Within Cases ANOVA Part One: Mixed Model and Multivariate Approaches[1]

STA441 Spring 2024

[1] See last slide for copyright information.

Contents

1. Overview of Within Cases
2. Mixed Model
3. Multivariate Approach

Within Cases

Example: A random sample of male and female university students is weighed midway through year 1, 2, 3 and 4. The explanatory variables are gender and year (time).

Gender is a between-cases factor and year is a within-cases factor.

- For a within-cases factor, a case contributes a response variable value for more than one value of the explanatory variable — usually all of them.
- It is natural to expect data from the same case to be correlated – *not* independent.
- For example, the same subject appears in several treatment conditions.
- Hearing study: How does pitch affect our ability to hear faint sounds? The same subjects will hear a variety of different pitch and volume levels (in a random order). They press a key when they think they hear something.

Student's Sleep Study (*Biometrika*, 1908)

First Published Example of a t-test

- Patients take two sleeping medicines several days apart.
- Half get A first, half get B first.
- Reported extra hours of sleep are recorded (difference from baseline).
- It's natural to subtract, and test whether the mean *difference* equals zero.
- That's what Gossett did.
- But some might do an independent t-test with $n_1 = n_2$.
- This assumes observations from the same person to be independent.
- It's unrealistic, but is it harmful?

Matched pairs, testing $H_0 : \mu_1 = \mu_2$

Independent *v.s.* Matched t-test

- If population covariance between the two measurements is positive, Type I error probability of both tests is 0.05, but matched t-test has better power.
- If population covariance between measurements is negative, matched t-test has Type I error probability of 0.05, but the independent t-test has Type I error probability greater than 0.05.

Why?

Why the matched t-test is better

- Numerator of both test statistics is $\overline{d} = \overline{y}_1 - \overline{y}_2$.
- Denominator is an estimate of the standard deviation of the difference.
- $Corr(\overline{y}_1, \overline{y}_2) = Corr(y_{i,1}, y_{i,2})$.
- So $Cov(\overline{y}_1, \overline{y}_2)$ has the same sign as $Cov(y_{i,1}, y_{i,2})$.
- $Var(\overline{y}_1 - \overline{y}_2) = Var(\overline{y}_1) + Var(\overline{y}_2) - 2Cov(\overline{y}_1, \overline{y}_2)$.
- If $Cov(\overline{y}_1, \overline{y}_2) > 0$, pretending independence results in overestimation of $Var(\overline{y}_1 - \overline{y}_2)$. Denominator is larger, so t statistic is smaller.
- If $Cov(\overline{y}_1, \overline{y}_2) < 0$, pretending independence results in underestimation of $Var(\overline{y}_1 - \overline{y}_2)$. Denominator is smaller, so t statistic is greater (too big).

Within-cases Terminology

You may hear terms like “longitudinal” and “repeated measures.”

- **Longitudinal**: The same variables are measured repeatedly over time. Usually there are lots of variables, including categorical ones, and large samples. If there’s an experimental treatment, its usually once at the beginning, like a surgery. Longitudinal studies basically track what happens over time.
- **Repeated measures**: Usually, the same subjects experience two or more experimental treatments. Usually quantitative response variables, and often small samples.

Wine Tasting Example

A single within-cases factor

In a taste test of wine, 6 professional judges judged 4 wines. The numbers they gave do not exactly represent quality. Instead, they are maximum prices in dollars per bottle that the judge thinks the company can charge and still sell most of the wine.

- Cases are judges: $n = 6$.
- Each judge tastes and rates all four wines.
- The single factor is Wine: Four categories.

Archery Example: Bow and Arrow

Two within-cases factors

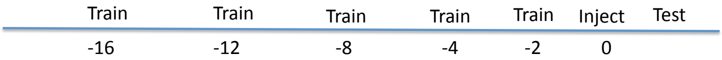
- Cases are archers. There are n archers.
- Test two bows, three arrow types.
- Warmup, then each archer takes 10 shots with each Bow-Arrow combination — 60 shots.
- In a different random order for each archer, of course.
- $y_{i,1}, \ldots, y_{i,6}$ are mean distances from arrow tip to centre of target, for $i = 1, \ldots, n$.
- Each $y_{i,j}$ is based on 10 shots.
- $E(y_{i,j}) = \mu_j$ for $j = 1, \ldots, 6$.

One Between, One Within

- Grapefruit study: Cases are n grocery stores.
- Within stores factor: Three price levels.
- Between-stores factor: Incentive program for produce managers (Yes-No).

Monkey Study

- Train monkeys on discrimination tasks, at 16, 12, 8, 4 and 2 weeks prior to treatment. Different task each time, equally difficult (randomize order).
- Treatment is to block function of the hippocampus (with drug, not surgery), re-test. Get 5 scores for each monkey.



- 11 randomly assigned to treatment, 7 to control.
- Treatment is between, time elapsed since training is within.

Advantages of Within-cases Designs

If measurement of the response variable does not mess things up too much

- Convenience (sometimes).
- Each case serves as its own control. A huge number of extraneous variables are automatically held constant. The result can be a very sensitive analysis.
- For some models, you can have lots of measurements on just a few subjects — if you are willing to make some assumptions.

Three main approaches for normal response variables

- Classical Mixed (Random Shock) model
- Multivariate
- Covariance Structure

- Randomization.

Random Shock Model

"Mixed" fixed and random effects

- Expect multiple measurements coming from the same individual to be correlated.
- Here's a model for how it could happen.
- In addition to an ordinary regression model, each person (case) brings his or her own individual contribution, a little piece of noise that pushes all his or her data values up or down by the same amount.
- Random shock — random because the individual is randomly sampled.

Mixed Model

Both fixed and random effects

- Random shock from each person.
- A random effect (for person).
- A mixed model because there are also fixed effects in the regression model.
- Fixed means the regression coefficients are constants.

Regression model

Cases $i = 1, \ldots, n$, with $j = 1, \ldots k$ values of y for each case

$$y_{ij} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \cdots + \beta_{p-1} x_{ij,p-1} + \epsilon_{ij} + \delta_i$$

- $\epsilon_{ij} \sim N(0, \sigma^2)$
- $\delta_i \sim N(0, \sigma_1^2)$
- Independent.

Example: Grapefruit sales

Three price levels within stores, one incentive program between

Store	p1	p2	Incent	Sales
1	1	0	1	18
1	0	1	1	27
1	0	0	1	14
2	1	0	0	32
2	0	1	0	11
2	0	0	0	9

$$
\begin{aligned}
y_{i1} &= \beta_0 + \beta_1 p_{i11} + \beta_2 p_{i12} + \beta_3 d_i + \epsilon_{i1} + \delta_i \\
y_{i2} &= \beta_0 + \beta_1 p_{i21} + \beta_2 p_{i22} + \beta_3 d_i + \epsilon_{i2} + \delta_i \\
y_{i3} &= \beta_0 + \beta_1 p_{i31} + \beta_2 p_{i32} + \beta_3 d_i + \epsilon_{i3} + \delta_i
\end{aligned}
$$

Variances and Covariances

$$\begin{aligned}
y_{i1} &= \beta_0 + \beta_1 p_{i11} + \beta_2 p_{i12} + \beta_3 d_i + \epsilon_{i1} + \delta_i \\
y_{i2} &= \beta_0 + \beta_1 p_{i21} + \beta_2 p_{i22} + \beta_3 d_i + \epsilon_{i2} + \delta_i \\
y_{i3} &= \beta_0 + \beta_1 p_{i31} + \beta_2 p_{i32} + \beta_3 d_i + \epsilon_{i3} + \delta_i
\end{aligned}$$

$$\begin{aligned}
Var(y_{i1}) &= Var(\epsilon_{i1}) + Var(\delta_i) \\
&= \sigma^2 + \sigma_1^2 \\
&= Var(y_{i2}) = Var(y_{i3})
\end{aligned}$$

$$\begin{aligned}
Cov(y_{i1}, y_{i2}) &= Cov(\epsilon_{i1} + \delta_i \, , \, \epsilon_{i2} + \delta_i) \\
&= Cov(\delta_i, \delta_i) = Var(\delta_i) = \sigma_1^2 \\
&= Cov(y_{i1}, y_{i3}) = Cov(y_{i2}, y_{i3})
\end{aligned}$$

Compound Symmetry

$$cov(\mathbf{y}_i) = \begin{pmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 \end{pmatrix}$$

- All variances equal.
- All covariances equal.
- All covariances positive.

Classical F-tests

Based on theory of mixed models

- No covariates allowed.
- F-tests exist only for balanced designs,
- And not for all models, even if balanced.

- Modern, high quality *approximate* F-tests are available.
- `proc mixed` in SAS and `lmer` in R.

Random Intercept

$$
\begin{aligned}
y_{ij} &= \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \cdots + \beta_{p-1} x_{ij,p-1} + \epsilon_{ij} + \delta_i \\
&= (\beta_0 + \delta_i) + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \cdots + \beta_{p-1} x_{ij,p-1} + \epsilon_{ij}
\end{aligned}
$$

- $(\beta_0 + \delta_i) \sim N(\beta_0, \sigma_1^2)$
- So the random shock model is sometimes called a “random intercept” model.
- That’s why the `lmer` syntax in R might look like

```
lmer(sales ~ price*incent + (1|store) )
```

Advantages and disadvantages of the random shock model

$$cov(\mathbf{y}_i) = \left(\begin{array}{cccc} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 \end{array} \right)$$

- Makes sense for some data sets more than others.
- I like it for experiments where the participants respond to several different stimuli, in a different random order for each participant.
- For longitudinal data, not so much.
- Should the correlation between time 1 and time 2 be the same as the correlation between time 1 and time 50?

Can be a life saver for small samples

Example: Ten men and ten women judge the beauty of randomly renerated shapes. There are seven levels of colour (roybgiv), five levels of complexity, and four levels of size. Each subject judges all $7 \times 5 \times 4 = 140$ pictures, in a different random order for each subject.

There are 3 within-subjects factors and one between-subjects factor.

- There are 280 β_j parameters in the regression model, but only $n = 20$.
- However, there are $20 \times 140 = 2,800$ data lines, and only $280 + 2 = 282$ parameters.
- Everything is fine, if you don't hate the model too much.

- General principle: There is a trade-off between sample size and assumptions.

If assumptions are violated

- Compound symmetry is assumed, but only “sphericity” is actually required.
- Sphericity means the variances of the differences are all the same.
- Without sphericity, the Type I error probability can be greater than 0.05.

- Two corrections of the p-values are available: Greenhouse-Geisser and Huynh-Feldt.
- There is also a test for sphericity.

Multivariate Approach to Repeated Measures

- Multivariate methods allow the analysis of more than one response variable at the same time.
- When a case (subject) provides data under more than one set of conditions, it is natural to think of the measurements as multivariate.
- The humble matched t-test has a multivariate version: Hotelling's T^2.
- Simultaneously test whether the means of several *differences* equal zero.
- Like rating of Wine One minus Wine Two, Wine Two minus Wine Three, and Wine Three minus Wine Four.
- When there are also between-subjects factors (like nationality of judge), use multivariate regression methods.

Pure within-cases: Multiple factors

Archery example

Each archer contributes 6 numbers:

	Arrow type		
Bow type	1	2	3
1	$E(y_{i,1}) = \mu_{11}$	$E(y_{i,2}) = \mu_{12}$	$E(y_{i,3}) = \mu_{13}$
2	$E(y_{i,4}) = \mu_{21}$	$E(y_{i,5}) = \mu_{22}$	$E(y_{i,6}) = \mu_{23}$

- Form (sets of) linear combinations of the response variables.
- Want to test main effect of Bow Type?
  - $H_0 : \mu_{11} + \mu_{12} + \mu_{13} = \mu_{21} + \mu_{22} + \mu_{23}$
  - Calculate $L_i = y_{i,1} + y_{i,2} + y_{i,3} - (y_{i,4} + y_{i,5} + y_{i,6})$.
  - $E(L_i) = \mu_{11} + \mu_{12} + \mu_{13} - (\mu_{21} + \mu_{22} + \mu_{23})$.
  - Test $H_0 : E(L_i) = 0$.
  - Could use an ordinary matched t-test for this one.

Main effect for arrow type

Differences between marginal means

	Arrow type		
Bow type	1	2	3
1	$E(y_{i,1}) = \mu_{11}$	$E(y_{i,2}) = \mu_{12}$	$E(y_{i,3}) = \mu_{13}$
2	$E(y_{i,4}) = \mu_{21}$	$E(y_{i,5}) = \mu_{22}$	$E(y_{i,6}) = \mu_{23}$

- $H_0 : \mu_{11} + \mu_{21} = \mu_{12} + \mu_{22}$ and $\mu_{12} + \mu_{22} = \mu_{13} + \mu_{23}$
- Calculate two linear combinations for each archer:
  - $L_{i,1} = y_{i,1} + y_{i,4} - (y_{i,2} + y_{i,5})$
  - $L_{i,2} = y_{i,2} + y_{i,5} - (y_{i,3} + y_{i,4})$
- Simultaneously test $H_0 : E(L_{i,1}) = 0$ and $E(L_{i,2}) = 0$.
- Use Hotelling's T^2.
- Or something equivalent.

Matched t-tests with `proc reg`

- Regression with no explanatory variables.
- $y_i = \beta_0 + \epsilon_i \sim N(\beta_0, \sigma^2)$.
- Test $H_0 : \beta_0 = 0$.

```
proc reg;
     model y =  ;
```

Hotelling's T-squared

Multivariate matched t-test

- Official SAS documentation claims that SAS won't calculate Hotelling's T-squared, but ...
- $T^2 = (n-1)\left(\frac{1}{\lambda}-1\right)$, so just get Wilks' Lambda from the `mtest` statement of `proc reg`. The p-value will be correct.
- In a regression model with *no explanatory variables*, and a vector of differences $\mathbf{d}_i$ for $i = 1, \ldots, n$, $E(\mathbf{d}_i) = \boldsymbol{\beta}_0$, so test $H_0 : \boldsymbol{\beta}_0 = \mathbf{0}$.

```
proc reg;
     model D1 D2 D3 =  ;
     Wine: mtest intercept=0;
```

- Or just use the test for Wilks' lambda directly.

Designs with both between and within cases factors

- Could have main effects and interactions for between-cases factors,
- Could have main effects and interactions for within-cases factors,
- Could have interactions of between by within.
- Again, observation from the same case are treated as multivariate.
- Again we form linear combinations of response variables and test hypothesis about them.
- **Recipe**: *Use a regression model with effect coding dummy variables for the between-cases factors* (if any). Use these same explanatory variables in every model.
- Response variables (linear combinations) will vary depending on the effect being tested.
- Null hypotheses for all the main effects and interactions are statements about the β values.

Main effects and interactions for the between-cases factors

- These are marginals, averaging μ parameters over the within-cases factors.
- Let $L_i =$ the mean (or sum) of the $y_{i,j}$ values averaging or adding over j.
- Do a standard between-cases analysis with L_i as the response variable.

Main effects and interactions for the within-cases factors

- Need to average μ parameters over the between-cases factors.
- Effect coding! β_0 is the grand mean.
- Form linear combinations as in the archery example.
- Test $H_0 : \beta_0 = 0$.
- Or test multiple $\beta_{0,j} = 0$ if need be.

Interactions of between by within

- The nature of a within-cases effect *depends* on a between-cases treatment combination.
- Take the linear combinations for the within-cases effect.
- Test the between-cases effect on those.
- For example, factors are Bow Type, Arrow Type and Gender.
- Want to test the Arrow Type by Gender interaction.
- Are the differences between arrow types (averaging over bow types) different for men and women?
- Simultaneously test for gender differences in the two linear combinations representing arrow type one versus two and two versus three.
- It's a standard multivariate test.

You could use `proc reg`

To test the arrow type by gender interaction

	Arrow type		
Bow type	1	2	3
1	$E(y_{i,1}) = \mu_{11}$	$E(y_{i,2}) = \mu_{12}$	$E(y_{i,3}) = \mu_{13}$
2	$E(y_{i,4}) = \mu_{21}$	$E(y_{i,5}) = \mu_{22}$	$E(y_{i,6}) = \mu_{23}$

```
L1 = y1+y4 - (y2+y5);
L2 = y2+y5 - (y3+y6);

proc reg;
     model L1 L2 = gender;
     arrow_by_sex: mtest gender=0;
```

Or you can let `proc glm` do the dummy variables and linear combinations for you.

If within-cases factors have just two levels

Like before and after, experimental vs. control

- You can always do it with a univariate analysis.
- No fancy software is needed.
- Make a sum variable and a difference variable.
- Salmon study: Fish are Canadian or Alaskan, Female or Male, Growth is measured in freshwater *and* marine environments.
- Three factors: Species by sex by environment – environment is within cases.
- Response variable is growth.

Salmon example
SAS code not tested

```
sumgrowth = freshgrowth + marinegrowth;
difgrowth = freshgrowth - marinegrowth;
```

Assume effect coding for country and sex, product term `cs`.

```
proc reg;
     title2 'Between-cases effects';
     model sumgrowth = country sex cs;
proc reg;
     title2 'Within and between-within';
     model difgrowth = country sex cs;
```

What do the t-tests give you?

Advantages of the multivariate approach

- Straightforward application of classical multivariate methods.
- No restrictions on the correlations of observations coming from the same case. Could have some positive, some negative, some zero. Let the data speak.
- Can be extended to repeated measures on *vectors* of observations (called doubly multivariate repeated measures).

Disadvantages of the multivariate approach

- Lots of unknown variance and covariance parameters. Small sample size and lots of experimental conditions will not work. Longitudinal at lots of time points is out.
- No way to model the dependence between observations.
- No time-varying covariates. Explanatory variable values must be the same for all y from a given case.

Copyright Information