The Magic of Pretest/Posttest & Longitudinal Data Analysis 6 Week Webinar Class

CLASS 3:

LONGITUDINAL ANALYSIS

Presented by

William M. Bannon, PhD

StatsWhisperer, Inc.

www.statswhisperer.com

Longitudinal Analysis

Welcome to **Week 3 Longitudinal Data Analysis** of our 6-week webinar class *The Magic of Pretest/Posttest & Longitudinal Data Analysis*. Recall, *The 7 Steps of Data Analysis* listed below. Notice our focus will be on Steps 3-6 as each pertains to performing Longitudinal Data Analysis.

The 7 Steps of Data Analysis:

1 Study Map

2 Data Entry

3 Check Data Integrity **FOCUS**

4 Univariate Analysis **FOCUS**

5 Bivariate Analysis **FOCUS**

6 Multivariate Analysis **FOCUS**

7 Write-up and Report

Keep in mind we will be applying these steps to examine sample study 2, which examines:

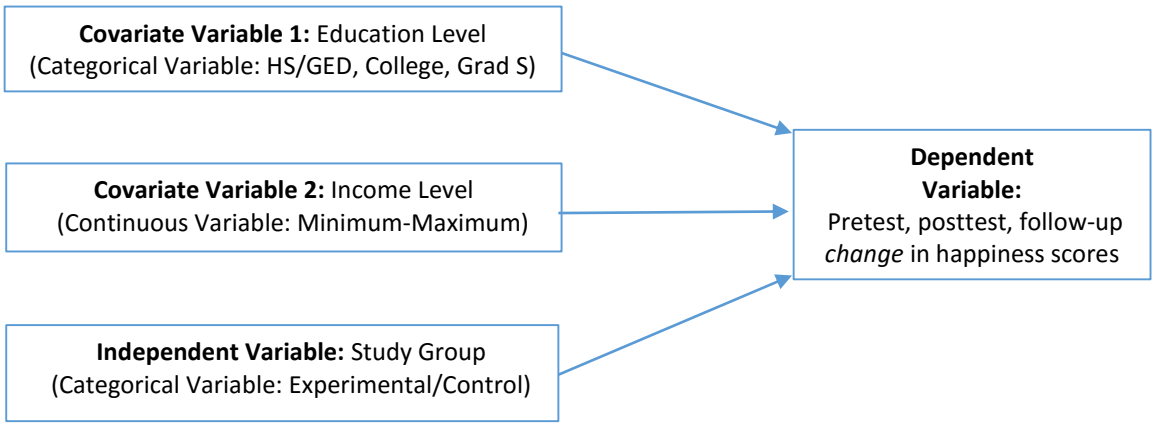
Research Question 1b: Is the *Happiness Intervention* associated with significantly higher happiness scores from pretest, posttest, and follow-up?

Hypothesis: Participants assigned to the *Happiness Intervention* will evidence a significant increase happiness scores, while those in the control will not.

Dependent Var.: Pre/Post/Fu happiness score changes

Independent Var. : Study Group (Experimental/Control)

Covariate Var.: Education, Income



Step 3: Check Data Integrity

What we are asking: Are the data appropriate for statistical analysis?

Recall last week we addressed the necessary checks of data integrity needed for our parametric statistical tests (which were **Multiple Linear Regression** and **Repeated Measures General Linear Model**) in the areas **of Check 2: Test Assumptions** and **Check 4: Measurement Tools**. Since we will use largely the same parametric tests this week in our longitudinal analysis, the **Checks of Data Integrity** this week will be an extension/modification of last week with some needed addendums.

Check 2: Test Assumptions

Recall that last week we describe that our parametric testing would require us meeting the test assumptions of:

A. Normal Distributions (Normality)

B. Influence of outlier scores

C. Linearity

D. Homoscedasticity

E. Multicollinearity

F. Others (Not addressed here)

Also recall that we checked these test assumptions through creating Pretest/Posttest **Happiness Difference Scores** (Posttest Scores – Pretest Scores) and conducting various tests covered on the week 2 lecture. Since we are adding another timepoint for our longitudinal analysis (i.e., follow-up scores), we can examine our tests assumptions through computing another difference, i.e., a posttest to follow-up difference score, and examining that score in the same context as we examined the creating Pretest/Posttest **Happiness Difference Scores** in week 2. To do this simply minus the Posttest Scores from the Follow-up scores via:

- In SPSS go to: *Transform, Compute Variable*
- Under *Target Variable* enter *differencepostfollowup*
- Under *Numeric Expression* enter *followup - posttest*
- Click *OK* or *PASTE* (if *Paste* then run) to produce **Difference Scores**

Next, repeat the same analysis described in week 2 to examine how these difference scores meet the parametric test assumptions.

Check 4: Measurement Tools

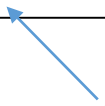
Recall, that in week 2 we describe the need to assess **internal consistency** reliability, which is most often expressed using the Cronbach's alpha, for a composite scale. Therefore, since we are adding a third timepoint measurement of *Happiness*, it would also be necessary to compute and report this metric for the *follow-up* measurement of *Happiness Scores*.

Step 4: Univariate Analysis

Recall in *Univariate Analysis* we are asking: **What are the characteristics of each individual study variable?** Logically, our longitudinal analysis involves adding a single timepoint measurement, so we would need to include that measure to our univariate (a.k.a., descriptive) analysis, as in Figure 1:

Figure 1 Univariate analysis of continuous outcome variables

Variable	Mean (SD)	Min/ MAX	Scale Range
Happiness Pretest	2.54 (.65)	1.0-3.5	1.0-5.0
Happiness Posttest	2.72 (.62)	1.5-3.8	1.0-5.0
Happiness Follow-Up	2.66 (.62)	1.5-3.9	1.0-5.0



Step 5: Bivariate Analysis

Recall, in the same manner as last week, we will use **Bivariate Analysis** to assess which predictor variables (Education Level, Income Level, & Study Group) are significantly associated with our dependent variable (i.e., for this analysis: changes in between Pretest, Posttest, and Follow-up scores over time). Keep in mind, all predictors that evidence a statistically significant relationship with the outcome will be included in **Step 6: Multivariate Analysis**.

As we approach inferential analysis, we might recall that in the week 2 session we analyzed 2 timepoint scores (Pretest & Posttest) using Repeated Measures GLM. This might suggest that for longitudinal analysis, we might just add a third timepoint (e.g., Follow-up) and repeat the Repeated Measures GLM. This would be quite correct.

However, we must also keep in mind that the Repeated Measures GLM does not address several issues in longitudinal analysis we must consider. In terms of magic, longitudinal analysis is a more complex trick. In order to address these issues in addition to the Repeated Measures GLM, we will also conduct a parallel *Linear Mixed-Effects Model* to carry out a *Repeated Measures Analysis*.

For example, in terms of shortcomings. the Repeated Measures GLM treats a participant's scores over time as independent. However, participant scores over time are frequently correlated (autocorrelation). The *Linear Mixed-Effects Model* will be able to control for this autocorrelation.

Furthermore, *Linear Mixed-Effects Model* can treat a variable as a random factor (e.g., Subject, Subject traits like Gender, etc.), which is also beneficial to an analysis. Consider that in our analysis we are looking at the *Direct Effects* between the independent Variable Study Group and Dependent Variable Changes in Happiness Scores over time. Within this direct relationship, there are all sorts of differences in the IV and DV by other factors, including the study participants. The *Linear Mixed-Effects Model* can adjust for these factors by treating them as random, so the direct effects between the IV/DV is clearer.

Bivariate Analysis with the Repeated Measures GLM

We will now use **Repeated Measures GLM** as in bivariate analysis to identify which predictors are related to pretest, posttest, and follow-up changes over time. Subsequently, for each predictor in SPSS by going to:

Education Level and Outcomes Over Time

- *Analyze, General Linear Model, Repeated Measures*
- *Enter 3 in number of levels & a variable name in the Within Subject Factor Name box, e.g., Time (no space)*
- *In Within Subjects Var., enter Pretest, Posttest, Follow-Up*
- ***In Covariate(s) box, enter Education Level***
- *Options = Descriptives, Effect Size, Obs. Power*
- *Click OK or PASTE (if Paste then run)*

Within the output, you will see within the box labeled *Multivariate Tests*, within the column labeled Sig., the line labeled Education*Time evidenced a statistically significant p value of .000 for all statistics including *Wilks Lambda* (which is a commonly reported statistic).

Income Level and Outcomes Over Time

- *Analyze, General Linear Model, Repeated Measures*
- *Enter 3 in number of levels & a variable name in the Within Subject Factor Name box, e.g., Time (no space)*
- *In Within Subjects Var., enter Pretest, Posttest, Follow-Up*
- ***In Covariate(s) box, enter Income Level***
- *Options = Descriptives, Effect Size, Obs. Power*
- *Click OK or PASTE (if Paste then run)*

Within the output, you will see within the box labeled *Multivariate Tests*, within the column labeled Sig., the line labeled Income*Time evidenced a statistically significant p value of .000 for all statistics including *Wilks Lambda* (which is a commonly reported statistic).

Study Group and Outcomes Over Time

- *Analyze, General Linear Model, Repeated Measures*
- *Enter 3 in number of levels & a variable name in the Within Subject Factor Name box, e.g., Time (no space)*
- *In Within Subjects Var., enter Pretest, Posttest, Follow-Up*
- ***In Between Subjects Factor(s) box, enter Study Group***
- *Options = Descriptives, Effect Size, Obs. Power*
- *Click OK or PASTE (if Paste then run)*

Within the output, you will see within the box labeled *Multivariate Tests*, within the column labeled Sig., the line labeled StudyGroup*Time evidenced a statistically significant p value of .000 for all statistics including *Wilks Lambda* (which is a commonly reported statistic).

Step 6: Multivariate Analysis

Next, we may approach multivariate analysis using the Repeated Measures GLM by going to:

Education Level, Income Level & Study Group

and Outcomes Over Time

- *Analyze, General Linear Model, Repeated Measures*
- *Enter 3 in number of levels & a variable name in the Within Subject Factor Name box, e.g., Time (no space)*
- *In Within Subjects Var., enter Pretest, Posttest, Follow-Up*
- ***In Covariate(s) box, enter Education Level & Income Level***
- ***In Between Subjects Factor(s) box, enter Study Group***
- *Options = Descriptives, Effect Size, Obs. Power*
- *Click OK or PASTE (if Paste then run)*

Figure 2 presents the **Descriptives Table** for the **Repeated Measures GLM** multivariate analysis. It is clear at pretest, the mean Happiness scores for the Experimental and Control group were alike (2.54 vs. 2.54). However, at posttest while the mean score for the Control group remained the same (2.54), the Experimental group increased to a mean score of 2.90. Finally, at posttest However, at posttest the mean score for the Control group remained the same (2.51), the Experimental group made a slight decrease with a mean score of 2.81.

Figure 2

Descriptives table for the Repeated Measures GLM

Study Group		Mean	Standard Deviation	N
Pretest Happiness Scores Total	Control	2.54	.64	50
	Experimental	2.54	.66	50
		2.54	.65	100
Posttest Happiness Scores Total	Control	2.54	.62	50
	Experimental	**2.90**	.57	50
		2.72	.62	100
Follow-Up Happiness Scores Total	Control	2.51	.61	50
	Experimental	**2.81**	.60	50
		2.66	.62	100

The main change in scores = Experimental group at Posttest

Figure 3 presents the **Statistical Findings of Repeated Measures GLM** multivariate analysis. We see all the predictors except *Education Level* are significantly related to pretest, posttest, and follow-up changes in happiness scores. Furthermore, the independent *Study Group* evidences the strangest and considerably large PES effect size upon the outcome variable. Lastly, we see that in terms of statistical power, only *Education Level* has less than 80% power.

Figure 3

Statistical Findings of Repeated Measures GLM

Effect	F	Sig.	Partial Eta Squared	Observed Power
PrePostChange	10.9	.001	.10	.90
PrePostChange*Income	11.4	.001	.11	.92
PrePostChange*Education	2.1	.15	.02	.30
PrePostChange*Studygroup	67.2	.000	.41	1.00

All Statistically Significant, Except Education

Study Group has the largest PES

Sufficient power (.80) for each predictor, Except Education

Next, we may approach multivariate analysis using a *Linear Mixed Effects Model* in SPSS by going to:

- *Data, Restructure,*
- ***Welcome to the Data Restructure Wizard:*** *Restructure selected variables into cases, Next*
- **Variables to Cases: Number of Variable Groups:** *How many variable groups do you want to restructure?* Select *One* (Happiness),
- Click *Next*

In the **Variables to Cases: Select Variables dialogue** box select:

- *Case Group Identification, Use Selected Variable, Variable:* Enter *Study Participant ID [ID]*
- *Variables to be Transposed, Target Variable,* Type *Happiness*
- Enter *Pretest, Posttest, Follow-up*
- *Fixed Variable(s): Education Level, Income Level, & Study Group*
- Click *Next*

In the **Variables to Cases: Create Index Variables** box select:

- *How many variables do you want to create?* Select *One*, Click *Next*
- *What kind of index values?* Click *Sequential Numbers*
- *Edit the index variable name and label,* Enter *Timepoint* (a label if you want)
- Click *Finish* (Save database: another name)

Once the data are restructured we can now implement *Mixed Effects Regression*. We can do this in SPSS by going to:

- *Analyze, Mixed Models, Linear*
- *Subjects Level Data* Enter *ID*
- *Repeated* Enter *Timepoint* (Index Variable)
- *Repeated Covariance Type* Enter *AR(1): Heterogeneous* (correlated timepoints)
- Click *Continue*

In the *Linear Mixed Models* box:

- Enter Dependent Variable *Happiness*
- In Factor(s) enter *Timepoint, Study Group*
- Click *Fixed* box, add all predictors, click *Continue*

Longitudinal Analysis

- Click *Random* box, check *Include Intercept*, under *Subject groupings* enter *ID* variable (Combinations), click *Continue*
- Click *Estimation* box, click *Restricted Maximum Likelihood* (w/random), *Continue*
- Click *Statistics* box, click *Parameter Estimates* & *Tests for Covariance Parameters*, click *Continue*
- Click *OK/Paste* (then run)

Figure 4 presents the results of the Linear Mixed-Effects model. In short we find once the autocorrelation and random effects are addressed, the independent variable *Study Group* remained statistically significant.

Figure 4

Statistical Findings of Repeated Measures GL

Type III Tests of Mixed Effects

Source	Numerator DF	Denominator DF	F	Sig.
Intercept	1	96.12	209.09	.000
Timepoint	2	99.94	28.61	.000
Study Group	1	95.96	6.10	**.015**
Education	1	95.96	1.02	.31
Income	1	95.96	.023	88

The IV Study Group Remains Statistically Significant,