GLM Repeated Measures ANOVA

Caspar J. van Lissa

2023-08-12

Within-participants Designs

- Participants are exposed to multiple experimental conditions
  - Type of stimulus, drug dosage
- The outcome is measured at two or more measurement occasions (*longitudinal design*).
  - Test-retest designs, panel studies, diary studies, repeated physiological assessments, etc

Advantages Within-participants Designs

Many individual differences are constant

- Variability due to individual differences is thus removed from the error term
- Each subject serves as its own control
- Greater statistical power
- More information per participant = more efficient; we can use a smaller sample than if we used a between-participants design

Limitations of Within-participants Designs

For experimental designs:

- Order effects: The order of conditions may have an effect
- **Differential** order effects: Order effects may differ across different orders
    - If participants take the drug before the placebo, they may still be under the influence during the placebo condition

Solution for Order Effects

Latin Square Design

- Experimentally controls for order effects
- In a Latin square, each condition appears in one position in the ordering:

1	2	3	4
A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

This is just one out of many (in fact 256; 44) possible Latin Squares. There are tools that randomly generate a Latin square.

Limitations of Within-participants Designs

For all designs, including non-experimental:

- Learning effects: People become familiar with your questionnaire
- Historical effects: Some event may happen during your study (fire alarm goes off, global pandemic breaks out, documentary on TV about your topic of study)
- Effect of time is confounded with effect of condition
- Effect of time may have a clearly defined functional form which RM-ANOVA is ignoring
    - E.g.: PTSD probably changes *after* deployment, and then increases or decreases over time
    - Techniques like “Structural Equation Modeling” allow you to describe this

Two Repeated Measurements

An intervention is imposed on ten people
Each person is measured twice:

- Before the intervention (pretest)
- After the intervention (posttest)

Resp.	Pretest	Posttest
1	2	5
2	3	4
3	4	6
4	5	5
5	6	8
6	7	10
7	8	9
8	9	11
9	10	9
10	11	15
Mean	6.5	8.2

How to Analyze These Data?

Problem: These data violate one assumption of the general linear model:

- Independence of errors

So we can’t use linear regression, or any of its “interfaces” like the independent samples t-test

Solution: Paired Samples t-test

Solution: With just two repeated measures, the *paired samples t-test*!
This is equivalent to calculating the *difference between the two measurements* and…

- Performing a one-sample t-test on that difference
- Performing linear regression with that difference score as outcome and only an intercept as predictor

Paired Samples t-test

Resp.	Pretest	Posttest	Difference (post-pre)
1	2	5	3
2	3	4	1
3	4	6	2
4	5	5	0
5	6	8	2
6	7	10	3
7	8	9	1
8	9	11	2
9	10	9	-1
10	11	15	4
Mean	6.5	8.2	1.7

Results

t-test:

Paired Samples Test

		Paired Differences					t	df	Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			One-Sided p	Two-Sided p
Pair 1	t1 - t2	-1.70000	1.49443	.47258	-2.76905	-.63095	-3.597	9	.003	.006

Regression:

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	int	1.700	.473	.768	3.597	.006

a. Dependent Variable: dt

b. Linear Regression through the Origin

More than Two Measurements

Example: PTSD in Military Personnel

- Data on 978 Dutch military personnel who have been deployed
- 4 repeated measures of PTSD symptoms on the SCL scale
  - 1 pre-deployment, 3 every 6 months post-deployment

3+ Repeated Measurements

Research question : is there a difference between those repeated measures?
$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4; H_1$: not H_0
Two approaches:

1. Univariate Approach (aka: Linear Mixed model)
    - Just linear regression
    - The repeated measures are treated as one outcome variable
    - Each participant has multiple rows
2. Multivariate Approach
    - Treating the repeated measures as different correlated outcomes

Liner Mixed Model

Uses Linear Regression:

- Treat all repeated measurements as a single variable with repeated observations
  - Convert data to “long format”
  - 4 repeated measures -> 4 data rows per participant
- Measurement occasion (or: condition) is a “fixed effect” (= limited number of discrete values)
- Participant ID is a “random effect”; each participant may vary around a person-specific mean

This only works if the so called **sphericity** assumption holds

Sphericity Assumption

> **Sphericity:** The variances of the differences between all combinations of repeated measures are equal.

This is analogous to the homoscedasticity assumption
Closely related to the notion of compound symmetry:

- All repeated measures have equal variance
- Each pair of repeated measures have the same correlation
- You see how this justifies treating repeated measures as a single long-format dependent variable

Multivariate Approach

If we can not / do not assume sphericity, we can use the *multivariate approach*

- Repeated measurements are treated as covariates of each other
  - I.e., effect on T1 controlling for T2, T3, …
  - Effect on T2 controlling for T1, T3, …
- Because of this, this approach has more predictors and thus much smaller df
  - Needs larger sample to get the same power

Mauchly’s test of Sphericity

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
time	.951	48.560	5	<.001	.967	.970	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

- H_0: sphericity holds, H_1: sphericity does not hold
- Significant test: evidence that the assumption is violated
- This is the case here, possibly because of the qualitative difference between pre- and post-deployment measurements

Mauchly’s test of Sphericity

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Epsilon[b] Huynh-Feldt	Epsilon[b] Lower-bound
time	.951	48.560	5	<.001	.967	.970	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

Epsilon is an estimate of departure of sphericity

- If sphericity holds, epsilon = 1
- Lower value of epsilon -> larger departure from sphericity
- Lower bound of epsilon: $1/(k-1)$
    - k is the number of repeated measures

Violation of Sphericity

- Remember: You don’t have to blindly adjust your test based on assumption checks
- You should definitely disclose it
- You can choose a test that is more robust to violations of sphericity

Mixed Model Results

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
time	Sphericity Assumed	87.432	3	29.144	7.285	<.001
	Greenhouse-Geisser	87.432	2.900	30.150	7.285	<.001
	Huynh-Feldt	87.432	2.909	30.051	7.285	<.001
	Lower-bound	87.432	1.000	87.432	7.285	.007
Error(time)	Sphericity Assumed	11726.318	2931	4.001		
	Greenhouse-Geisser	11726.318	2833.201	4.139		
	Huynh-Feldt	11726.318	2842.538	4.125		
	Lower-bound	11726.318	977.000	12.002		

- Notice error df: 2931
- You can use a corrected test
- Trade-off between Type I error (false-positives) and Type II error (false negatives) by adjusting df:
  - Sphericity assumed: Highest Type I error, lowest Type II error
  - Huynh-Feldt: Slightly lower Type I error slightly higher Type II error

- Greenhouse Geisser: Slightly lower Type I error, slightly higher Type II error
- Lower bound: Lowest Type I error, highest Type II error

Multivariate Approach

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
time	Pillai's Trace	.021	7.079[b]	3.000	975.000	<.001
	Wilks' Lambda	.979	7.079[b]	3.000	975.000	<.001
	Hotelling's Trace	.022	7.079[b]	3.000	975.000	<.001
	Roy's Largest Root	.022	7.079[b]	3.000	975.000	<.001

- Notice error df: 975 (much lower)

Mixed Design

Mixed Design

Within-participants Factor:

- Time, experimental condition, etc
- In our example: Time

Between-participants Factor:

- Sex, age, major, etc
- In our example: Whether the participant was exposed to high-intensity combat action (1) or not (0)

Mixed Design ctd

This is essentially a factorial design: 4(Time: T1, T2, T3, T4) x 2(Combat exposure: Low vs. High)

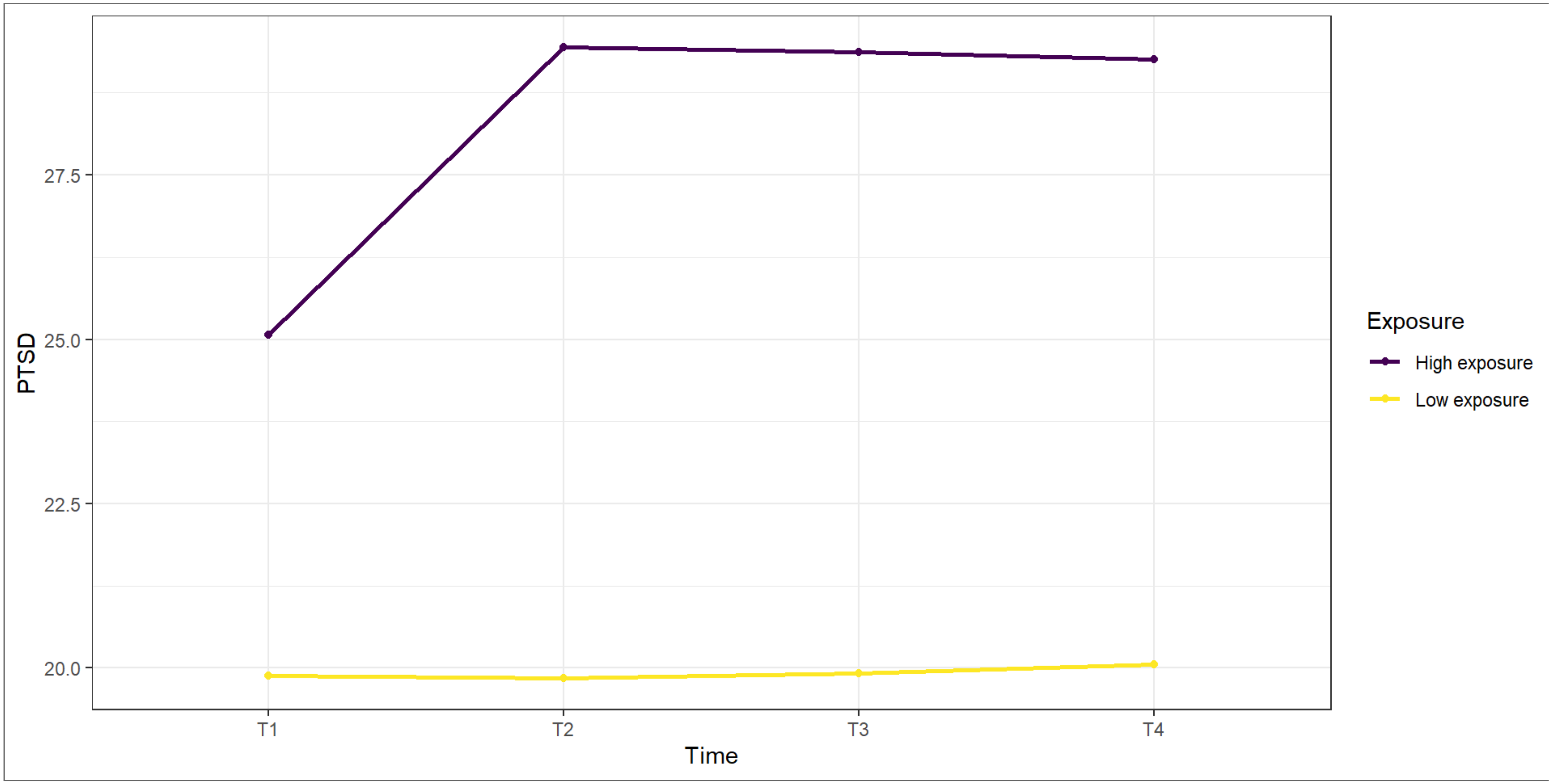
	T1	T2	T3	T4
Low exposure	19.89	19.84	19.92	20.06
High exposure	25.07	29.44	29.37	29.25

Interactions

This requires testing an interaction effect between time and exposure

- If there is a significant interaction, you can use simple effects analysis:
  - Test whether the within-participants factor has an effect for each level of the between-participants factor.
  - Test whether the between-participants factor has an effect for each level of the within-participants factor.

Graphical display



Test for interaction

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
time	Sphericity Assumed	783.088	3	261.029	69.637	<.001
	Greenhouse-Geisser	783.088	2.948	265.593	69.637	<.001
	Huynh-Feldt	783.088	2.961	264.432	69.637	<.001
	Lower-bound	783.088	1.000	783.088	69.637	<.001
time * Exposure	Sphericity Assumed	750.910	3	250.303	66.776	<.001
	Greenhouse-Geisser	750.910	2.948	254.679	66.776	<.001
	Huynh-Feldt	750.910	2.961	253.566	66.776	<.001
	Lower-bound	750.910	1.000	750.910	66.776	<.001
Error(time)	Sphericity Assumed	10975.407	2928	3.748		
	Greenhouse-Geisser	10975.407	2877.694	3.814		
	Huynh-Feldt	10975.407	2890.323	3.797		
	Lower-bound	10975.407	976.000	11.245		

Significant interaction, we could perform simple effects tests

Follow-up: Simple Effect of Time

Multivariate Tests

Exposure		Value	F	Hypothesis df	Error df	Sig.
Low	Pillai's trace	.007	2.148[a]	3.000	974.000	.093
	Wilks' lambda	.993	2.148[a]	3.000	974.000	.093
	Hotelling's trace	.007	2.148[a]	3.000	974.000	.093
	Roy's largest root	.007	2.148[a]	3.000	974.000	.093
High	Pillai's trace	.172	67.614[a]	3.000	974.000	<.001
	Wilks' lambda	.828	67.614[a]	3.000	974.000	<.001
	Hotelling's trace	.208	67.614[a]	3.000	974.000	<.001
	Roy's largest root	.208	67.614[a]	3.000	974.000	<.001

- Not significant in low exposure group
- Significant in high exposure group

Follow-up: Simple Effect of Exposure

Univariate Tests

Measure: MEASURE_1

time		Sum of Squares	df	Mean Square	F	Sig.
1	Contrast	1488.785	1	1488.785	368.540	<.001
	Error	3942.732	976	4.040		
2	Contrast	5108.932	1	5108.932	718.764	<.001
	Error	6937.348	976	7.108		
3	Contrast	4959.123	1	4959.123	755.419	<.001
	Error	6407.176	976	6.565		
4	Contrast	4691.217	1	4691.217	722.939	<.001
	Error	6333.356	976	6.489		

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

- Significant difference at each time
- We could apply Bonferroni correction, but these p-values are very small (< .001)

Error

×