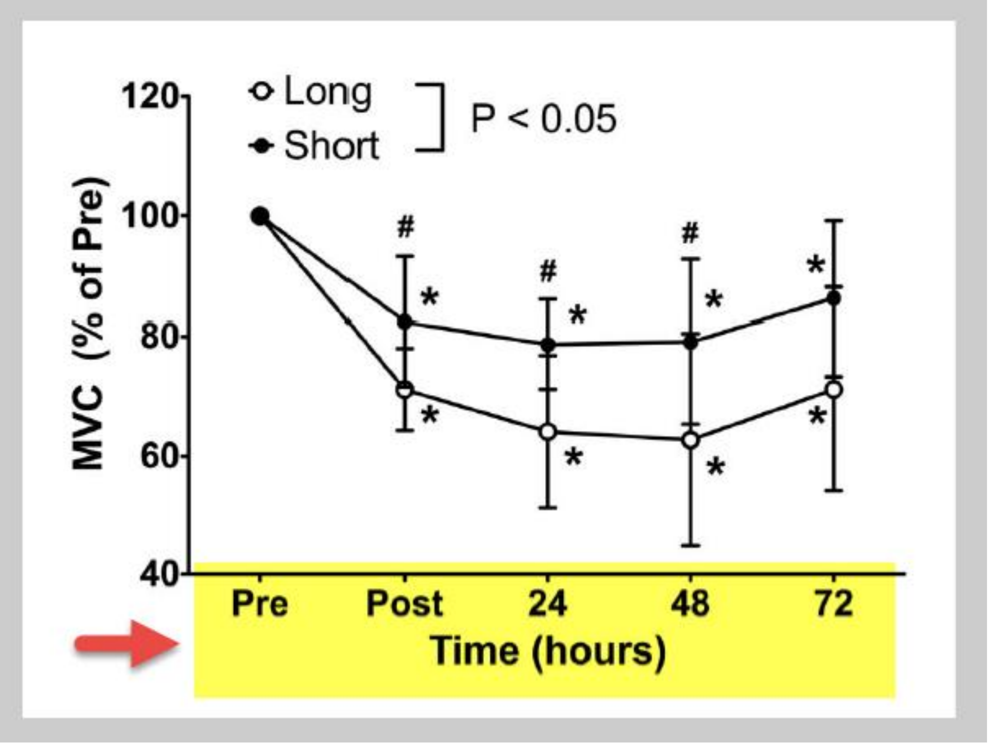
- It's usually an observational research, although it may also take the form of a longitudinal randomized experiment

- In a longitudinal study, researchers repeatedly examine the same individuals to detect any changes that might occur over a period of time.



Effects of Eccentric Cycling Performed at Long vs. Short Muscle Lengths on Heart Rate, Rate Perceived Effort, and Muscle Damage Markers

Luis Peñailillo,[1] Carolina Aedo,[1] Mayari Cartagena,[1] Alejandra Contreras,[1]
Alvaro Reyes,[2] Rodrigo Ramirez-Campillo,[3] Jacob E. Earp,[4] and Hermann Zbinden-Foncea[1]



Organisational Stressors, Coping, and Coping Effectiveness: A Longitudinal Study with an Elite Coach

Andrew Levy[1], Adam Nicholls[2], David Marchant[3] and Remco Polman[4]

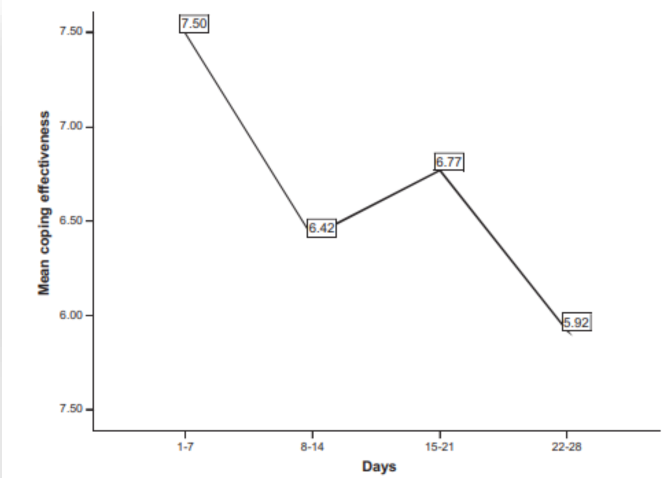
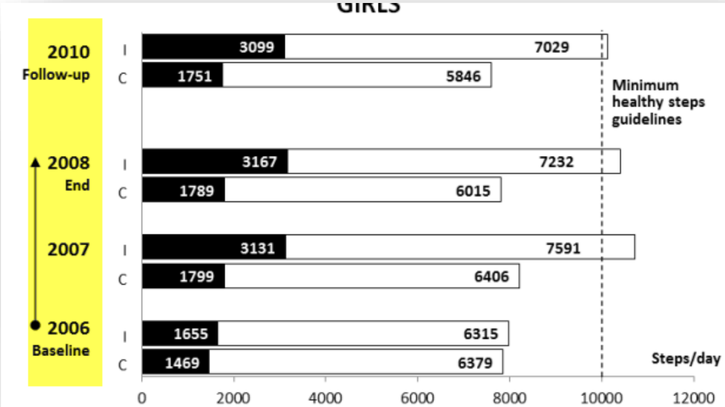


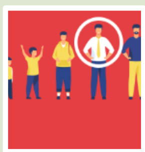
Figure 3. Coping Effectiveness Reported Over 28-Day Period

Longitudinal 2-Year Follow-up on the Effect of a Non-Randomised School-Based Physical Activity Intervention on Reducing Overweight and Obesity of Czech Children Aged 10–12 Years

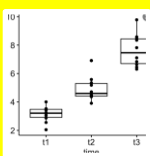
Erik Sigmund [†,*] and Dagmar Sigmundová [†]



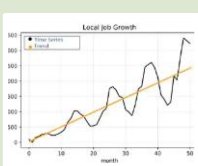
Longitudinal and Time-Series Analysis



Longitudinal analysis is concerned with studying the progression of the values of a variable over time for the members of a population.

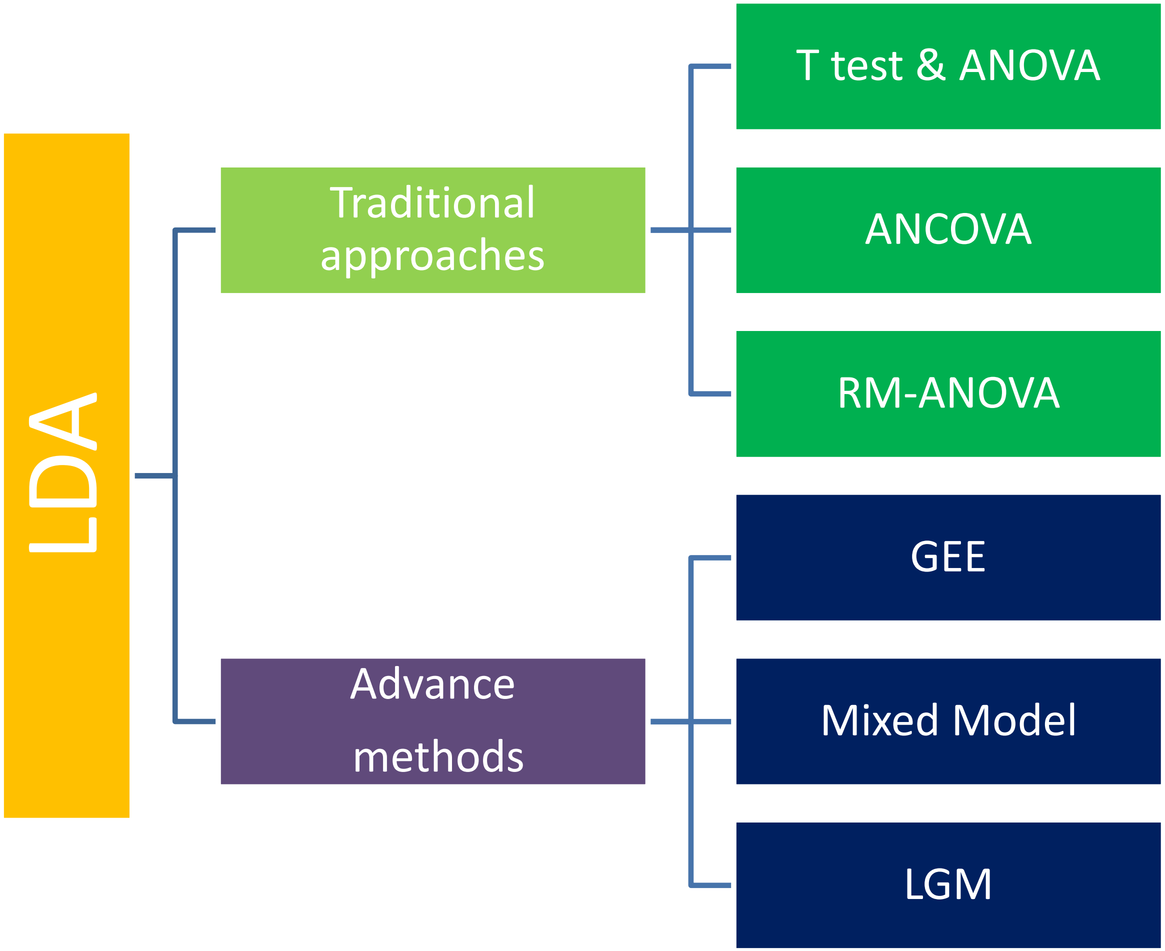


If time is defined as a **categorical variable**, longitudinal analysis is closely related to multivariate analysis, studying **vectors of outcomes**.



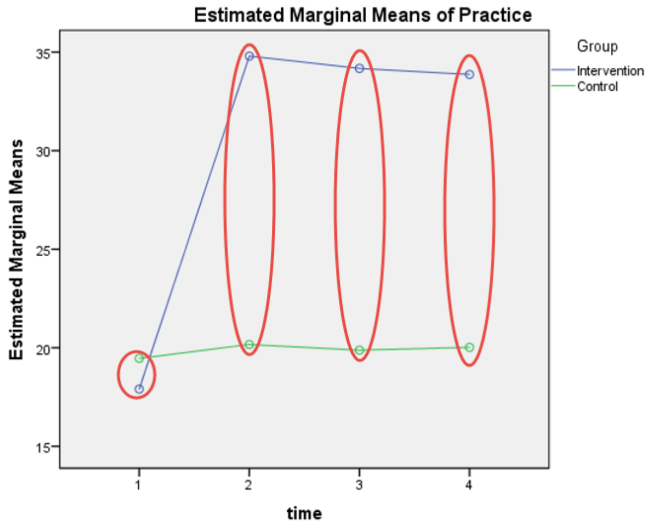
When time is a **continuous variable**, longitudinal analysis studies the subjects' curves (trajectories), and random coefficient models are well suited for this purpose

Data analysis approaches

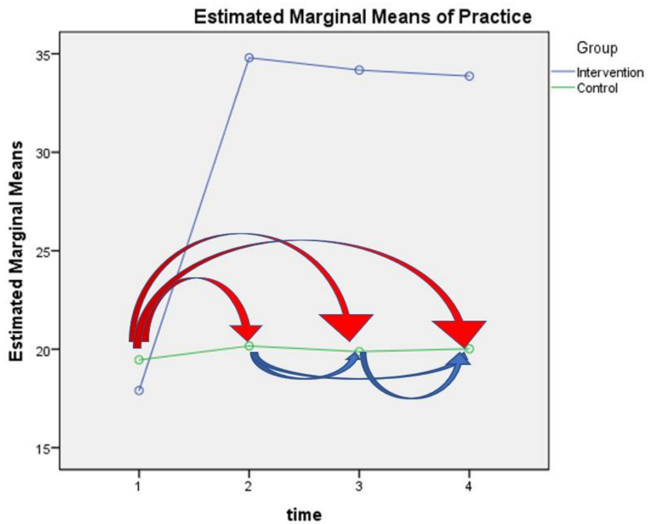


Types of comparisons in repeated measure designs

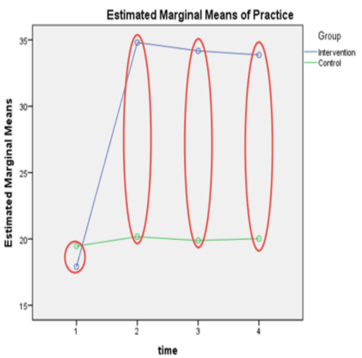
Between-group comparison

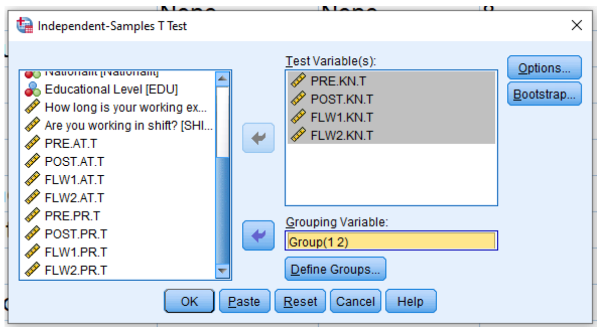


Within-group comparison



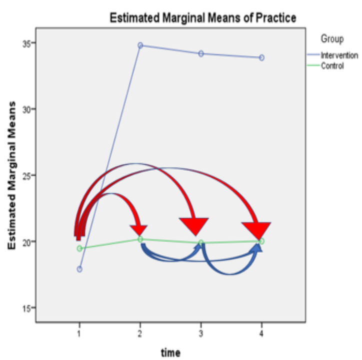
T test & One way ANOVA (method 1)

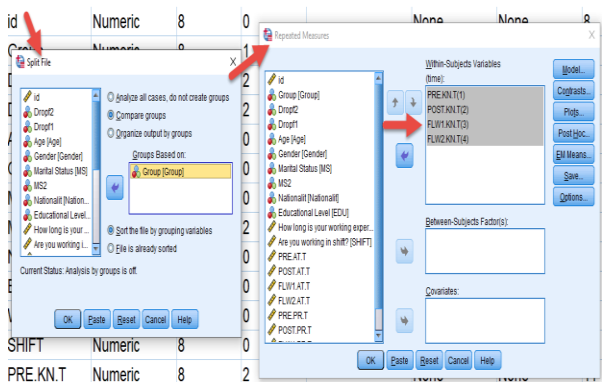




Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
PRE.KN.T	Equal variances assumed	.720	.398	-.275	138	.783	-.07143	.25932
	Equal variances not assumed			-.275	135.615	.783	-.07143	.25932
POST.KN.T	Equal variances assumed	.003	.954	23.021	138	.000	7.11429	.30904
	Equal variances not assumed			23.021	137.986	.000	7.11429	.30904
FLW1.KN.T	Equal variances assumed	.205	.651	28.793	129	.000	7.10233	.24667
	Equal variances not assumed			28.778	128.144	.000	7.10233	.24680
FLW2.KN.T	Equal variances assumed	1.151	.285	27.050	123	.000	6.91398	.25560
	Equal variances not assumed			27.025	120.895	.000	6.91398	.25584





Pairwise Comparisons

Measure: MEASURE_1

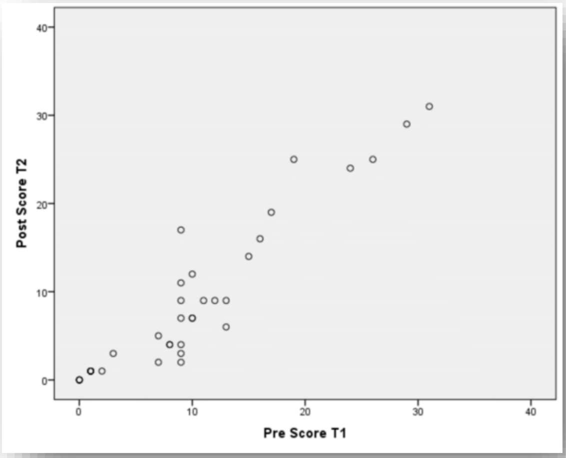
(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-4.168*	.369	.000	-5.158	-3.178
	3	-3.840*	.359	.000	-4.803	-2.877
	4	-3.792*	.354	.000	-4.742	-2.842
2	1	4.168*	.369	.000	3.178	5.158
	3	.328	.180	.421	-.153	.809
	4	.376	.190	.302	-.134	.886
3	1	3.840*	.359	.000	2.877	4.803
	2	-.328	.180	.421	-.809	.153
	4	.048	.085	1.000	-.180	.276
4	1	3.792*	.354	.000	2.842	4.742
	2	-.376	.190	.302	-.886	.134
	3	-.048	.085	1.000	-.276	.180

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

Analysis of Covariance (Method 2)

- ANCOVA is commonly used for analysis of **pretest-posttest designs** in which groups are compared at posttest, using pretest scores as the covariate to control for ***pre-existing differences*** on the dependent variable

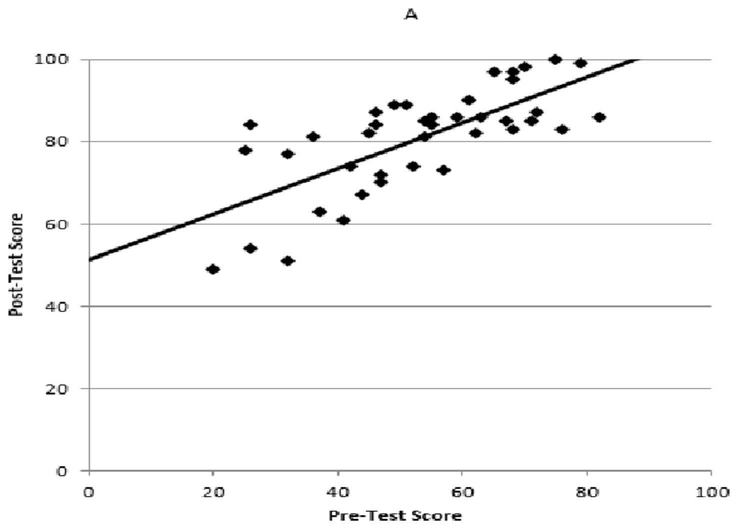


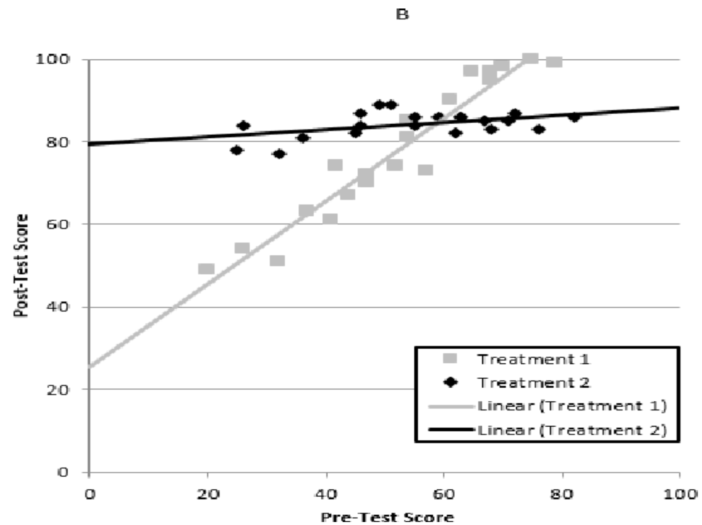
- **ANCOVA is sometimes recommended with experimental research (e.g., Huitema, 2011; Maxwell &Delaney, 1990), where inclusion of covariates can have the effect of reducing the mean square error and**
- **thus increasing the power of the analysis**

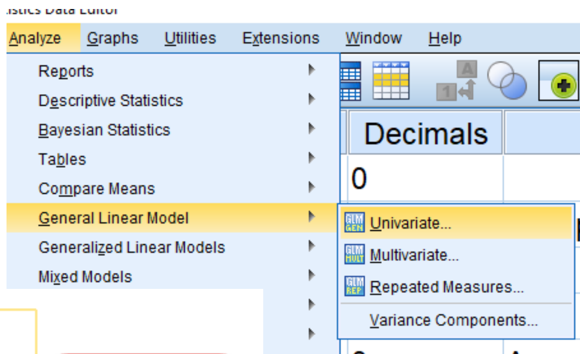
- **Among the methods, ANCOVA-POST is generally regarded as the preferred approach, given that it typically leads to unbiased treatment effect estimate with the lowest variance relative to ANOVA-POST or ANOVA-CHANGE**

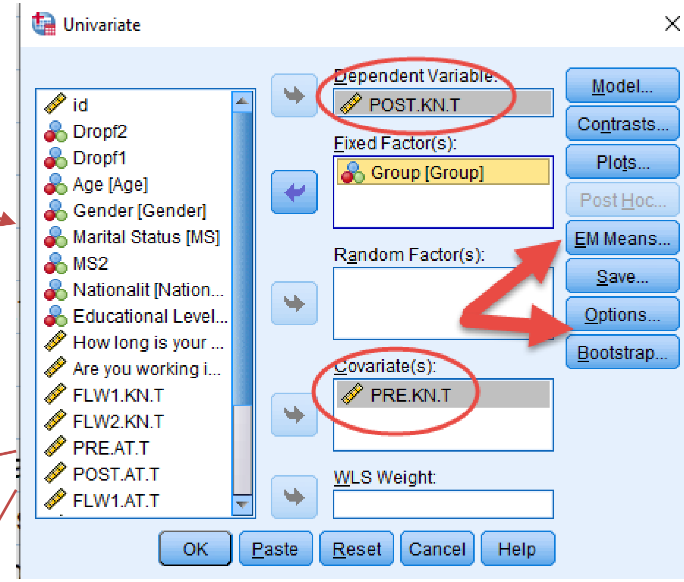
Assumption of ANCOVA

- linearity of regression
- Homogeneity of error variances
- Independence of error terms
- Normality of error terms
- Homogeneity of regression slopes (when you have more than 1 group)









Descriptive Statistics

Dependent Variable: POST.KN.T

Group	Mean	Std. Deviation	N
1.0	19.0143	1.83755	70
2.0	11.9000	1.81898	70
Total	15.4571	4.00785	140

Unadjusted mean score at post test

Estimated Marginal Means

Group

Estimates

Dependent Variable: POST.KN.T

Group	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1.0	19.022[a]	.215	18.597	19.448
2.0	11.892[a]	.215	11.466	12.318

a. Covariates appearing in the model are evaluated at the following values: PRE.KN.T = 11.2071.

Adjusted mean after excluding the effect of pre test

Tests of Between-Subjects Effects

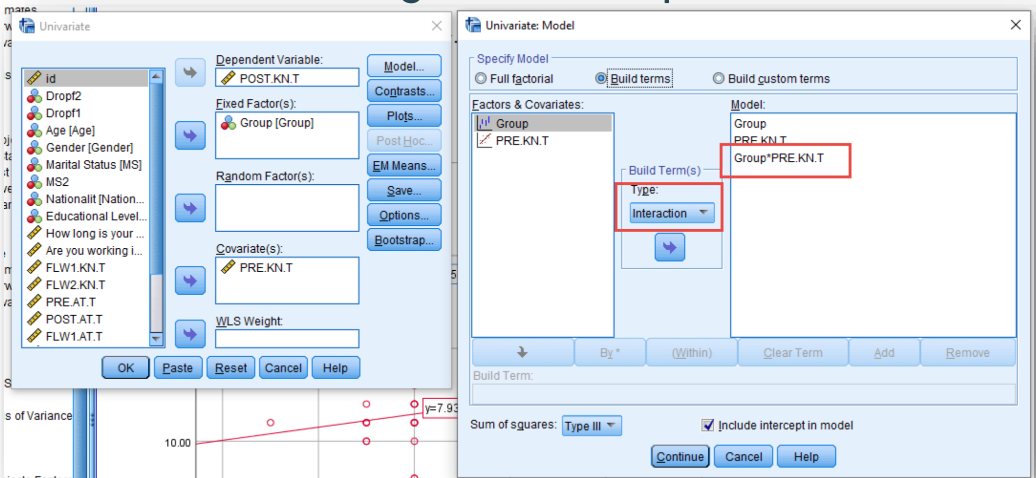
Dependent Variable: POST.KN.T

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	1788.102[a]	2	894.051	275.469	.000	.801	550.939	1.000
Intercept	423.870	1	423.870	130.600	.000	.488	130.600	1.000
PRE.KN.T	16.645	1	16.645	5.128	.025	.036	5.128	.614
Group	1778.541	1	1778.541	547.993	.000	.800	547.993	1.000
Error	444.641	137	3.246					
Total	35682.000	140						
Corrected Total	2232.743	139						

a. R Squared = .801 (Adjusted R Squared = .798)

b. Computed using alpha = .05

Test of Regression Slopes



Dependent Variable: POST.KN.T

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1792.078[a]	3	597.359	184.360	.000
Intercept	405.327	1	405.327	125.094	.000
Group	57.848	1	57.848	17.853	.000
PRE.KN.T	18.561	1	18.561	5.728	.018
Group * PRE.KN.T	3.976	1	3.976	1.227	.270
Error	440.665	136	3.240		
Total	35682.000	140			
Corrected Total	2232.743	139			

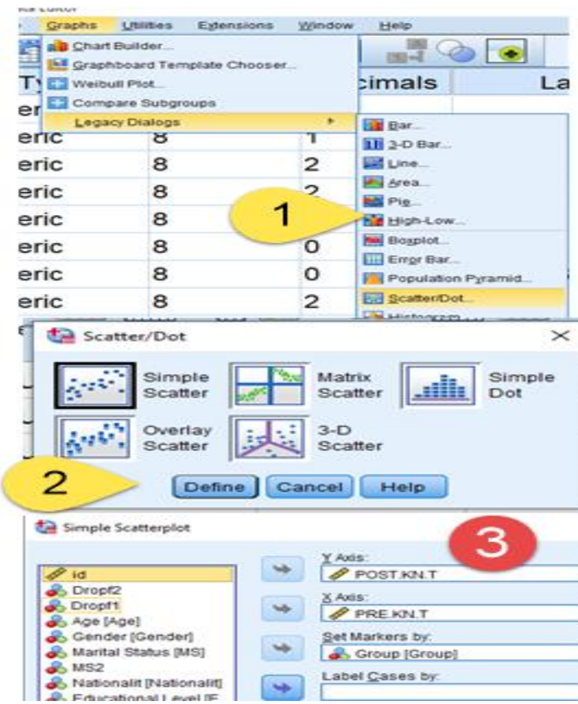
2016 Vol. 12 no. 3

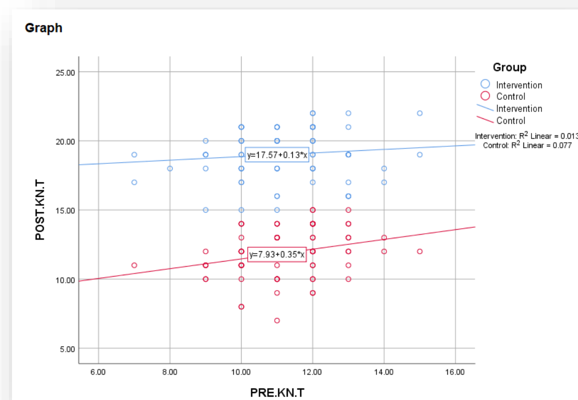
Violation of the homogeneity of regression slopes assumption in ANCOVA for two-group pre-post designs: Tutorial on a modified Johnson-Neyman procedure

Teresa R. Johnson[a]
[a]Johns Hopkins University School of Medicine, Baltimore

https://www.tqmp.org/RegularArticles/vol12-3/p253/p253.pdf

Drawing of Regression Slopes





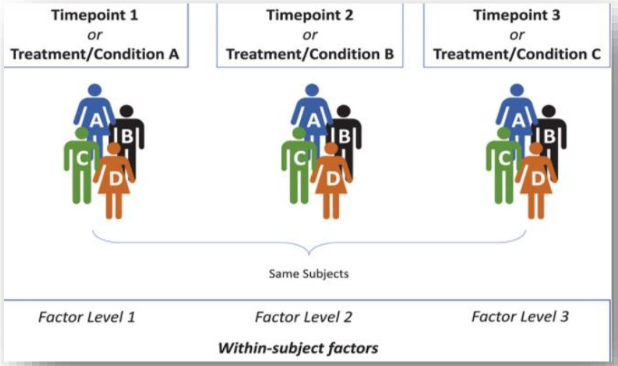
Repeated Measures ANOVA (Method 3)

Repeated Measures ANOVA – one-way ANOVA but the same subjects are measured in each group.

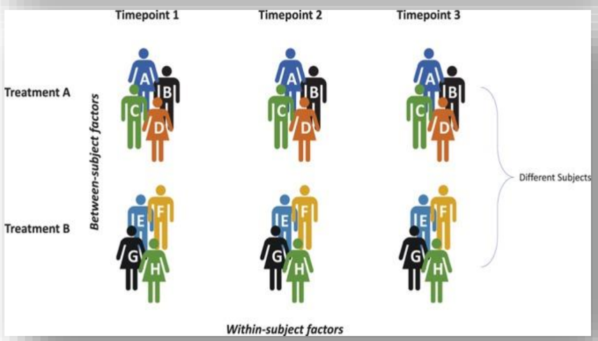
In repeated measures subjects serve as their own controls.own controls.

- ***Differences in means must be due to:***
- ***Treatment variations***
- ***Within subject's variations***
- ***within subject's error (unexplained variation)error***

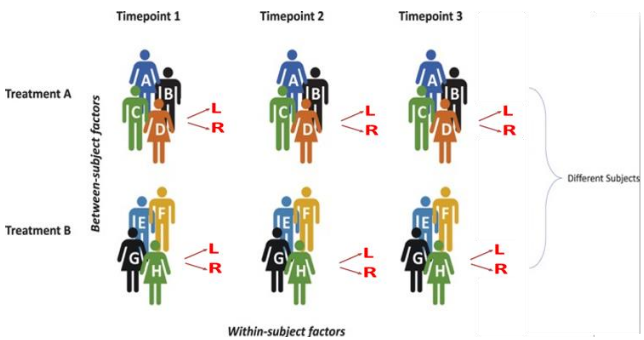
Repeated measures designs are more powerful than independent groups designs.



One-way RM-ANOVA/ MANOVA



Two-way RM-ANOVA /MANOVA



Three-way RM-ANOVA / MANOVA

Between-Subjects Design

SS TOTAL

SS Effect (SS between-groups) | SS Error (SS within-groups)

SS TOTAL = SS Effect + SS Error

(SS between groups) (SS within groups)

Repeated-Measures Design

SS TOTAL

SS Effect (SS between-groups) | SS Error (SS within-groups)

SS Error (SS within-groups) → SS Subjects | SS Error

SS TOTAL = SS Effect + SS Error

(SS between groups) (SS within groups)

This get's split up into two parts

split up

SS TOTAL = SS Effect + (SS Subjects + SS Error)

(SS within groups)

(SS between groups) (SS Subjects) (SS Left-over Error)

Comparison to One-Way ANOVA

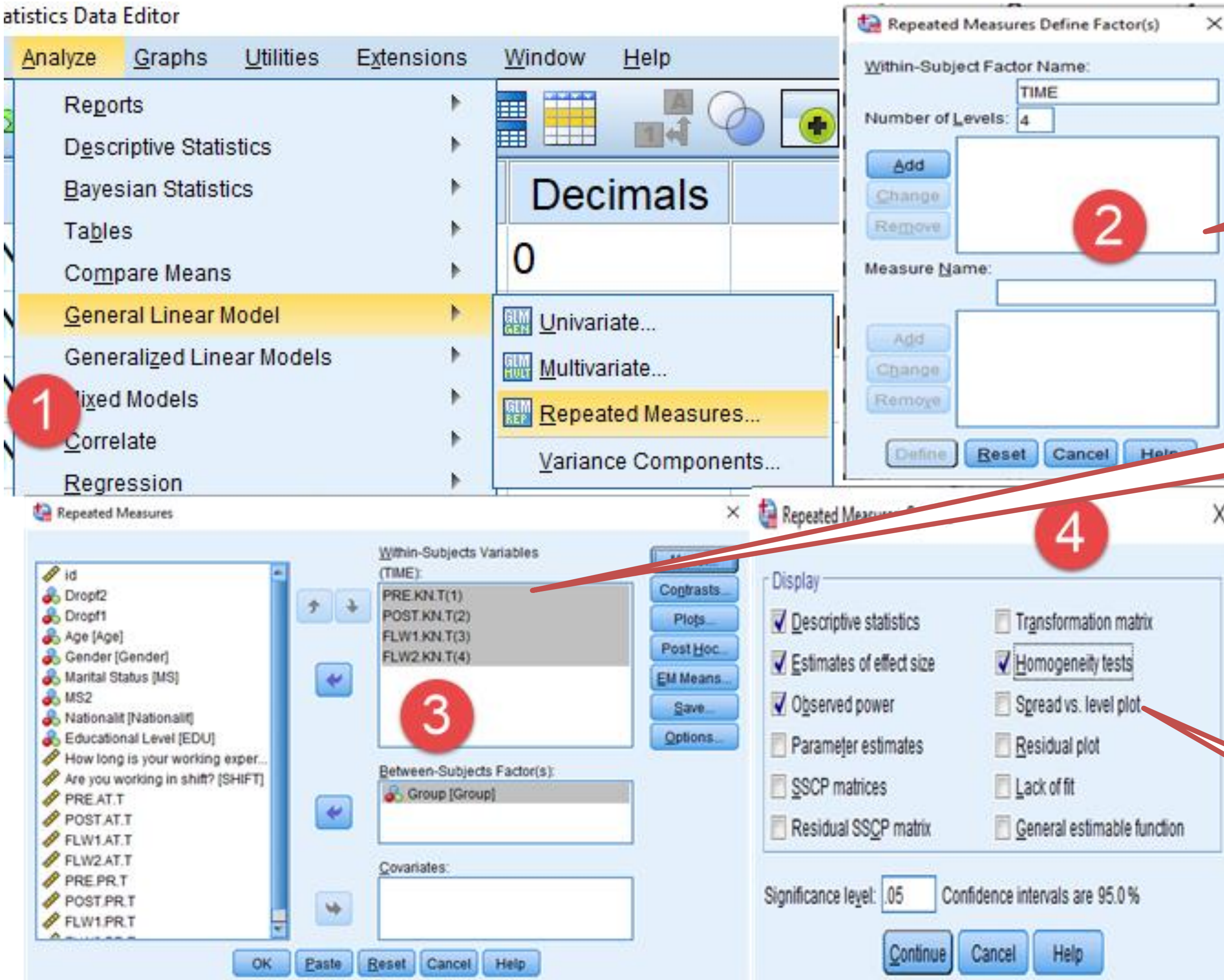
- Other assumptions hold (e.g., normality, equal variance), but sphericity is an added assumption.
  - Sphericity means data are uncorrelated.
- Counterbalancing may be needed.
- η^2 is interpreted the same way as for one-way ANOVA.
- Post-hoc t-tests need to be for paired samples, not independent groups.

Assumptions

The following assumptions are made when using the F-test.

- 1. The response variable is continuous.
- 2. The eijk follow the normal probability distribution with mean equal to zero.
- 3. The variances of the eijk are equal for all values of i, j, and k.
- 4. The individuals are independent

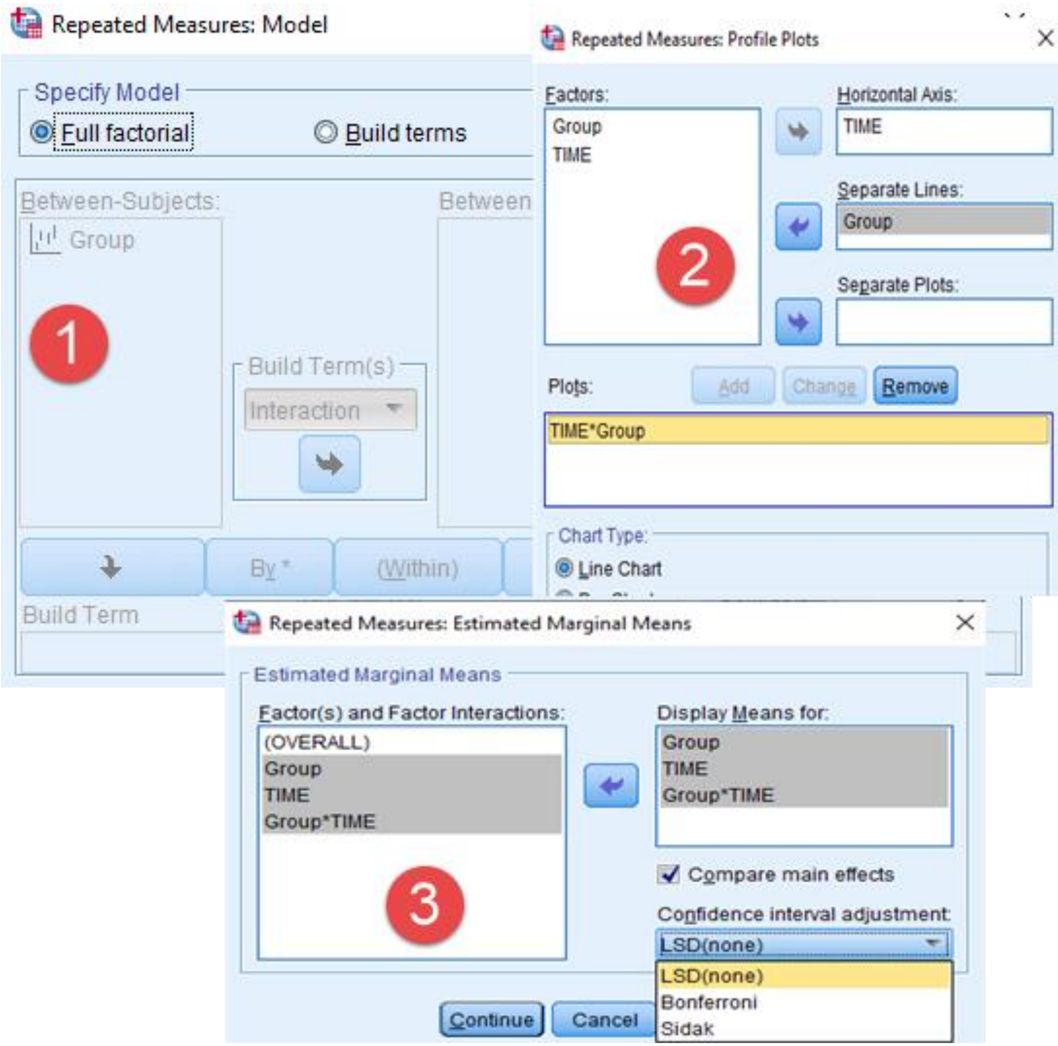
RM-ANOVA using SPSS



Define the within subject factors and level

Enter your DV and define the between subject variable

Define the Analysis properties



1. You can specify and customize the model
2. Defining the graph (line & bar)
3. EM – Mean calculation and comparison

Modified syntax for pairwise comparison

- DATASET ACTIVATE DataSet1.
- GLM PRE.KN.T POST.KN.T FLW1.KN.T FLW2.KN.T BY Group
- /WSFACTOR=TIME 4 Polynomial
- /METHOD=SSTYPE(3)
- /PLOT=PROFILE(TIME*Group) TYPE=LINE ERRORBAR=NO MEANREFERENCE=NO YAXIS=AUTO
- /EMMEANS=TABLES(Group*TIME) COMPARE (Group) ADJ(LSD)
- /EMMEANS=TABLES(Group*TIME) COMPARE (Time) ADJ(LSD)
- /EMMEANS=TABLES(Group*TIME)
- /PRINT=DESCRIPTIVE ETASQ OPOWER HOMOGENEITY
- /CRITERIA=ALPHA(.05)
- /WSDESIGN=TIME
- /DESIGN=Group.

Pairwise Comparisons

Measure: MEASURE_1

TIME	(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	1.0	2.0	.037	.258	.887	-.474	.548
	2.0	1.0	-.037	.258	.887	-.548	.474
2	1.0	2.0	7.064*	.338	.000	6.394	7.734
	2.0	1.0	-7.064*	.338	.000	-7.734	-6.394
3	1.0	2.0	7.139*	.245	.000	6.654	7.624
	2.0	1.0	-7.139*	.245	.000	-7.624	-6.654
4	1.0	2.0	6.914*	.256	.000	6.408	7.420
	2.0	1.0	-6.914*	.256	.000	-7.420	-6.408

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Univariate Tests

Measure: MEASURE_1

TIME		Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
1	Contrast	.042	1	.042	.020	.887	.000	.020	.052
	Error	256.406	123	2.085					
2	Contrast	1559.280	1	1559.280	436.016	.000	.780	436.016	1.000
	Error	439.872	123	3.576					
3	Contrast	1592.458	1	1592.458	848.114	.000	.873	848.114	1.000
	Error	230.950	123	1.878					
4	Contrast	1493.751	1	1493.751	731.715	.000	.856	731.715	1.000
	Error	251.097	123	2.041					

Each F tests the simple effects of Group within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Computed using alpha = .05

Pairwise Comparisons

Measure: MEASURE_1

Group	(I) TIME	(J) TIME	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
1.0	1	2	-7.710*	.273	.000	-8.250	-7.169
		3	-7.419*	.235	.000	-7.885	-6.953
		4	-7.258*	.247	.000	-7.748	-6.768
	2	1	7.710*	.273	.000	7.169	8.250
		3	.290	.256	.259	-.216	.797
		4	.452	.271	.098	-.085	.988
	3	1	7.419*	.235	.000	6.953	7.885
		2	-.290	.256	.259	-.797	.216
		4	.161	.120	.182	-.077	.399
	4	1	7.258*	.247	.000	6.768	7.748
		2	-.452	.271	.098	-.988	.085
		3	-.161	.120	.182	-.399	.077
2.0	1	2	-.683*	.271	.013	-1.219	-.146
		3	-.317	.234	.177	-.780	.145
		4	-.381	.245	.123	-.867	.105
	2	1	.683*	.271	.013	.146	1.219
		3	.365	.254	.153	-.137	.868
		4	.302	.269	.264	-.231	.834
	3	1	.317	.234	.177	-.145	.780
		2	-.365	.254	.153	-.868	.137
		4	-.063	.119	.595	-.299	.172
	4	1	.381	.245	.123	-.105	.867
		2	-.302	.269	.264	-.834	.231
		3	.063	.119	.595	-.172	.299

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

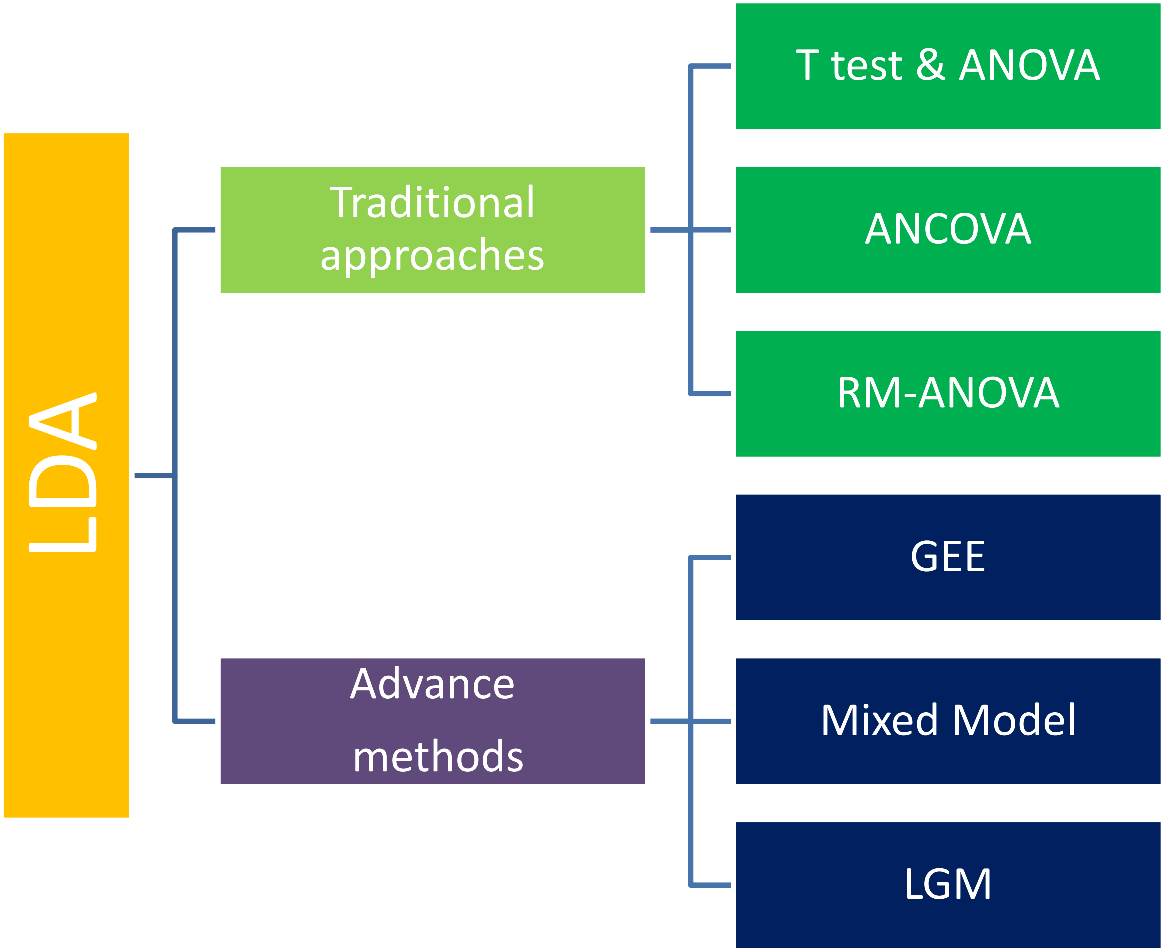
Multivariate Tests

Group		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
1.0	Pillai's trace	.908	395.946[a]	3.000	121.000	.000	.908	1187.837	1.000
	Wilks' lambda	.092	395.946[a]	3.000	121.000	.000	.908	1187.837	1.000
	Hotelling's trace	9.817	395.946[a]	3.000	121.000	.000	.908	1187.837	1.000
	Roy's largest root	9.817	395.946[a]	3.000	121.000	.000	.908	1187.837	1.000
2.0	Pillai's trace	.052	2.223[a]	3.000	121.000	.089	.052	6.669	.552
	Wilks' lambda	.948	2.223[a]	3.000	121.000	.089	.052	6.669	.552
	Hotelling's trace	.055	2.223[a]	3.000	121.000	.089	.052	6.669	.552
	Roy's largest root	.055	2.223[a]	3.000	121.000	.089	.052	6.669	.552

Each F tests the multivariate simple effects of TIME within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

GLM
• t-test S
• Linear Regression (Simple, Multiple)
• ANOVA /MANOVA
• One way ANOVA , Two Way ANOVA , ...(FACTORIAL)
• ANCOVA /MANCOVA
• RM ANOVA /MANOVA /ANCOVA/MANCOVA

Data analysis approaches



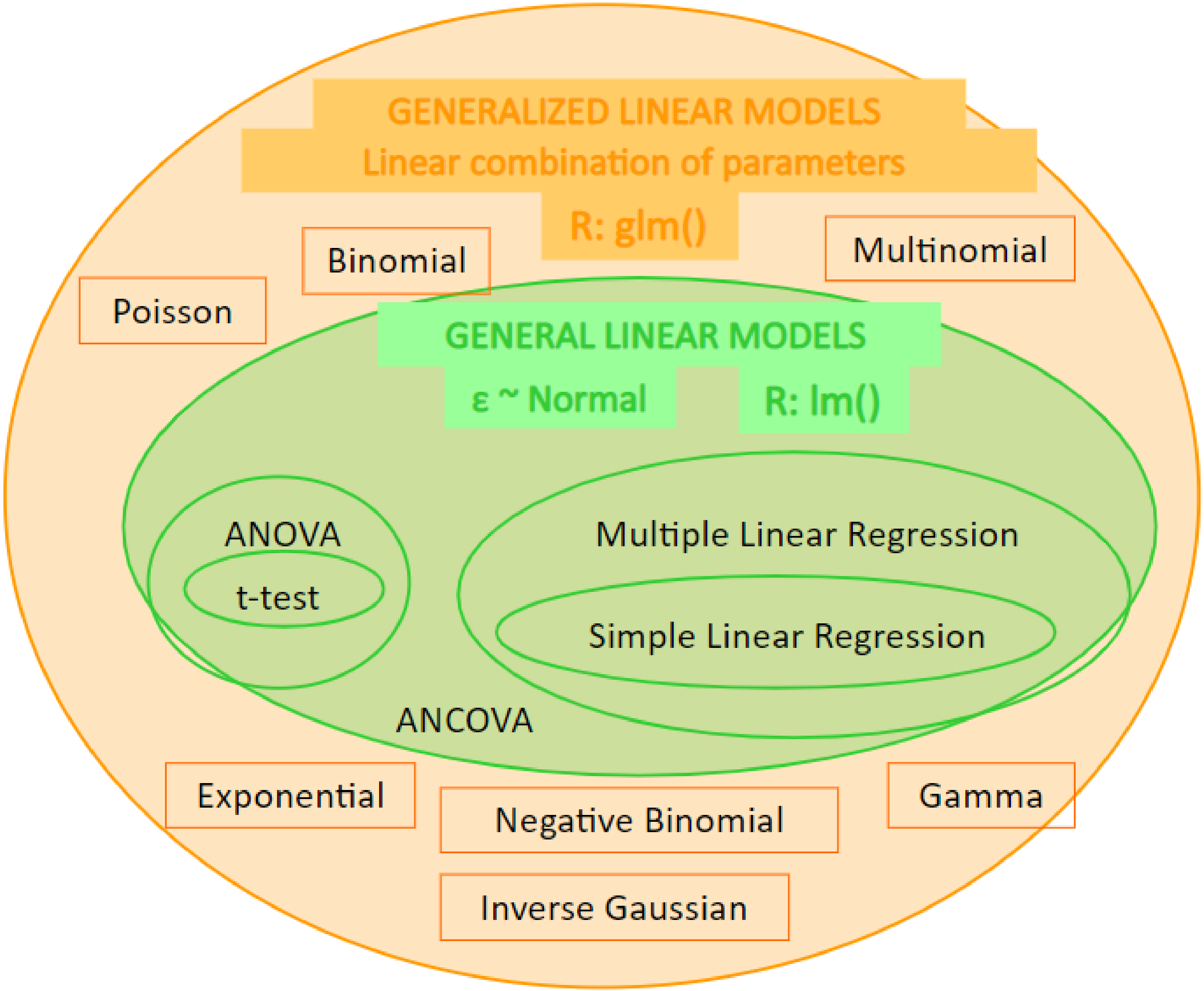
Comparison of traditional and new methods

Table 1. Comparison of Traditional and Mixed-Effects Approaches for the Analysis of Repeated-Measures Data

	End-Point Analysis	rANOVA	rMANOVA	Mixed-Effects Analysis
Complete data required on every subject	Yes	No*	Yes	No
Possible effect of omitting subjects with missing values	Sample bias	Sample bias	Sample bias	Not applicable†
Possible effects of imputation of missing data	Estimation bias	Estimation bias	Estimation bias	Not applicable†
Subjects measured at different time points	Yes	No	No	Yes
Description of time effect	Simple	Flexible	Flexible	Flexible
Estimation of individual trends	No	No	No	Yes
Restrictive assumptions about correlation pattern	Not applicable	Yes	No	No
Time-dependent covariates	No	Yes	No	Yes
Ease of implementation	Very easy	Easy	Easy	Hard
Computational complexity	Low	Low	Medium	High

Abbreviations: rANOVA, univariate repeated-measures analysis of variance; rMANOVA, multivariate repeated-measures analysis of variance.
*Subjects with missing data are often omitted from the analysis.
†It is not necessary to omit subjects with missing values from the analysis or to impute missing values.

FROM:
Ralitza Gueorguieva, PhD; John H. Krystal, MD Move Over ANOVA : Progress in Analyzing Repeated-Measures Data and Its Reflection in Papers Published in the Archives of General Psychiatry. *Arch Gen Psychiatry*. 2004;61:310-317.


GENERALIZED LINEAR MODELS
Linear combination of parameters
R: glm()
Binomial
Multinomial
Poisson
GENERAL LINEAR MODELS
ε ~ Normal
R: lm()
ANOVA
t-test
Multiple Linear Regression
Simple Linear Regression
ANCOVA
Exponential
Negative Binomial
Gamma
Inverse Gaussian

Generalized Estimating Equations

Generalized Linear Models

- The generalized linear model expands the general linear model so that the dependent variable is linearly related to the factors and covariates via a specified link function. Moreover, the model allows for the dependent variable **to have a non-normal distribution**

- It covers widely used statistical models, such as linear regression for normally distributed responses, **logistic models for binary data, loglinear models for count data, complementary log-log models for interval-censored survival data**, plus many other statistical models through its very general model formulation.\

The general linear model make the 5 assumptions below. When these assumptions are met, OLS regression coefficients are MVUE (Minimum Variance Unbiased Estimators) and BLUE (Best Linear Unbiased Estimators).

1. Exact X: The IVs are assumed to be known exactly (i.e., without measurement error)
2. Independence: Residuals are independently distributed (prob. of obtaining a specific observation does not depend on other observations)
3. Normality: All residual distributions are normally distributed

4. Constant variance: All residual distributions have a constant variance, SEE^2
5. Linearity: All residual distributions (i.e., for each Y') are assumed to have means equal to zero

Problems and Solutions

Normality: **Inefficient (with large N). Use power transformations, generalized linear models**

Constant variance: **Inefficient and inaccurate standard errors. Use power transformations, SE corrections, weighted least squares , generalized linear models**

Linearity: **Biased parameter estimates. Use power transformations, polynomial regression, generalized linear models**

- The generalized linear model expands the general linear model so that the dependent variable is linearly related to the factors and covariates via a specified link function.
- The link function is a transformation of the dependent variable that allows estimation of the model. The following functions are available

- Identity. $f(x)=x$. The dependent variable is not transformed. This link can be used with any distribution.
- Complementary log-log. $f(x)=\log(-\log(1-x))$. This is appropriate only with the binomial distribution.
- Cumulative complementary log-log. $f(x)=\ln(-\ln(1-x))$, applied to the cumulative probability of each category of the response. This is appropriate only with the multinomial distribution.
- Cumulative logit. $f(x)=\ln(x / (1-x))$, applied to the cumulative probability of each category of the response. This is appropriate only with the multinomial distribution.

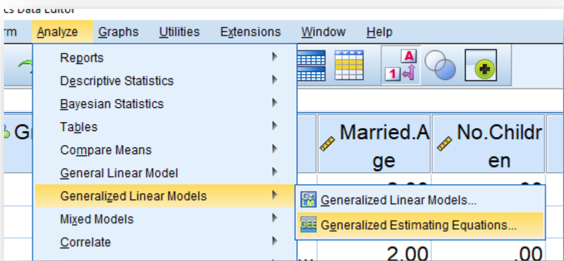
- Cumulative negative log-log. $f(x) = -\ln(-\ln(x))$, applied to the cumulative probability of each category of the response. This is appropriate only with the multinomial distribution.
- Cumulative probit. $f(x) = \Phi{-}1(x)$, applied to the cumulative probability of each category of the response, where $\Phi{-}1$ is the inverse standard normal cumulative distribution function. This is appropriate only with the multinomial distribution.
- Log. $f(x) = \log(x)$. This link can be used with any distribution.
- Log complement. $f(x) = \log(1-x)$. This is appropriate only with the binomial distribution.
- Logit. $f(x) = \log(x / (1-x))$. This is appropriate only with the binomial distribution.

- Negative binomial. $f(x)=\log(x / (x+k-1))$, where k is the ancillary parameter of the negative binomial distribution. This is appropriate only with the negative binomial distribution.
- Negative log-log. $f(x)=-\log(-\log(x))$. This is appropriate only with the binomial distribution.
- Odds power. $f(x)=[(x/(1-x))\alpha-1]/\alpha$, if $\alpha \neq 0$. $f(x)=\log(x)$, if $\alpha=0$. α is the required number specification and must be a real number. This is appropriate only with the binomial distribution.
- Probit. $f(x)=\Phi-1(x)$, where $\Phi-1$ is the inverse standard normal cumulative distribution function. This is appropriate only with the binomial distribution.
- Power. $f(x)=x\ \alpha$, if $\alpha \neq 0$. $f(x)=\log(x)$, if $\alpha=0$. α is the required number specification and must be a real number. This link can be used with any distribution.

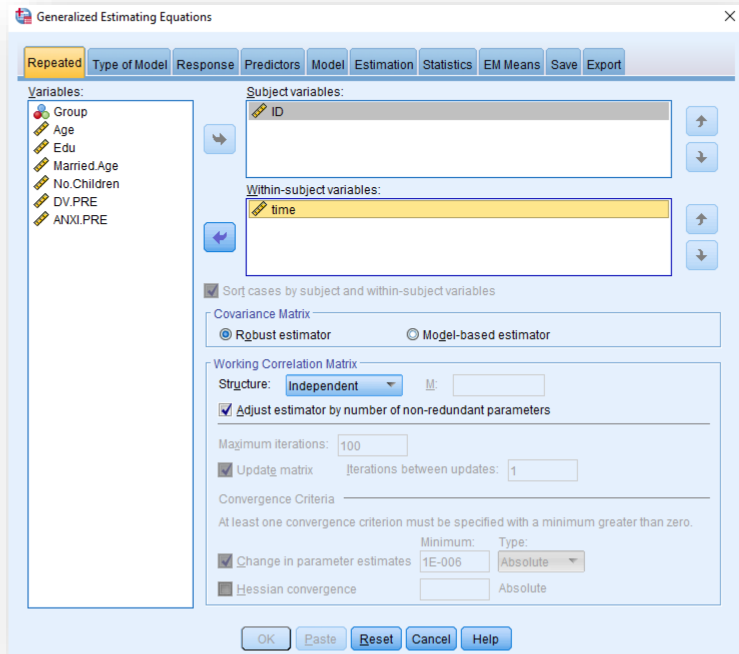
Types of outcome/DV in GEE

- **Scale Response.** The following options are available:
- **Linear.** Specifies Normal as the distribution and Identity as the link function.
- **Gamma with log link.** Specifies Gamma as the distribution and Log as the link function.
- **Ordinal Response.** The following options are available:
- **Ordinal logistic.** Specifies Multinomial (ordinal) as the distribution and Cumulative logit as the link function.
- **Ordinal probit.** Specifies Multinomial (ordinal) as the distribution and Cumulative probit as the link function.
- **Counts.** The following options are available:
- **Poisson loglinear.** Specifies Poisson as the distribution and Log as the link function.
- **Negative binomial with log link.** Specifies Negative binomial (with a value of 1 for the ancillary parameter) as the distribution and Log as the link function. To have the procedure estimate the value of the ancillary parameter, specify a custom model with Negative binomial distribution and select **Estimate value** in the Parameter group.
- **Binary Response or Events/Trials Data.** The following options are available:
- **Binary logistic.** Specifies Binomial as the distribution and Logit as the link function.
- **Binary probit.** Specifies Binomial as the distribution and Probit as the link function.
- **Interval censored survival.** Specifies Binomial as the distribution and Complementary log-log as the link function.
- **Mixture.** The following options are available:
- **Tweedie with log link.** Specifies Tweedie as the distribution and Log as the link function.
- **Tweedie with identity link.** Specifies Tweedie as the distribution and Identity as the link function.

GEE



Analyze
Graphs
Utilities
Extensions
Window
Help
Reports
Descriptive Statistics
Bayesian Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Generalized Linear Models...
Generalized Estimating Equations...
Married.Age
No.Children
2.00
.00



Generalized Estimating Equations
Repeated
Type of Model
Response
Predictors
Model
Estimation
Statistics
EM Means
Save
Export
Variables:
Group
Age
Edu
Married.Age
No.Children
DV.PRE
ANXI.PRE
Subject variables:
ID
Within-subject variables:
time
Sort cases by subject and within-subject variables
Covariance Matrix
Robust estimator
Model-based estimator
Working Correlation Matrix
Structure:
Independent
M:
Adjust estimator by number of non-redundant parameters
Maximum iterations:
100
Update matrix
Iterations between updates:
1
Convergence Criteria
At least one convergence criterion must be specified with a minimum greater than zero.
Minimum:
Type:
Change in parameter estimates
1E-006
Absolute
Hessian convergence
Absolute
OK
Paste
Reset
Cancel
Help

Working Correlation Matrix

- This **correlation matrix** represents the within-subject dependencies.

- Its size is determined by the number of measurements and thus the combination of values of within-subject variables.

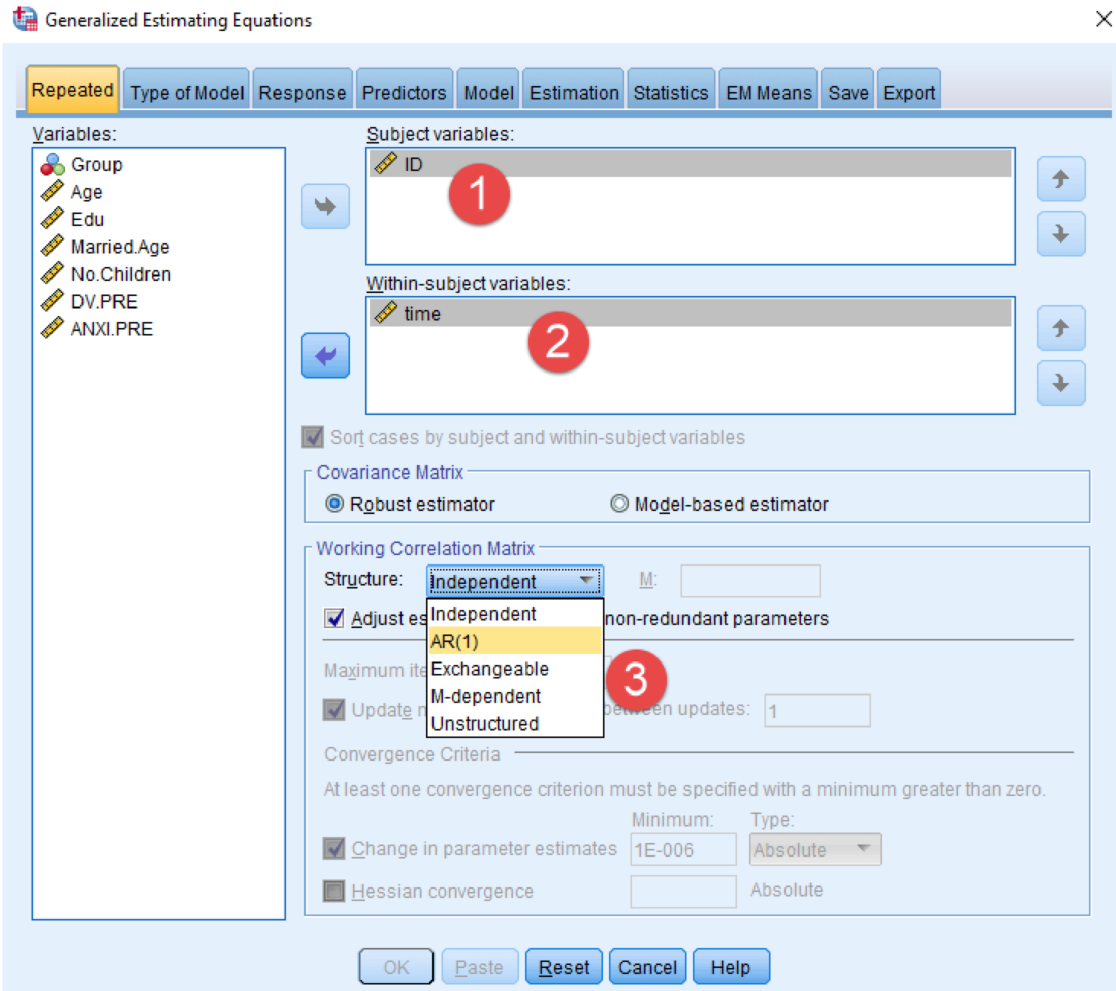
Repeated measurements have a first-order **autoregressive relationship**. The correlation between any two elements is equal to rho for adjacent elements, rho2 for elements that are separated by a third, and so on.

Independence, $$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Exchangeable, $$\begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

Autoregressive order 1, $$\begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}$$

Unstructured, $$\begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}$$

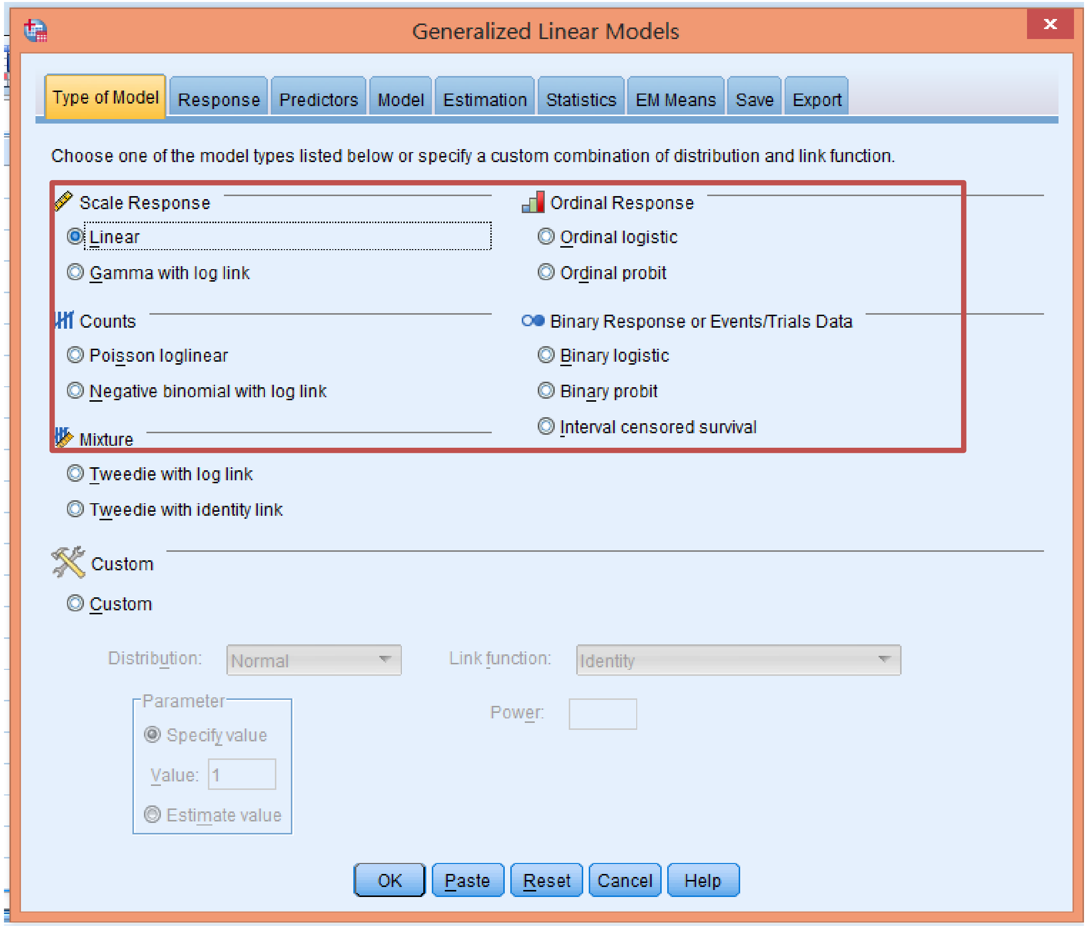


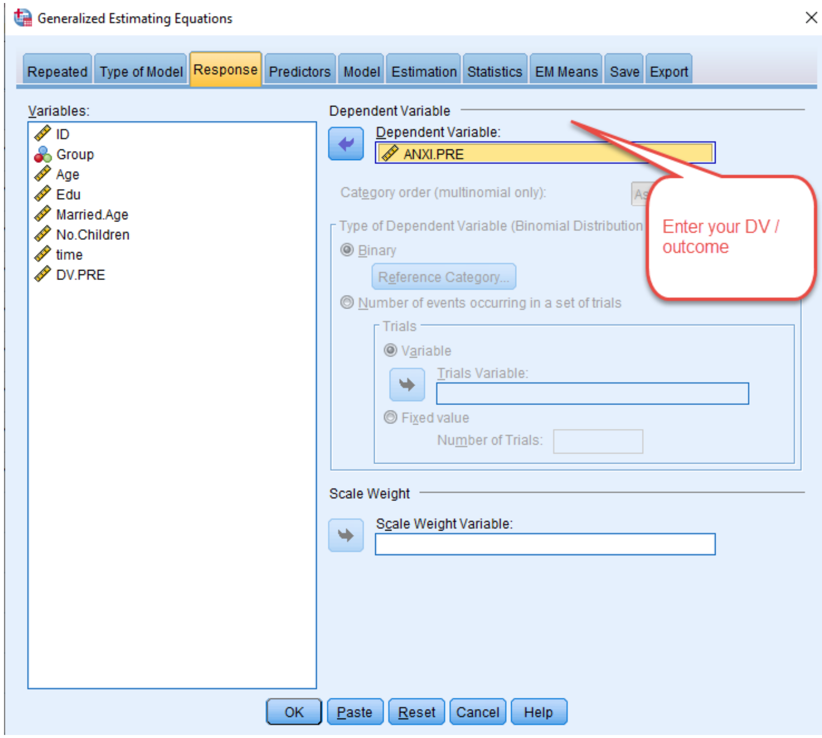
1 : Enter your subject's ID

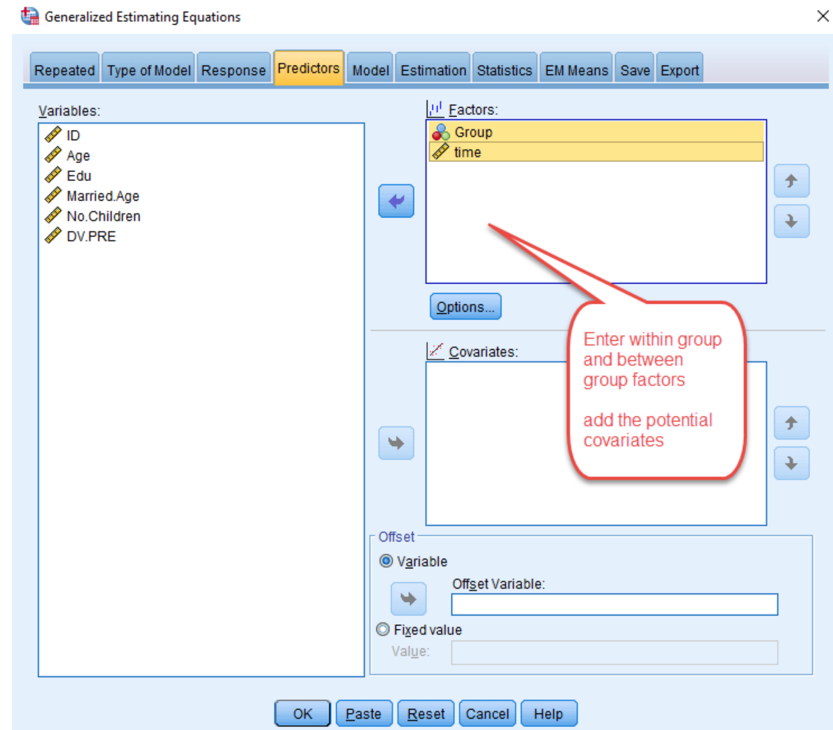
2- Within subject variable (Time)

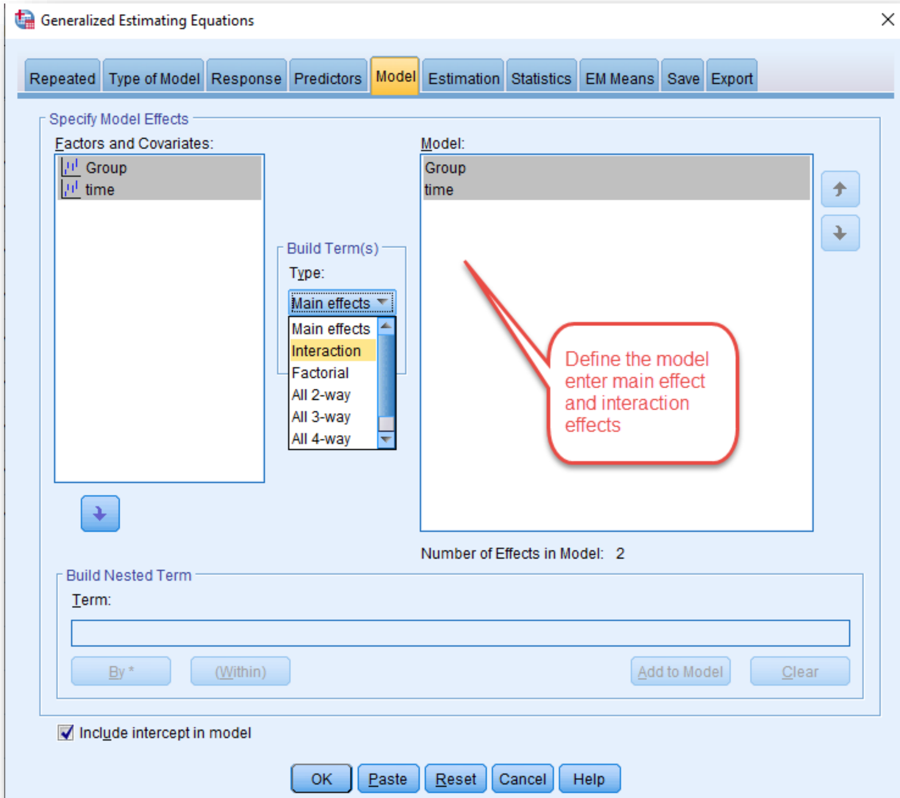
3 : define the type of correlation matrix

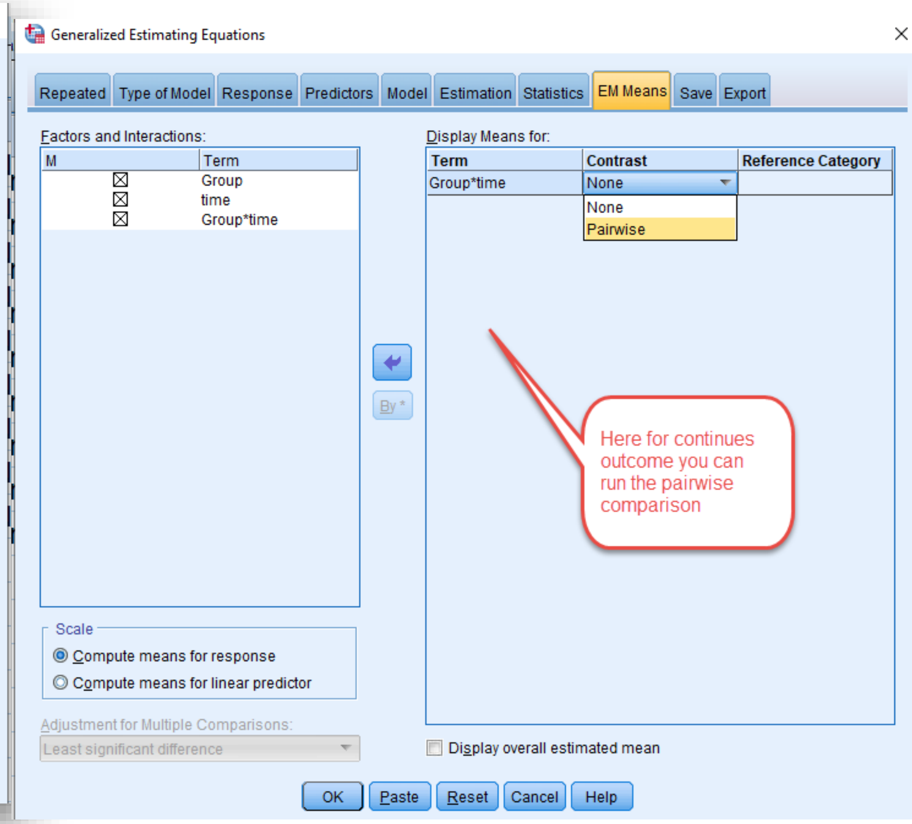
The response can be scale, counts, binary, or events-in-trials. Factors are assumed to be categorical. The covariates, scale weight, and offset are assumed to be scale.




Generalized Estimating Equations
Repeated
Type of Model
Response
Predictors
Model
Estimation
Statistics
EM Means
Save
Export
Variables:
ID
Group
Age
Edu
Married.Age
No.Children
time
DV.PRE
Dependent Variable
Dependent Variable:
ANXI.PRE
Category order (multinomial only):
Type of Dependent Variable (Binomial Distribution
Binary
Reference Category...
Number of events occurring in a set of trials
Trials
Variable
Trials Variable:
Fixed value
Number of Trials:
Scale Weight
Scale Weight Variable:
Enter your DV / outcome
OK
Paste
Reset
Cancel
Help


Generalized Estimating Equations
Repeated
Type of Model
Response
Predictors
Model
Estimation
Statistics
EM Means
Save
Export
Variables:
ID
Age
Edu
Married.Age
No.Children
DV.PRE
Factors:
Group
time
Options...
Covariates:
Enter within group and between group factors
add the potential covariates
Offset
Variable
Offset Variable:
Fixed value
Value:
OK
Paste
Reset
Cancel
Help


Generalized Estimating Equations
Repeated
Type of Model
Response
Predictors
Model
Estimation
Statistics
EM Means
Save
Export
Specify Model Effects
Factors and Covariates:
Group
time
Model:
Group
time
Build Term(s)
Type:
Main effects
Main effects
Interaction
Factorial
All 2-way
All 3-way
All 4-way
Define the model enter main effect and interaction effects
Number of Effects in Model: 2
Build Nested Term
Term:
By *
(Within)
Add to Model
Clear
Include intercept in model
OK
Paste
Reset
Cancel
Help


Generalized Estimating Equations
Repeated
Type of Model
Response
Predictors
Model
Estimation
Statistics
EM Means
Save
Export
Factors and Interactions:
M
Term
Group
time
Group*time
Display Means for:
Term
Contrast
Reference Category
Group*time
None
None
Pairwise
By *
Here for continues outcome you can run the pairwise comparison
Scale
Compute means for response
Compute means for linear predictor
Adjustment for Multiple Comparisons:
Least significant difference
Display overall estimated mean
OK
Paste
Reset
Cancel
Help

Goodness of Fit[a]

	Value
Quasi Likelihood under Independence Model Criterion (QIC)[b]	110.552
Corrected Quasi Likelihood under Independence Model Criterion (QICC)[b]	110.552

Dependent Variable: ANXI.PRE
Model: (Intercept), Group, time, Group * time

a. Information criteria are in smaller-is-better form.

b. Computed using the full log quasi-likelihood function.

This is same as ANOVA table , both main effects and also interaction effect are tested using Wald test

Tests of Model Effects

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	3199.338	1	.000
Group	14.801	1	.000
time	24.327	2	.000
Group * time	16.517	2	.000

Dependent Variable: ANXI.PRE
Model: (Intercept), Group, time, Group * time

Pairwise Comparisons

(I) Group*time	(J) Group*time	Mean Difference (I-J)	Std. Error	df	Sequential Bonferroni Sig.	95% Wald Confidence Interval for Difference[b]	
						Lower	Upper
[Group=1.00]*[time=1.00]	[Group=1.00]*[time=2.00]	.8750[a]	.14038	1	.000	.4630	1.2870
	[Group=1.00]*[time=3.00]	.6622[a]	.13303	1	.000	.2746	1.0498
	[Group=2.00]*[time=1.00]	.0972	.15267	1	1.000	-.2422	.4366
	[Group=2.00]*[time=2.00]	.2081	.14256	1	1.000	-.1754	.5915
	[Group=2.00]*[time=3.00]	.1203	.14443	1	1.000	-.2151	.4556
[Group=1.00]*[time=2.00]	[Group=1.00]*[time=1.00]	-.8750[a]	.14038	1	.000	-1.2870	-.4630
	[Group=1.00]*[time=3.00]	-.2128	.16381	1	1.000	-.6365	.2109
	[Group=2.00]*[time=1.00]	-.7778[a]	.17401	1	.000	-1.2764	-.2792
	[Group=2.00]*[time=2.00]	-.6669[a]	.16521	1	.001	-1.1357	-.1982
	[Group=2.00]*[time=3.00]	-.7547[a]	.16683	1	.000	-1.2369	-.2725
[Group=1.00]*[time=3.00]	[Group=1.00]*[time=1.00]	-.6622[a]	.13303	1	.000	-1.0498	-.2746
	[Group=1.00]*[time=2.00]	.2128	.16381	1	1.000	-.2109	.6365
	[Group=2.00]*[time=1.00]	-.5650[a]	.16700	1	.007	-1.0309	-.0991
	[Group=2.00]*[time=2.00]	-.4542[a]	.15781	1	.032	-.8857	-.0227
	[Group=2.00]*[time=3.00]	-.5419[a]	.15951	1	.007	-.9897	-.0942
[Group=2.00]*[time=1.00]	[Group=1.00]*[time=1.00]	-.0972	.15267	1	1.000	-.4366	.2422
	[Group=1.00]*[time=2.00]	.7778[a]	.17401	1	.000	.2792	1.2764
	[Group=1.00]*[time=3.00]	.5650[a]	.16700	1	.007	.0991	1.0309
	[Group=2.00]*[time=2.00]	.1108	.16260	1	1.000	-.2542	.4759
	[Group=2.00]*[time=3.00]	.0231	.15611	1	1.000	-.2912	.3373
[Group=2.00]*[time=2.00]	[Group=1.00]*[time=1.00]	-.2081	.14256	1	1.000	-.5915	.1754
	[Group=1.00]*[time=2.00]	.6669[a]	.16521	1	.001	.1982	1.1357
	[Group=1.00]*[time=3.00]	.4542[a]	.15781	1	.032	.0227	.8857
	[Group=2.00]*[time=1.00]	-.1108	.16260	1	1.000	-.4759	.2542
	[Group=2.00]*[time=3.00]	-.0878	.17201	1	1.000	-.4599	.2844
[Group=2.00]*[time=3.00]	[Group=1.00]*[time=1.00]	-.1203	.14443	1	1.000	-.4556	.2151
	[Group=1.00]*[time=2.00]	.7547[a]	.16683	1	.000	.2725	1.2369
	[Group=1.00]*[time=3.00]	.5419[a]	.15951	1	.007	.0942	.9897
	[Group=2.00]*[time=1.00]	-.0231	.15611	1	1.000	-.3373	.2912
	[Group=2.00]*[time=2.00]	.0878	.17201	1	1.000	-.2844	.4599

Pairwise comparisons of estimated marginal means based on the original scale of dependent variable ANXI.PRE

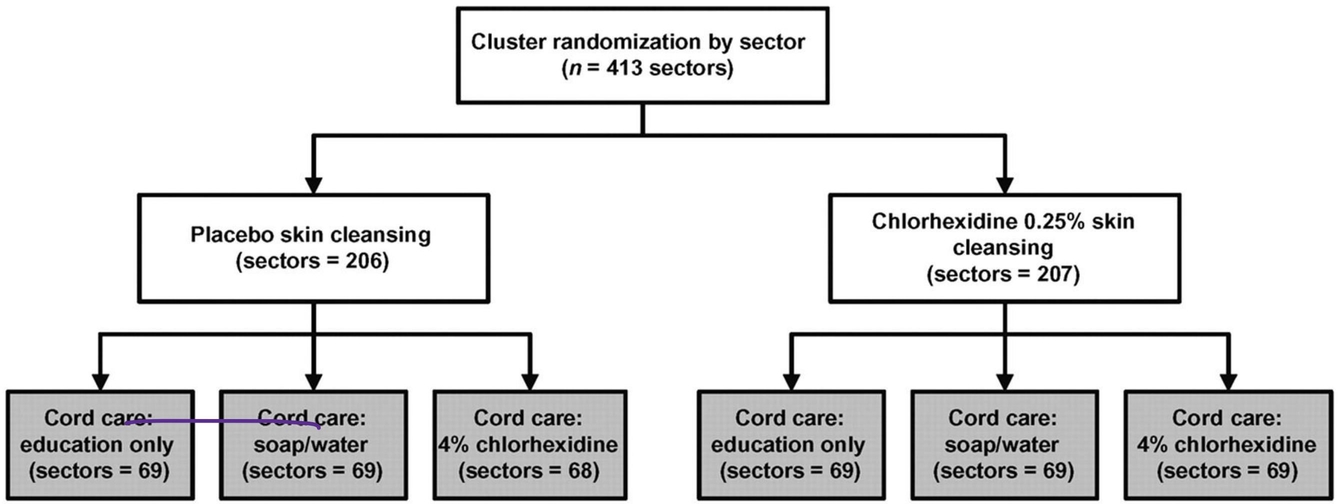
a. The mean difference is significant at the .05 level.

Cluster design

- A cluster randomised controlled trial is a type of randomised controlled trial in which groups of subjects (as opposed to individual subjects) are randomised.

- Cluster randomised controlled trials are also known as cluster randomised trials, group-randomised trials and place-randomized trials.

- Cluster trials originated from educational research. Intact classes or schools were randomised to an intervention or no intervention.


Cluster randomization by sector
(n = 413 sectors)
Placebo skin cleansing
(sectors = 206)
Chlorhexidine 0.25% skin cleansing
(sectors = 207)
Cord care: education only (sectors = 69)
Cord care: soap/water (sectors = 69)
Cord care: 4% chlorhexidine (sectors = 68)
Cord care: education only (sectors = 69)
Cord care: soap/water (sectors = 69)
Cord care: 4% chlorhexidine (sectors = 69)


Cluster/ Group A
Cluster/ Group B
Cluster/ Group C
Intervention A
Intervention B
Intervention C

The responses from individuals within a cluster are likely to be more similar than those from different clusters. This is because individuals within a cluster may share similar characteristics or be exposed to the same external factors associated with membership to a particular cluster.

Observations on participants in the **same cluster tend to be correlated** (non-independent). • Degree of correlation within clusters is known as intracluster correlation coefficient (ρ).

Published by Oxford University Press on behalf of the International Epidemiological Association *International Journal of Epidemiology* 2006;**35**:1292–1300

Advance Access publication 30 August 2006 doi:10.1093/ije/dyl129

METHODOLOGY

Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method

Sandra M Eldridge,[1]* Deborah Ashby[2] and Sally Kerry[3]

design effect can be calculated as was calculated as:

$DE = 1 + \rho (m-1)$

Where m = number of subjects in a cluster and ρ = intra cluster correlation coefficient

ICC= 0.01 which usually, values of between 0.01 and 0.02 in human studies

Cluster allocation

- Need to use some form of stratification.
  - Pairing is often used – match clusters on an important co-variate and randomly allocate a member of each pair to the intervention.
  - Stratification using blocking or the use of minimisation is an alternative.

geepack: Generalized Estimating Equation Package

Package ‘geepack’

December 18, 2020

Version 1.3-2

Title Generalized Estimating Equation Package

Maintainer Søren Højsgaard <sorenh@math.aau.dk>

Description Generalized estimating equations solver for parameters in mean, scale, and correlation structures, through mean link, scale link, and correlation link. Can also handle clustered categorical responses.

Encoding UTF-8

LazyData true

License GPL (>= 3)

NeedsCompilation yes

Depends R (>= 3.5.0), methods

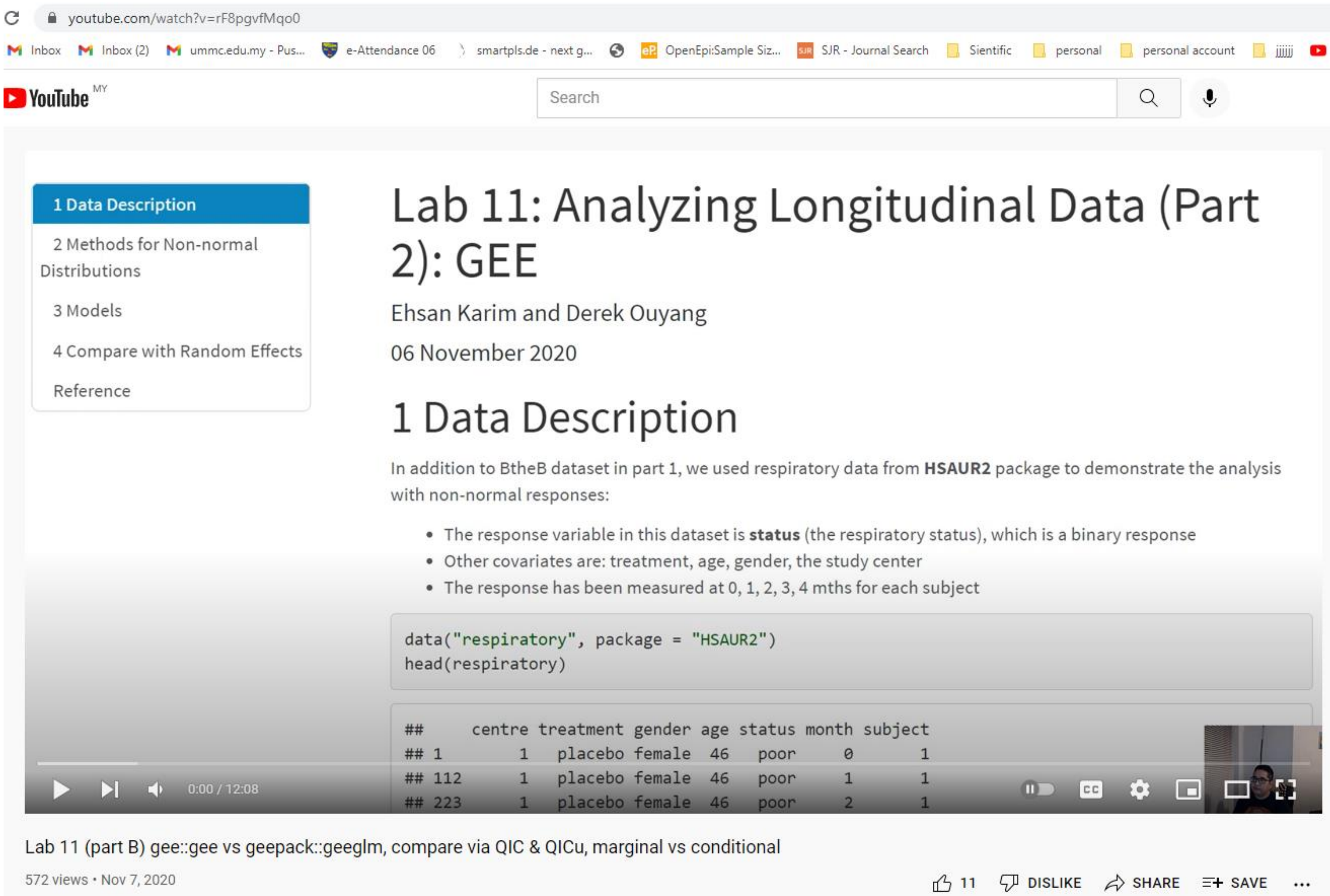
Imports MASS, broom, magrittr

RoxygenNote 7.1.1

Author Søren Højsgaard [aut, cre, cph],
Ulrich Halekoh [aut, cph],
Jun Yan [aut, cph],
Claus Ekstrøm [ctb]

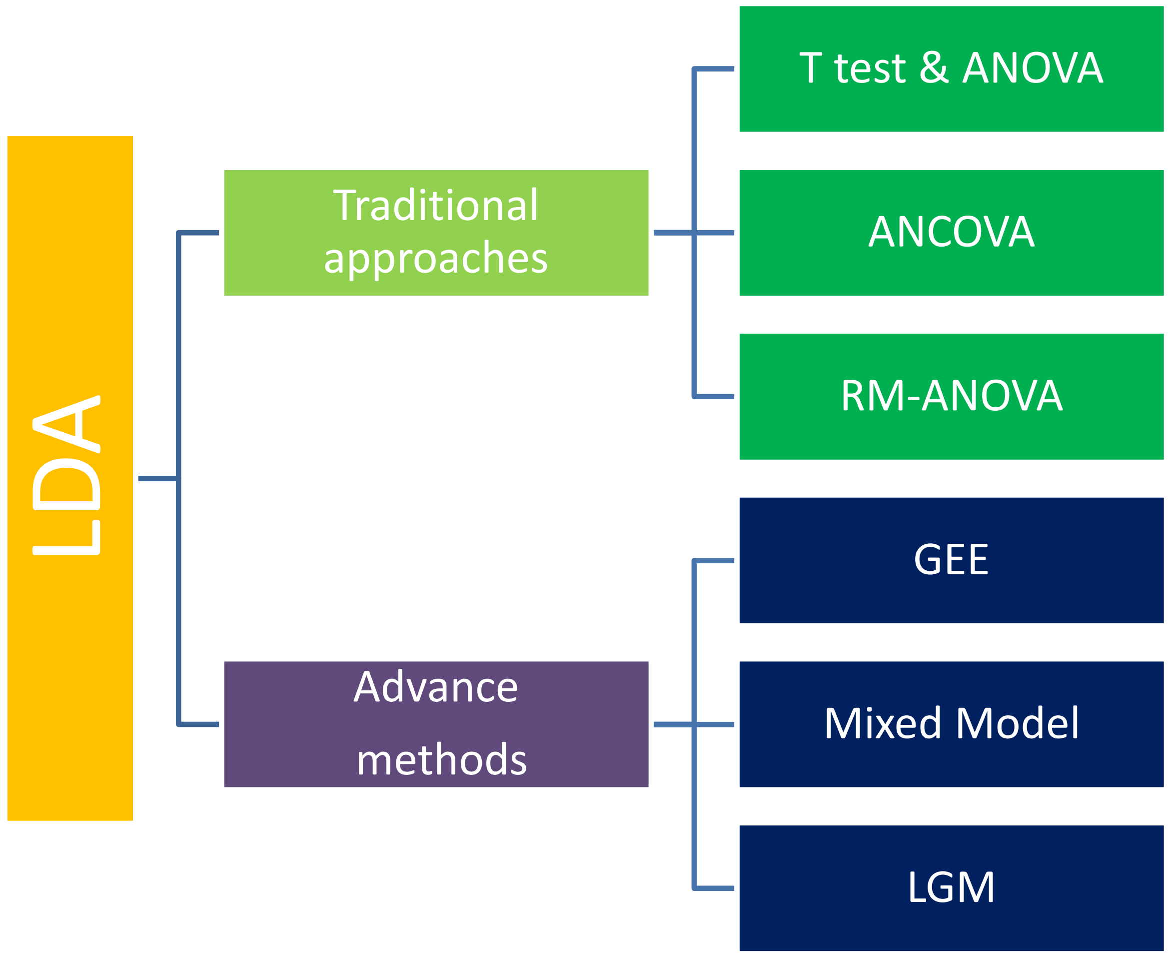
Repository CRAN

Date/Publication 2020-12-18 06:20:11 UTC



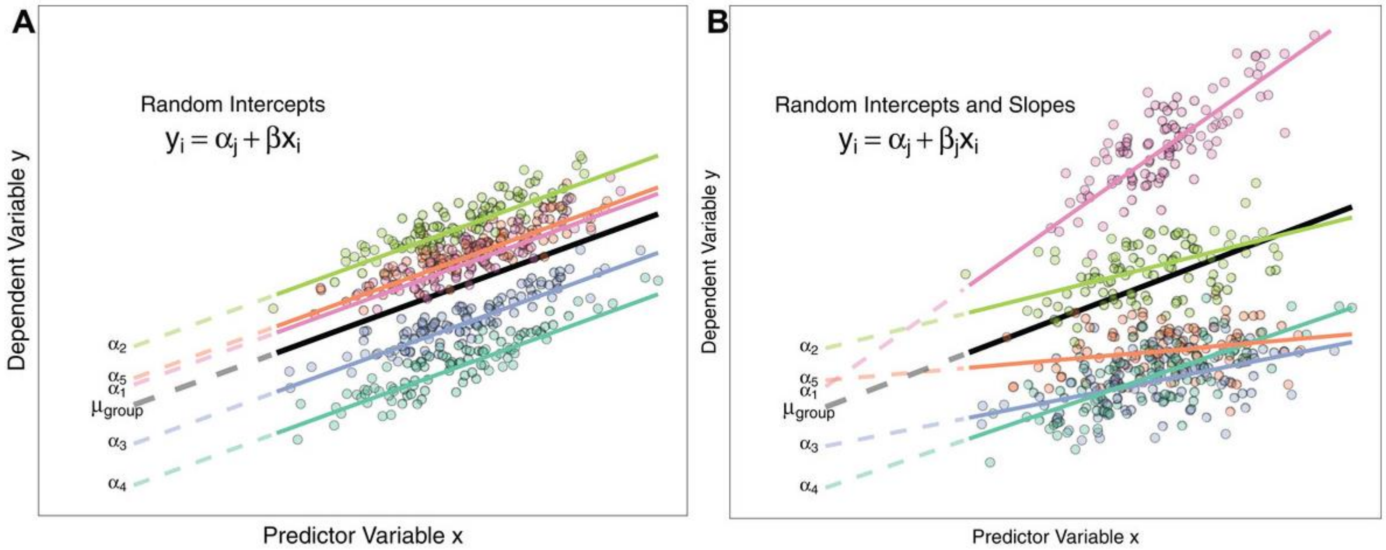
https://www.youtube.com/watch?v=rF8pgvfMqo0

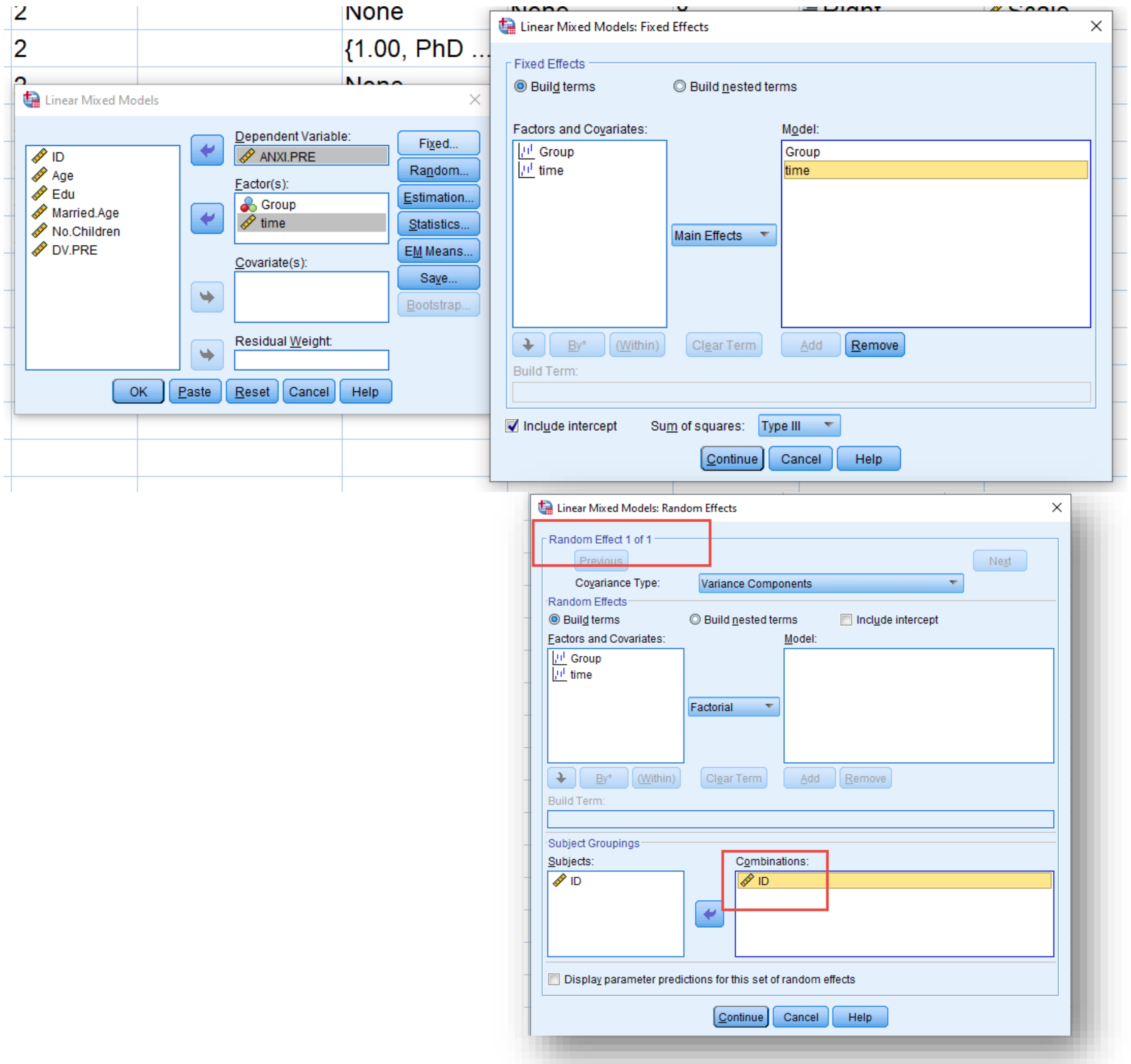
Data analysis approaches



GLMM (Mixed model)

- Mixed effects models are useful **when we have data with more than one source of random variability**.
- For example, an outcome may be measured more than once on the same person (repeated measures taken over time).
- When we do that we have to account for both within-person and across-person variability.


A
Random Intercepts
$y_i = \alpha_j + \beta x_i$
Dependent Variable y
α_2
α_5
α_1
μ_{group}
α_3
α_4
Predictor Variable x
B
Random Intercepts and Slopes
$y_i = \alpha_j + \beta_j x_i$
Dependent Variable y
α_2
α_5
α_1
μ_{group}
α_3
α_4
Predictor Variable x

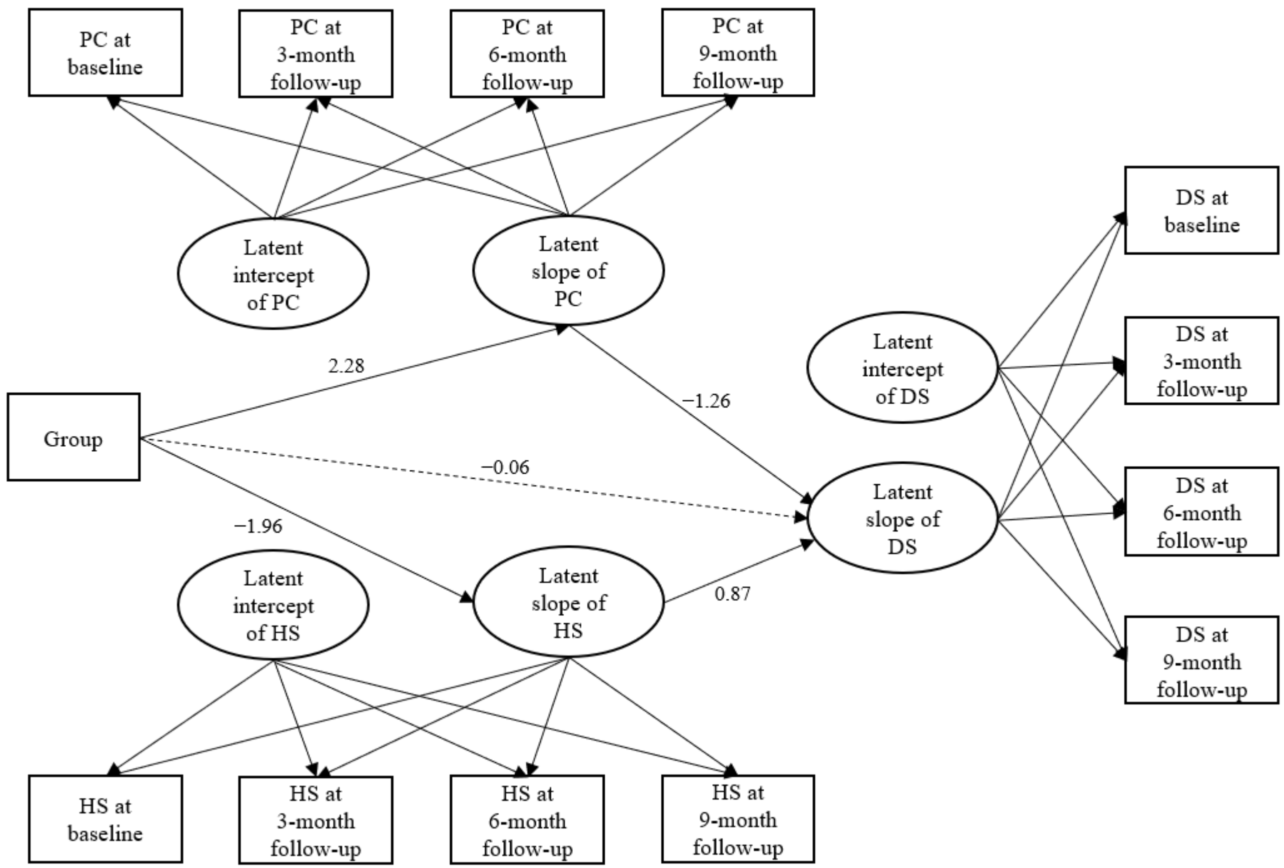

Linear Mixed Models
ID
Age
Edu
Married.Age
No.Children
DV.PRE
Dependent Variable:
ANXI.PRE
Factor(s):
Group
time
Covariate(s):
Residual Weight:
Fixed...
Random...
Estimation...
Statistics...
EM Means...
Save...
Bootstrap...
OK
Paste
Reset
Cancel
Help
Linear Mixed Models: Fixed Effects
Fixed Effects
Build terms
Build nested terms
Factors and Covariates:
Group
time
Main Effects
Model:
Group
time
By*
(Within)
Clear Term
Add
Remove
Build Term:
Include intercept
Sum of squares:
Type III
Continue
Cancel
Help
Linear Mixed Models: Random Effects
Random Effect 1 of 1
Previous
Next
Covariance Type:
Variance Components
Random Effects
Build terms
Build nested terms
Include intercept
Factors and Covariates:
Group
time
Factorial
Model:
By*
(Within)
Clear Term
Add
Remove
Build Term:
Subject Groupings
Subjects:
ID
Combinations:
ID
Display parameter predictions for this set of random effects
Continue
Cancel
Help

Latent growth curve modeling (LGC)

- Latent growth curve modeling is currently one of the most popular approaches used to study longitudinal patterns of change over time
- LGM is a methodology that uses **structural equation modeling** techniques to model individual change, assess treatment effects and the relationship between multiple outcomes simultaneously, and model measurement error.
- The growth curve model (GCM), or latent curve model (Meredith & Tisak, 1990), has been one of the most widely adopted statistical techniques in longitudinal studies to investigate progression over time

Mediators of Intervention Effects on Depressive Symptoms Among People Living With HIV: Secondary Analysis of a Mobile Health Randomized Controlled Trial Using Latent Growth Curve Modeling

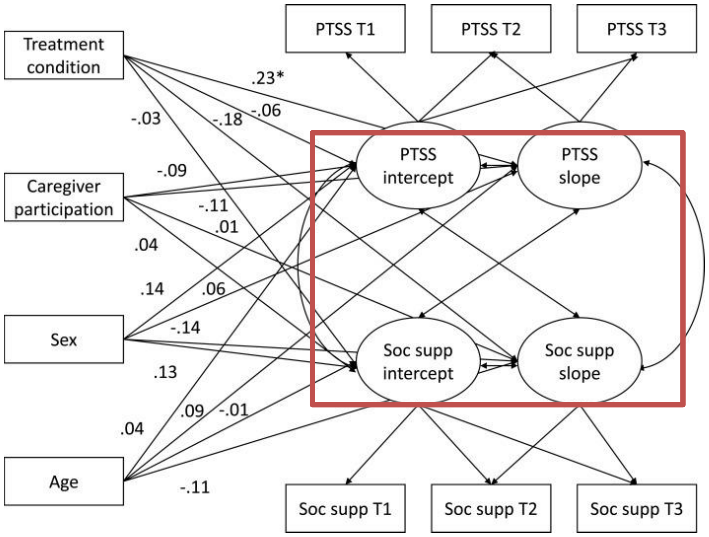
Mengting Zhu [1]; Weiping Cai [2]; Linghua Li [2]; Yan Guo [1, 3, 4]; Aliza Monroe-Wise [5]; Yiran Li [1]; Chengbo Zeng [6, 7]; Jiaying Qiao [1]; Zhimeng Xu [1]; Hanxi Zhang [8]; Yu Zeng [1]; Cong Liu [2]

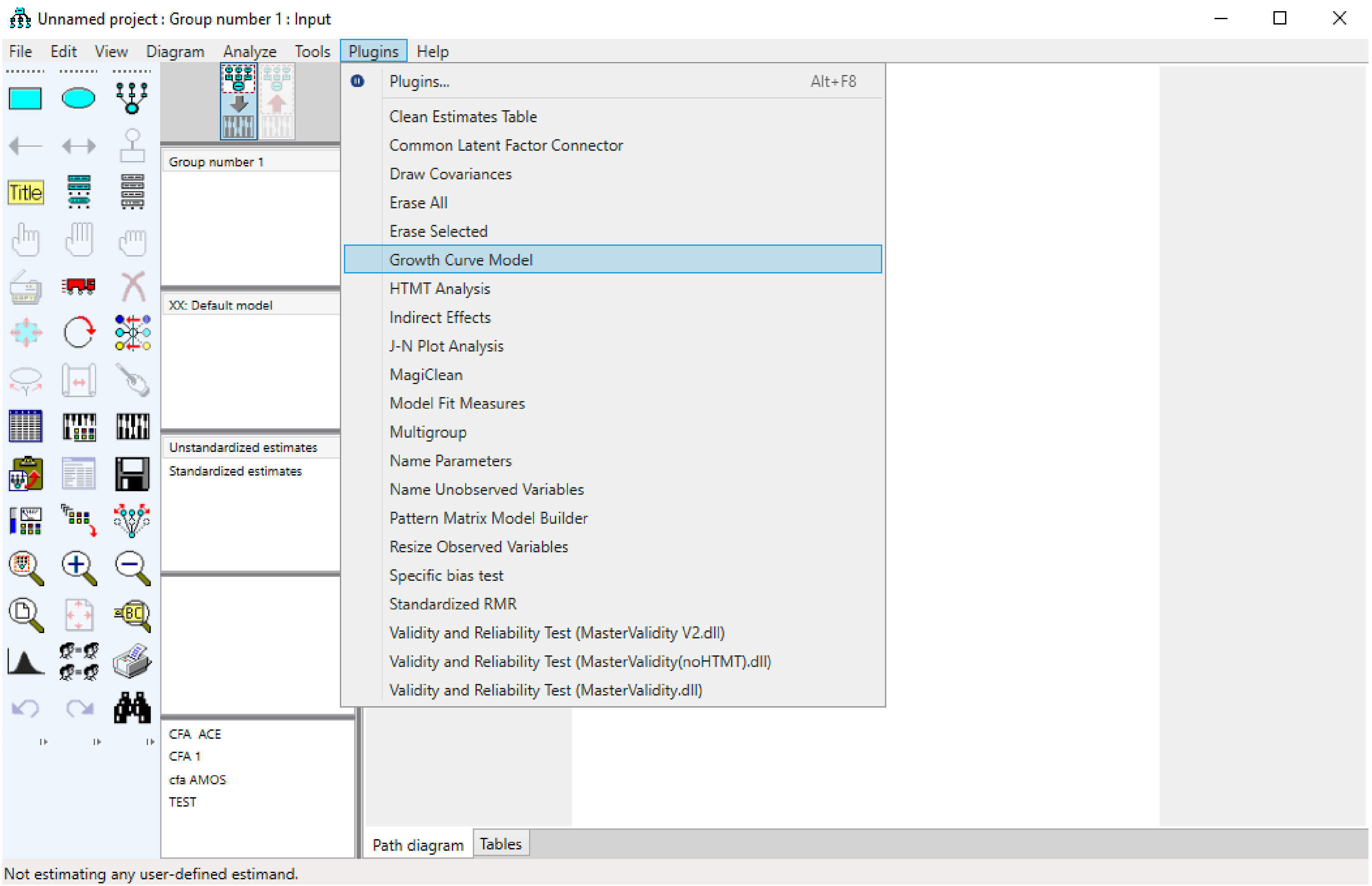


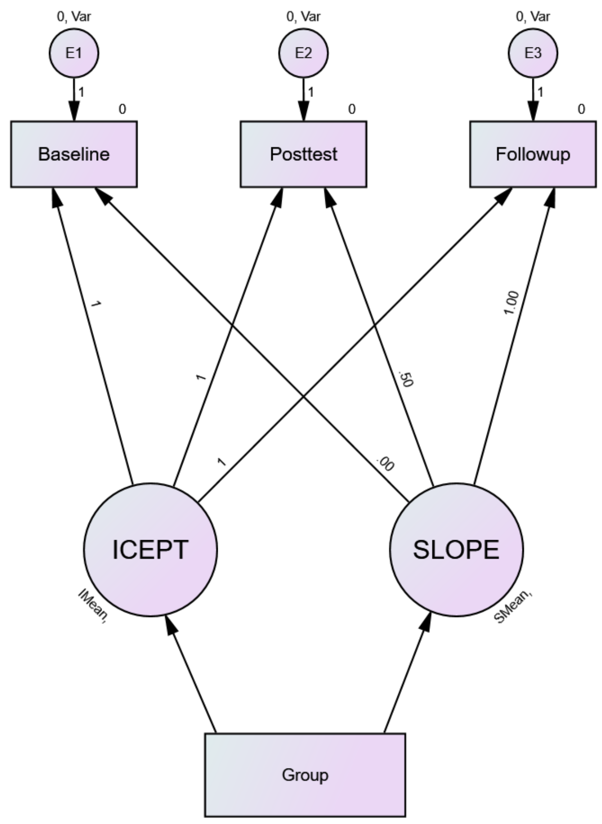
	Mean			Variance		
	Estimate	SE	*p*	Estimate	SE	*p*
Perceived social support						
Intercept	28.00	0.52	<0.001	32.00	3.80	<0.001
Slope T1-T3	1.05	0.29	<0.001	6.25	1.27	<0.001
Slope T3-T5	0.42	0.24	0.077	0	–	
Posttraumatic stress symptoms						
Intercept	27.01	0.61	<0.001	58.22	5.30	<0.001
Slope T1-T3	−6.52	0.48	<0.001	25.14	3.41	<0.001
Slope T3-T5	−1.11	0.46	0.016	0	–	

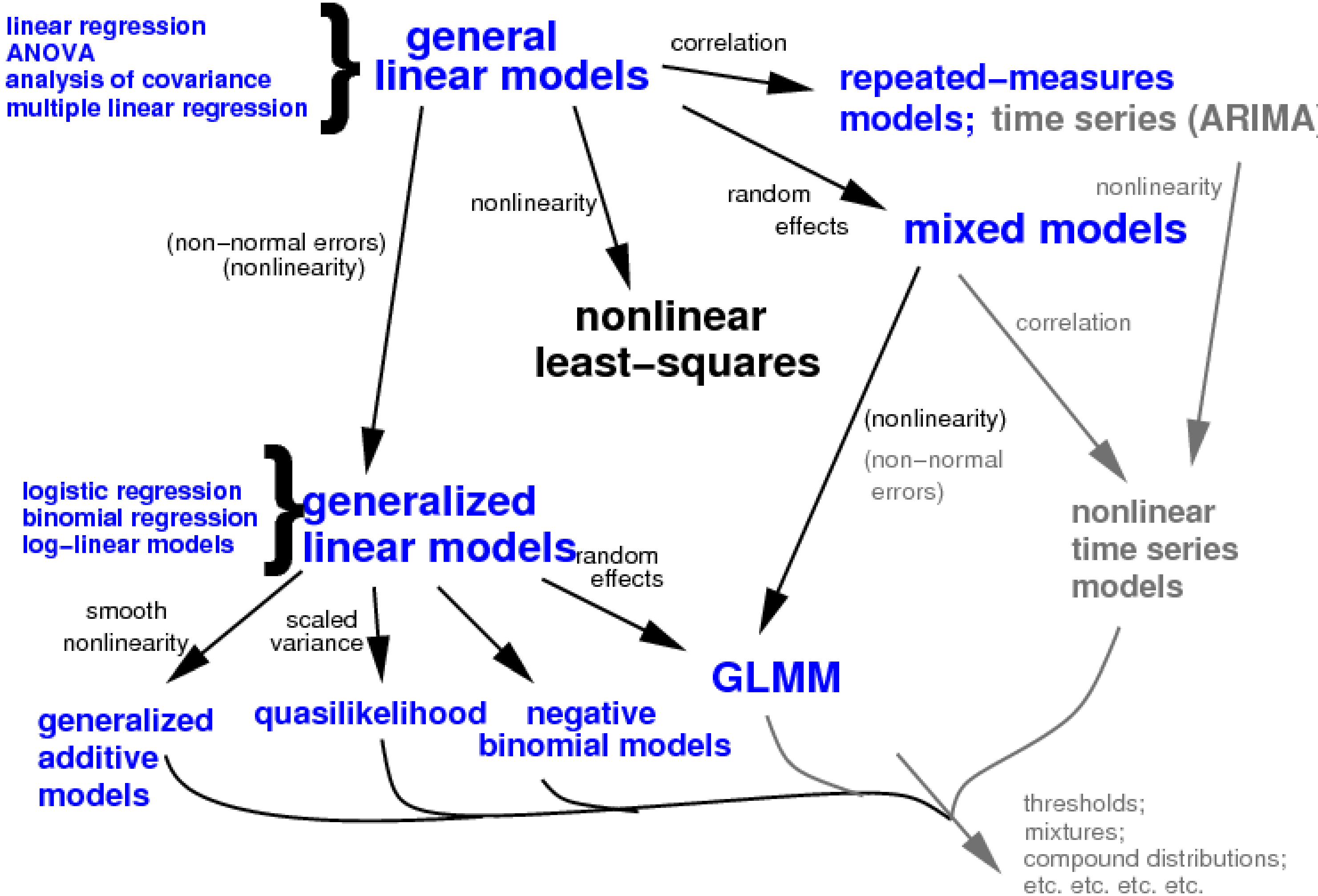
Note: T1 = pretreatment, T3 = post-treatment, T5 = 18 months after post-treatment measure.

	Perceived social support			Posttraumatic stress symptoms		
	Estimate	SE	*p*	Estimate	SE	*p*
Treatment[a] → intercept	−0.03	0.10	0.742	−0.06	0.09	0.509
Treatment[a] → slope	−0.18	0.10	0.197	0.23	0.10	0.025
Caregiver[b] → intercept	0.04	0.10	0.694	−0.09	0.10	0.334
Caregiver[b] → slope	0.01	0.13	0.942	−0.11	0.11	0.326
Sex[c] → intercept	−0.14	0.10	0.742	0.14	0.08	0.091
Sex[c] → slope	0.13	0.10	0.197	0.06	0.10	0.557
Age → intercept	−0.01	0.01	0.900	0.04	0.09	0.648
Age→ slope	−0.11	0.13	0.445	0.09	0.10	0.334




Unnamed project : Group number 1 : Input
File
Edit
View
Diagram
Analyze
Tools
Plugins
Help
Plugins... Alt+F8
Clean Estimates Table
Common Latent Factor Connector
Draw Covariances
Erase All
Erase Selected
Growth Curve Model
HTMT Analysis
Indirect Effects
J-N Plot Analysis
MagiClean
Model Fit Measures
Multigroup
Name Parameters
Name Unobserved Variables
Pattern Matrix Model Builder
Resize Observed Variables
Specific bias test
Standardized RMR
Validity and Reliability Test (MasterValidity V2.dll)
Validity and Reliability Test (MasterValidity(noHTMT).dll)
Validity and Reliability Test (MasterValidity.dll)
Group number 1
XX: Default model
Unstandardized estimates
Standardized estimates
CFA ACE
CFA 1
cfa AMOS
TEST
Path diagram
Tables
Not estimating any user-defined estimand.


0, Var
E1
0, Var
E2
0, Var
E3
1
1
1
0
0
0
Baseline
Posttest
Followup
1
1
1
.00
.50
1.00
ICEPT
SLOPE
IMean,
SMean,
Group



Coral protection from seastars (Culcita) by symbionts (McKeon et al., 2012)

Thank you

- **Dr. Mahmoud Danaee**
- **mdanaee@um.edu.my**

Research method and statistical consultation service (RMC)

- ***Department of Social and Preventive Medicine,***
- ***Faculty of Medicine,University of Malaya,50603 Kuala Lumpur,***
- ***MALAYSIA.***