Longitudinal Data - Features, Vocabulary, Models

Level of Analysis: Between-Person and Within-Person Relationships

Between-Person refers to the existence of interindividual variation

↳ individuals can differ on "stable" attributes
eg: ethnicity, biological sex

↳ individuals can also differ on attributes that can potentially change over time
eg: intelligence, height, personality

↳ BUT if these attributes are assessed at a single time point, we assumed they are also "stable"

Stable attributes are considered time-invariant

↳ Between-person relationships refer to interindividual differences in variables (attributes) that are time-invariant

We can't always assume variables remain constant over time, or we specifically want to measure a change within an individual over time

Within-Person refers to intraindividual variation when a person is measured over time (or under some other condition)

↳ this variation is only directly observable when each person is measured more than once (i.e. longitudinal study)

Attributes that are expected to change over time are called time-varying

↳ Within-person relationships refer to intraindividual differences in variables that are time-varying

Some common model terminology:

between-person analysis ⟶ level 2 or macro-level analysis

within-person analysis ⟶ level 1 or micro-level analysis

Longitudinal variables (i.e. variables we measure multiple times) usually contain both between-person and within-person variation

↳ variables measured over time are usually really two variables (not just one)

↳ allows us to test hypotheses at multiple levels (macro + micro) simultaneously

Example: We are looking into the link between physical activity and bone density. This can be explored at a macro ~~[illegible]~~ and micro level:

Macro level (between-persons)

↳ We find people who are more active generally have denser bones

This is an interperson relationship where we are comparing a cross-section of individuals

↳ we ~~[illegible]~~ could have assessed each person once, or taken multiple measures over time and averaged them (this is still between-person)

Micro level (within-person)

↳ We collect repeated measures over time and as people change their activity level

↳ with this information we could examine the extent to which bone density changes with more or less activity

This is an intraperson relationship where each person's baseline serves as his or her own reference.

Note

Fundamentally, we must be aware that relationships observed at the within-person level do not necessarily mirror those observed at the between-person level of analysis

↳ in longitudinal analysis we aim to separate these two phenomena of interest.

Breakdown of within-person variation

- Total variation in data
  - Intraindividual Variation
    - Within-person Change
      - Systematic change we expect to observe as a result of the "meaningful" passage of time (eg: time of treatment)
        These changes may manifest in different patterns or rates across people.
        The goal is to predict and describe individual differences in change over time (eg: which people benefit the most from a treatment?)
    - Within-person Fluctuation
      - undirected variation over repeated assessments when one would not expect any systematic change (like within-person error)
        Eg: we measure someone's reaction time 10 times without any changes to the person or study design. We expect some fluctuation in their reaction times
  - Interindividual Variation
    - discussed in other models

Features of Longitudinal Models

- these models have <u>two sides</u>:
  1) Model for the Means
  2) Model for the Variance

- Model for the Means states how the expected (or predicted) outcome (Y) for each person varies as a function of his or her predictor values (X)
  - if you know nothing about X, a naive guess for Y would be the grand mean $\hat{Y}$ of the whole sample
  - if you know one or more predictors (X) you can now expect their Y will be the group mean $\hat{Y}_k$ (also called conditional mean)
  - Predictors can be continuous, categorical, or a mix of both.

- Model for the variance refers to stochastic error in the model, and describes how the residuals of the Y outcomes are distributed and related across observations

  usually we just make assumptions about our model for the variance. ie we assume homogeneity of variance, normal distribution of residuals, and unrelated across persons.

Repeated Measures ANOVA - this is the simplest longitudinal model.

↳ also referred to as a within-subjects ANOVA or ANOVA of correlated samples

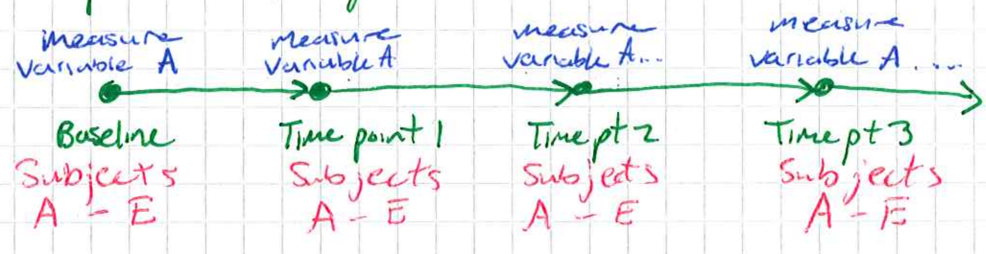
This model is essentially equivalent to the one-way ANOVA, but for related, not independent groups

↳ repeated measures of the same subjects over time

For the one-way repeated measures ANOVA, we require that:

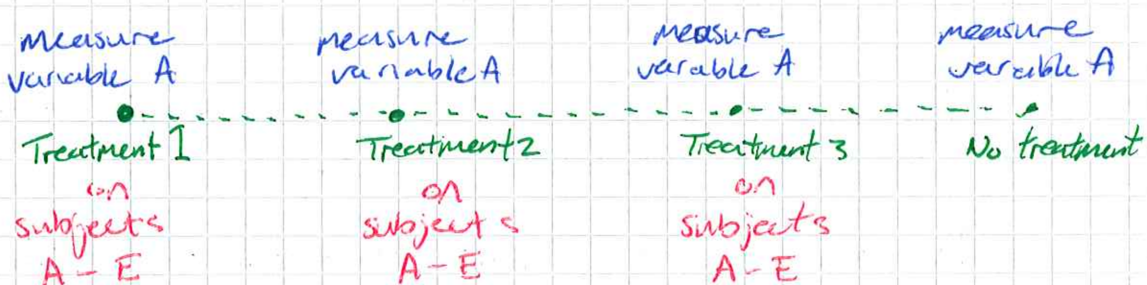
- The dependent variable needs to be continuous
- The independent variable needs to be categorical
  - ↳ called levels or related groups
  - ↳ This is usually time, and each level is a specific time point of interest.

this is a longitudinal study set-up



↳ the independent variable can also be a different condition, and each condition is a level (sometimes called "treatment")

this could be a setup for drug clinical trial where patients change treatments (double-dummy)



In either study set-up the same subjects are measured multiple times (either under different treatments/conditions or at different time points) and the same continuous variable is measured.

The study design ~~schematic~~ for repeated measures ANOVA is usually shown in a table like this:

Subjects	Time / condition		
	T_1	T_2	T_3
S_1	S_1	S_1	S_1
S_2	S_2	S_2	S_2
S_3	S_3	S_3	S_4
S_4	S_4	S_4	S_4
S_5	S_5	S_5	S_5

There are 5 subjects

The variable values at each time/condition for each subject would be listed in the table.

There are 3 conditions or time points

Hypothesis set-up for repeated-measures ANOVA

↳ just like one-way ANOVA, we are testing if there are any differences between related groups

↳ in this case between time points or conditions

$$H_0: \mu_{T_1} = \mu_{T_2} = \mu_{T_3} = \ldots = \mu_{T_K}$$

H_A: at least two means are significantly different

The repeated measures ANOVA is an omnibus test, it does not tell you where the difference lies.

You would need to run post-hoc tests to identify where these differences occur.

SS = sum of squared residuals.

MS = mean sum of squares.

d.f. = degrees of freedom.

~~How a Repeated Measure ANOVA works~~

First Recall Logic of the ~~one-way~~ ANOVA

SS_T

Total variability

Between-Groups Variability

SS_b

with $d.f._b = G-1$

$MS_b = \frac{SS_b}{df_b}$

this was our leftover unexplained variability

Within-Groups Variability

SS_w

with $d.f._w = N - G$

where G = # groups

N = # subjects/observations

$MS_w = \frac{SS_w}{df_w}$

$F = \frac{MS_b}{MS_w}$ or MS_{error}

the F-ratio tells us how much variance we have accounted for in our model for between-groups.

we apply this same principle to Repeated Measures.

Logic of Repeated Measures ANOVA

Total Variability

SS_T

same as SS_b

Time/Conditions Variability

$SS_{conditions}$

with $df_b = G-1$

groups are the different time points or conditions

Within-Groups Variability

SS_w

Independent-Samples One-way ANOVA

Repeated Measures ANOVA

we're further partitioning the previously unexplained error using this term.

Subject Variability

$SS_{subjects}$

Error (Unexplained) Variability

SS_{error}

with ~~(N-G)~~ $df_{err} = (N-1)(G-1)$

$MS_b = \frac{SS_b}{df_b}$

$MS_{err} = \frac{SS_{err}}{df_{err}}$

$F = \frac{MS_b}{MS_{err}}$

notice this is the same as one-way ANOVA except we've removed within-subject variability from the denominator

Example on Calculating Repeated Measure ANOVA

- Same approach as one-way ANOVA, we must calculate our SS for between-groups (now between-time or between-conditions) and SSerror (which was previously SS within-groups, but now it is slightly different)

$$SS_{error} = SS_{within\text{-}group} - SS_{subjects}$$

OR

$$SS_{error} = SS_{Total} - SS_{conditions/time} - SS_{subjects}$$

1) First Calculate SS_{time}

$$SS_{time} = SS_b = \sum_{k=1}^{G} N_k(\bar{Y}_k - \bar{Y})^2$$

OR

$$SS_{time} = \sum_{k=1}^{G} \sum_{i=1}^{N_k} (\bar{Y}_k - \bar{Y})^2$$

where
G = number of time points (groups)
N_k = number of subjects under each (i^{th}) time point
$\bar{Y}_k$ = mean for each k^{th} time point
$\bar{Y}$ = grand mean

2) Second Calculate SS_w (this is the combined variance of within-subjects and our unexplained variance)

$$SS_w = \sum_{k=1}^{G} \sum_{i=1}^{N_k} (Y_{ik} - \bar{Y}_k)^2$$

same as ANOVA :)

3) Calculate $SS_{subjects}$
↳ here we treat each subject as it's own block
↳ each <u>subject is a level</u> of an independent factor called <u>subjects</u>

$$SS_{subjects} = G \sum_{i=1}^{N} (\bar{Y}_i - \bar{Y})^2$$

sum over all subjects

this is the mean of the subject

Numerical Example:

Study ID	Baseline	2wk	4wk	6wk	Subject Avg
1	197	195	225	201	205
2	248	287	294	277	276
3	239	246	242	222	237
4	238	266	278	273	264
5	244	280	285	283	273
Time Average	233	255	265	251	251

Grand Average

$SS_{Time} = 2{,}567$

$SS_{within} = 16{,}176$

$SS_{subjects} = 14{,}571$

4) Finally Calculate SS_{error}

$$SS_{error} = SS_W - SS_{subjects}$$

$$SS_{error} = 16{,}176 - 14{,}571$$

$$SS_{error} = 1{,}604$$

notice how much of the variance this accounts for!

<u>Next</u> Determine Mean Sum of Squares

↳ need to establish degrees of freedom

$$d.f._{time} = d.f._b = G - 1 \quad = 4 - 1 \quad = 3$$

i.e. number of time points minus one

$$\therefore MS_{time} = \frac{SS_{time}}{d.f._{time}} = 856$$

$$d.f._{error} = (N-1)(G-1) = (5-1)(4-1) = 12$$

$$\therefore MS_{error} = \frac{SS_{error}}{d.f._{error}} = \frac{1{,}604}{12} = 134$$

Then we compute our F-ratio! $F = \frac{MS_{b/time}}{MS_{error}} = 6.4$!

$F_{(3,12)}$ $\quad p = 0.008$

Effect Size

$$\eta^2_{partial} = \frac{SS_{conditions}}{(SS_{conditions} + SS_{error})} = \frac{2567}{2567 + 1604} = 0.615$$

→ > 0.14 is considered large

0.06 is medium

0.01 is small

$$\omega^2 = \frac{SS_{time} - (d.f._{time} \cdot MS_{error})}{SS_{total} + MS_{error}}.$$