Experts Event

Statistics – Examining changes

Jorge Tendeiro

11 April 2018

university of groningen

Overview

1. Within-subjects design with $k = 2$ levels: Difference scores *vs* ANCOVA
   - One-within
   - One-between, one-within
2. Within-subjects design with $k > 2$ levels
   - RM-ANOVA (within-subjects ANOVA)
   - RM-MANOVA (profile analysis)
   - RM-Multilevel analysis (linear mixed model)
3. Missing data (brief)

Within-subjects design with $k = 2$ levels: Difference scores vs ANCOVA

Pretest-posttest design (one-within)

- Two repeated measures: Pretest and posttest (i.e., one within-subjects factor with $k = 2$ levels).
- For now assume a one-group sample (i.e., no between-subjects factors).

Pretest y_0	Posttest y_1
s_1	s_1
s_2	s_2
$\dots$	$\dots$
s_n	s_n

Pretest-posttest design (one-within) – Possible analyses

1. Paired t-test
   - This is equivalent to RM-ANOVA when $k = 2$.
   - Consider difference scores: $d = y_1 - y_0$.
     Then paired t-test is equivalent to one-sample t-test on d_i.
   - In regression terms, this consists of fitting a model without predictors:

$$\underbrace{y_{1i} - y_{0i}}_{d_i} = \beta_0 + \varepsilon_i.$$

   Paired t-test $=$ t-test associated to β_0.

2. ANCOVA
   - Regress posttest on pretest.

$$\begin{aligned} y_{1i} &= \beta_0 + \beta_1 y_{0i} + \varepsilon_i \\ \underbrace{y_{1i} - \beta_1 y_{0i}}_{d_i^*} &= \beta_0 + \varepsilon_i. \end{aligned}$$

ANCOVA test $=$ t-test associated to β_0.

Pretest-posttest design (one-within) – Comparison

Test	Model	H_0
Paired t-test	$y_{1i} - y_{0i} = \beta_0 + \varepsilon_i$	$\mu_d = \mu_1 - \mu_0 = 0$
ANCOVA	$y_{1i} - \beta_1 y_{0i} = \beta_0 + \varepsilon_i$	$\mu_d^* = \mu_1 - \beta_1 \mu_0 = 0$

- Paired t-test is a constrained version of ANCOVA ($\beta_1 = 1$).
- $\beta_1 = 1$ is a strong assumption in some cases.
- Thus, ANCOVA is more flexible:
  Smaller error variance, larger power.
- Price to pay for ANCOVA: Loss of 1df.

Observe that the paired t-test and ANCOVA test slightly different H_0s:

- Paired t-test: Population mean of difference scores is zero.
- ANCOVA: Population mean posttest score, adjusted for pretest scores, is zero.

Pretest-posttest design (one-between, one-within)

- Two repeated measures: Pretest and posttest (i.e., one within-subjects factor with $k = 2$ levels).
- More than one group of subjects (i.e., one between-subjects factor with g levels).

This is a mixed between-within subjects design.

Group	Pretest y_0	Posttest y_1
1	s_1	s_1
	s_2	s_2
	...	...
...	...	...
g	...	...
	s_{n-1}	s_{n-1}
	s_n	s_n

Pretest-posttest design (one-between, one-within) – Possible analyses

1. Paired t-test
   - This is equivalent to RM ANOVA when $k = 2$, with one between-subjects factor.
   - Consider difference scores: $d = y_1 - y_0$.
     Then paired t-test is equivalent to between-subjects ANOVA on d_i.
   - In regression terms:

$$\underbrace{y_{1i} - y_{0i}}_{d_i} = \beta_0 + \underbrace{(\beta_1 D_1 + \cdots + \beta_{g-1} D_{g-1})}_{\text{between-subjects factor}} + \varepsilon_i.$$

2. ANCOVA
   - Regress posttest on pretest and covariates (dummy variables).

$$y_{1i} = \beta_0 + (\beta_1 D_1 + \cdots + \beta_{g-1} D_{g-1}) + \beta_g y_{0i} + \varepsilon_i$$
$$\underbrace{y_{1i} - \beta_g y_{0i}}_{d_i^*} = \beta_0 + (\beta_1 D_1 + \cdots + \beta_{g-1} D_{g-1}) + \varepsilon_i.$$

Pretest-posttest design (one-between, one-within) – Comparison

Paired t-test	$y_{1i} - \quad y_{0i} = \beta_0 + (\beta_1 D_1 + \cdots + \beta_{g-1} D_{g-1}) + \varepsilon_i$
ANCOVA	$y_{1i} - \beta_g y_{0i} = \beta_0 + (\beta_1 D_1 + \cdots + \beta_{g-1} D_{g-1}) + \varepsilon_i$

- As before, the paired t-test is a constrained version of ANCOVA ($\beta_g = 1$).
- Price to pay for ANCOVA:
  - Loss of 1df.
  - The ANCOVA assumption of equality of regression slopes needs to be assessed when $g > 1$.
    (Look at size of interaction effect between both effects.)
- As before, paired t-test and ANCOVA test slightly different H_0s (difference scores *vs* adjusted posttest scores).

Pretest-posttest design – How to choose?

Is design experimental or quasi-experimental/ observational?

Pretest-posttest design – How to choose?

Is design experimental or quasi-experimental/ observational?

Experimental

- Randomized groups imply no systematic initial differences between groups.
- Thus, pretest scores on average are equal across groups.
- ANCOVA adjusted means are basically equal to the unadjusted means.
- In this case, paired t-test and ANCOVA test the same hypothesis and estimate the same group differences.
- Still, ANCOVA provides more power and precision than ANOVA scores (smaller error variance), as long as ANCOVA assumptions hold:
  - Usual ANOVA assumptions (independence, normality, homoscedasticity).
  - Linearity = Linear relation between pretest and posttest scores.
  - Homogeneity of slopes = The linear relation is the same across groups.
  - Covariate (i.e., pretest scores) measured without error (humm. . .).

Pretest-posttest design – How to choose?

Is design experimental or quasi-experimental/ observational?

Quasi-experimental/ observational

- Non-randomized: Groups may display mean differences on pretest scores.
- In this case, ANCOVA's adjusted means may differ from non-adjusted means.
- In this case, paired t-test and ANCOVA test different hypotheses.
- Furthermore, ANCOVA may be possibly invalid (*natural* groups).
  - Groups may differ due to factors not considered in the experiment.
  - Such differences might matter (i.e., removing them is a bad idea).
  - Is group membership unrelated to pretest scores?
    - Yes: ANCOVA is OK.
    - No: ANCOVA is doubtful.
      Results from paired t-test and ANCOVA differ (sometimes a lot!; Lord's paradox).
      Paired t-test might be better if the adjusted means are unrealistic.

Within-subjects design with $k > 2$ levels

Within-subjects design with $k > 2$ levels

Poor models to use:

1. Separate ANOVAs for each time point.
   - Allows studying between-subjects effect at each time point separately.
   - Does not allow studying within-subjects effect.
2. Paired t-test for each pair of time points.
   - $k - 1$ tests performed $\longrightarrow$ chance capitalization
   - Reduced power (also when controlling familywise error)
   - More complex relations between time points ignored.

Viable model options:

1. RM-ANOVA or within-subjects ANOVA.
2. RM-MANOVA or ‘profile analysis’.
3. RM-Multilevel analysis.

RM-ANOVA (within-subjects ANOVA)

Main idea

- Block on subjects: Subject as a random effects factor.
  ('random' because subjects are a SRS from the population)
- Thus, remove within-subjects variability from error variance.
- RM-ANOVA is an instance of a mixed model.

Between-subjects ANOVA:

$$SS_T = SS_{\text{Between}} + \underbrace{SS_{\text{Within}}}$$

Within-subjects ANOVA:

$$SS_T = SS_{\text{Between}} + \overbrace{SS_{\text{Subjects}} + SS_{\text{Error}}}$$

RM-ANOVA (within-subjects ANOVA)

Assumptions:

- Independent observations (across subjects).
- Normality.
- Sphericity (for $k > 2$):
  Variances for differences of y scores for any time point pairs are equal.
  Mauchly's test for sphericity, but it is poor:
  - Hope not to reject H_0 (lack of stat. sig. $\neq H_0$ holds).
  - Too sensitive to violations of normality.

RM-ANOVA (within-subjects ANOVA)

Q: Sphericity violated, so what?
A: Test biased (inflated Type I error rate).

What to do?

- Simply ignore the unadjusted test (Maxwell & Delaney, p. 545).
  Thus, ignore (please!) Mauchly's test.
- Use epsilon-correction, from conservative to liberal:
  - Lower-bound correction (overly conservative).
  - Greenhouse-Geisser; preferable for small *n*.
  - Huynh-Feldt.

  For large *n* it matters little which correction to use (G-G $\simeq$ H-F).
- Use other models:
  - RM-MANOVA (profile analysis).
  - RM-Multilevel analysis.

RM-MANOVA (profile analysis)

- RM-MANOVA = MANOVA of $k - 1$ transformed scores (e.g., $Y_2 - Y_1, \ldots, Y_k - Y_{k-1}$).
- This is the same idea as creating difference scores in the pretest-posttest design.
- The omnibus multivariate F tests are invariant across sets of (linearly independent) transformations.
- Assumptions: Those from MANOVA (multivariate normality, homogeneity of variance-covariance matrix)

RM-ANOVA *vs* RM-MANOVA – How to choose?

‘Sphericity’ criterion

- **If sphericity is violated ($\varepsilon < .7$) and sample size is ‘large’:** RM-MANOVA.
- **If sphericity holds ($\varepsilon > .7$) or sample size is ‘small’:** RM-ANOVA.

What is ‘small’ or ‘large’? That is unfortunately debatable.
(see Algina & Keselman, 1997; Maxwell & Delaney, 2004; Keppel & Wickens, 2004).

’Type I error rate’ criterion

RM-ANOVA $\approx$ RM-MANOVA.

‘Constrasts’ criterion

RM-MANOVA preferred because it offers a consistent approach with the omnibus test (Maxwell & Delaney, 2004, p. 672).

RM-Multilevel analysis (linear mixed model)

Repeated measurements (level 1) nested within subjects (level 2).

Very flexible:

- Different number of measurements across subjects.
- Measurements at different time points (unlike RM-(M)ANOVA).
- NAs allowed; much more flexible than e.g. RM-MANOVA (which requires listwise deletion).
- More than two levels allowed (unlike RM-(M)ANOVA).
- Relations between groups and within groups modeled simultaneously.
- Predictors at each level allowed:
  - Level 1: Time-dependent variables.
  - Level 2: Individual characteristics.
- Cross-level interactions possible.
  E.g.: Do patterns across time differ between genders?
- Regression models per subject.

RM-(M)ANOVA *vs* RM-Multilevel – How to choose?

- If sphericity is violated $\longrightarrow$ Discard RM-ANOVA.
  RM-Multilevel analysis models the var-cov matrix, much more flexible than RM-ANOVA.
- Problems with NAs $\longrightarrow$ Discard RM-MANOVA.
  - Avoid 'saving the day' by resorting to poor missing values tricks (like listwise deletion or mean imputation; more below).
  - RM-Multilevel very flexible (assumes MAR).
- Unequal time points across subjects $\longrightarrow$ RM-Multilevel analysis.
- The data hierarchical structure involved more than 2 levels $\longrightarrow$ RM-Multilevel analysis.
- Time-level covariates $\longrightarrow$ RM-Multilevel analysis.

For completeness, also keep in mind that SEM and its latent growth curve model is a viable option (no details today).

Missing data (brief)

Missing data – Mechanisms

Three common missing data mechanisms (Rubin, 1976):

- Missing completely at random (MCAR)
  - NA unrelated to observed and missing data.
- Missing at random (MAR)
  - NA unrelated to missing data, but related to observed data.
  - Thus, nonresponse can be predicted by observed data.
- Missing not at random (MNAR)
  - NA related to missing data.

These mechanisms are assumptions:

- Only MCAR can be empirically tested.
  However, these tests are typically low powered (Baraldi & Enders, 2010).
- MAR and MNAR depend on the unobserved data, thus cannot be verified.

Missing data – Classical tecniques

- Deletion: Listwise, pairwise.
  - ✓ Complete data sets (for listwise).
  - ✗ (Much) smaller $N \longrightarrow$ low power.
  - ✗ MCAR assumed.
- Single imputation: Mean, regression, stochastic regression.
  - ✓ Complete data sets.
  - ✓ No reduction in N.
  - ✓ MAR assumed (regression, stochastic regression).
  - ✗ Correlations attenuated (mean) or overestimated (regression).
  - ✗ Variance attenuated (mean, regression).
  - ✗ MCAR assumed (mean).

Stochastic regression is the best of the above options. But:

✗ SEs are too small (because uncertainty in the imputted missing data is ignored in its computation), thus Type I error rates can be unnaceptably high.

Missing data – Modern tecnniques

- Maximum likelihood estimation (MLE).
- Multiple imputation (MI).

These techniques are preferable over traditional ones.

✓ Unbiased under MCAR and MAR.

✓ No reduction in $N \longrightarrow$ larger power.

But:

✗ MAR is untestable.

✗ MI: It can be difficult to pool estimates together.

✗ Based on assumptions (e.g., multivariate normality).