Session 9: Repeated Measures and Longitudinal Analysis I

Levi Waldron

Learning objectives and outline

Intro: hierarchical and longitudinal data

Fecal Fat example

Correlations within subjects (ICC)

Random and fixed effects

Session 9: Repeated Measures and Longitudinal Analysis I

Levi Waldron

CUNY SPH Biostatistics 2

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

Learning objectives and outline

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

Learning objectives

Learning objectives:

1. Identify and define hierarchical and longitudinal data
2. Analyze correlated data using Analysis of Variance
3. Define and calculate Intraclass Correlation
4. Identify and define random and fixed effects

Textbook sections:

- Vittinghoff sections 7.1 (7.2-7.3 next class)

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

Outline

1. Introduction to hierarchical and longitudinal data
2. Fecal Fat example
3. Correlations within subjects (ICC)
4. Random and fixed effects

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

Intro: hierarchical and longitudinal data

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

What are hierarchical and longitudinal data?

- Knee radiographs are taken yearly in order to understand the onset of osteoarthritis
- An indicator of heart damage is measured at 1, 3, and 6 days following a brain hemorrhage.
- Groups of patients in a urinary incontinence trial are assembled from different treatment centers
- Susceptibility to tuberculosis is measured in family members
- A study of the choice of type of surgery to treat a brain aneurysm either by clipping the base of the aneurysm or implanting a small coil. The study is conducted by measuring the type of surgery a patient receives from a number of surgeons at a number of different institutions.

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

What is the distinction between hierarchical and longitudinal data?

- Longitudinal data are repeated measures over time
- Longitudinal data are a type of hierarchical data
  - repeated measures are correlated, and nested within the observational unit (individual)
- Other non-longitudinal data can also be hierarchical

Definition: Hierarchical data are data (responses or predictors) collected from or specific to different levels within a study.

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

Important features of this type of data

1 The outcomes are correlated across observations

2 The predictor variables can be associated with different levels of a hierarchy. *e.g.* we might be interested in:
   - the volume of operations at the hospital,
   - whether it is a for-profit or not-for-profit hospital,
   - years of experience of the surgeon or where surgeons were trained,
   - how the choice of surgery type depends on the age and gender of the patient.

Session 9:
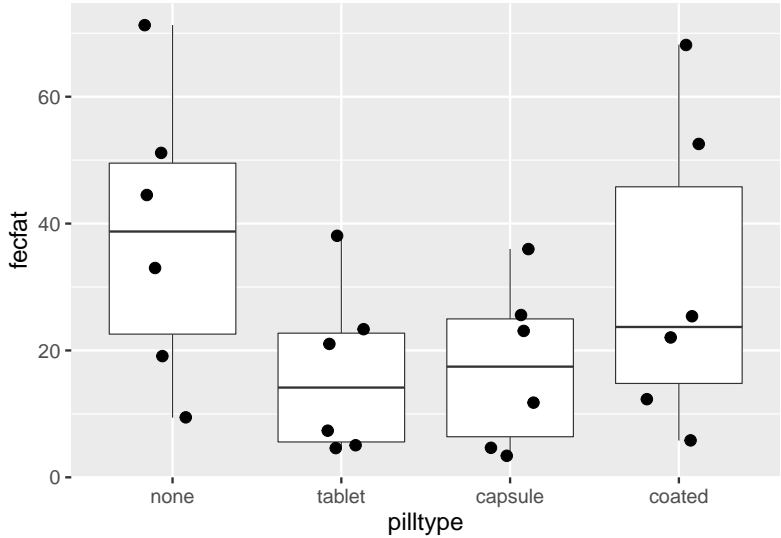Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

Fecal Fat example

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

A Repeated Measures Example

- Lack of digestive enzymes in the intestine can cause bowel absorption problems.
  - This will be indicated by excess fat in the feces.
  - Pancreatic enzyme supplements can alleviate the problem.
  - fecfat.csv: a study of fecal fat quantity (g/day) for individuals given each of a placebo and 3 types of pills

Table 7.1 Fecal fat (g/day) for six subjects

Subject number	Pill type				Subject Average
	None	Tablet	Capsule	Coated	
1	44.5	7.3	3.4	12.4	16.9
2	33.0	21.0	23.1	25.4	25.6
3	19.1	5.0	11.8	22.0	14.5
4	9.4	4.6	4.6	5.8	6.1
5	71.3	23.3	25.6	68.2	47.1
6	51.2	38.0	36.0	52.6	44.5
Pill type average	38.1	16.5	17.4	31.1	25.8

**Session 9:
Repeated
Measures and
Longitudinal
Analysis I**

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

**Fecal Fat
example**

Correlations
within
subjects
(ICC)

Random and
fixed effects

Option 1: non-hierarchical analysis (wrong)

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

Option 1: non-hierarchical analysis (wrong)

```
fit1way <- lm(fecfat ~ pilltype, data=dat)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
pilltype	3	2008.60	669.53	1.86	0.1687
Residuals	20	7193.36	359.67		

Table 1: One-way analysis of variance table for fecal fat dataset

- Does not account for similarity of measurements within individual
- Would be correct if each treatment were given to a different individual

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

Option 2: 2-way AOV

- Accounts for individual differences in mean fecal fat
- Fits a coefficient for mean fecal fat per individual
- Getting closer
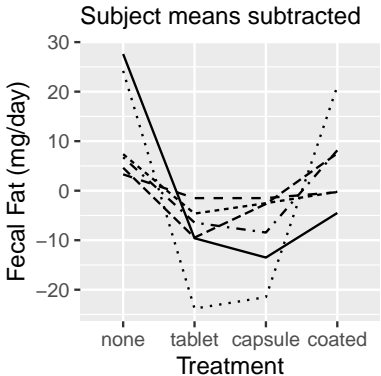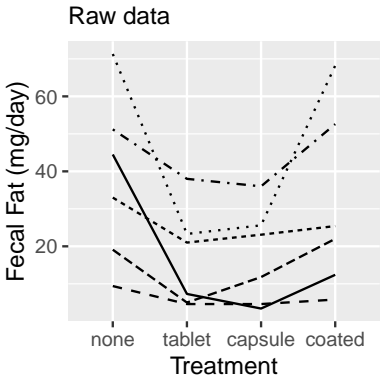
Session 9: Repeated Measures and Longitudinal Analysis I

Levi Waldron

Learning objectives and outline

Intro: hierarchical and longitudinal data

Fecal Fat example

Correlations within subjects (ICC)

Random and fixed effects

Option 2: 2-way AOV

```
fit1way <- lm(fecfat ~ pilltype, data=dat)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
pilltype	3	2008.60	669.53	1.86	0.1687
Residuals	20	7193.36	359.67		

Table 2: One-way analysis of variance table for fecal fat dataset

```
fit2way <- lm(fecfat ~ subject + pilltype, data=dat)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
subject	5	5588.38	1117.68	10.45	0.0002
pilltype	3	2008.60	669.53	6.26	0.0057
Residuals	15	1604.98	107.00		

Table 3: Two-way analysis of variance table. Note the similarity of the pilltype row.

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

What happened??

- 1-way ANOVA correctly estimates the effect of pill type
- However, 1-way ANOVA fails to accommodate the correlation within subjects
- 1-way ANOVA over-estimates the residual variance
  - under-estimates the significance of pill type

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

Regression models for 1 and 2-way ANOVA

- Recall for ordinary multiple linear regression:

$$E[y|x] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p$$

  - x_p are the predictors or independent variables
  - y is the outcome, response, or dependent variable
  - $E[y|x]$ is the expected value of y given x
  - β_p are the regression coefficients

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

Regression models for 1 and 2-way ANOVA

- One-way ANOVA (person i with pill type j):

  $FECFAT_{ij}$ = fecal fat measurement for person i with pill typ
  $$= \mu + PILLTYPE_j + \epsilon_{ij}$$

- Two-way ANOVA:

  $$FECFAT_{ij} = \mu + SUBJECT_i + PILLTYPE_j + \epsilon_{ij}$$

Assumption: $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

**Correlations
within
subjects
(ICC)**

Random and
fixed effects

Correlations within subjects (ICC)

Session 9: Repeated Measures and Longitudinal Analysis I

Levi Waldron

Learning objectives and outline

Intro: hierarchical and longitudinal data

Fecal Fat example

Correlations within subjects (ICC)

Random and fixed effects

Correlations within subjects

- One-way ANOVA fails because it does not account for the correlation of measurements within-person
- How highly correlated are measurements on the same person? Consider subject i, pill types j and k:

$$corr(FECFAT_{ij}, FECFAT_{ik}) = \frac{cov(FECFAT_{ij}, FECFAT_{ik})}{sd(FECFAT_{ij})sd(FECFAT_{ik})}$$

* This is a measure of how large the subject effect is, in relation to the error term

Session 9: Repeated Measures and Longitudinal Analysis I

Levi Waldron

Learning objectives and outline

Intro: hierarchical and longitudinal data

Fecal Fat example

Correlations within subjects (ICC)

Random and fixed effects

Correlation within subjects

$$cov(FECFAT_{ij}, FECFAT_{ik}) = cov(SUBJECT_i, SUBJECT_i)$$
$$= var(SUBJECT_i)$$
$$= \sigma^2_{subject}.(\text{definition})$$

- Equality 1:
  - μ and *pilltype* terms are assumed to be constant, so do not enter into covariance calculation
  - residuals ϵ are assumed to be independent
- Equality 2:
  - covariance with self is variance

Recall $SUBJECT_i$ is the term for individual in 2-way AOV. Now $\beta_i * subjectID$, will later be treated as a **random variable**

Session 9: Repeated Measures and Longitudinal Analysis I

Levi Waldron

Learning objectives and outline

Intro: hierarchical and longitudinal data

Fecal Fat example

Correlations within subjects (ICC)

Random and fixed effects

Correlation within subjects

Previous slide calculated *covariance* for numerator of correlation. Now calculate *variance* for the denominator $(sd(FECFAT_{ij}) * sd(FECFAT_{ik}) = var(FECFAT_{ij}))$

$$var(FECFAT_{ij}) = var(SUBJECT_i, SUBJECT_i) + var(\epsilon_{ij})$$
$$= \sigma^2_{subject} + \sigma^2_{\epsilon}.(\text{definition})$$

- Difference is that the independent residuals do contribute to $var(FECFAT_{ij})$
- Variance is broken into componenets due to *subject* and *residual* variance

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

Intraclass Correlation

The correlation between two treatments j and k across subjects i is:

$$corr(FECFAT_{ij}, FECFAT_{ik}) = \frac{cov(FECFAT_{ij}, FECFAT_{ik})}{sd(FECFAT_{ij})sd(FECFAT_{ik})}$$

$$= \frac{\sigma^2_{subj}}{\sigma^2_{subj} + \sigma^2_\epsilon}$$

$$ICC = \frac{\tau^2_{00}}{\tau^2_{00} + \sigma^2_\epsilon}$$

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

**Correlations
within
subjects
(ICC)**

Random and
fixed effects

Intuition behind correlations within subjects

Table 7.1 Fecal fat (g/day) for six subjects

Subject number	Pill type				Subject Average
	None	Tablet	Capsule	Coated	
1	44.5	7.3	3.4	12.4	16.9
2	33.0	21.0	23.1	25.4	25.6
3	19.1	5.0	11.8	22.0	14.5
4	9.4	4.6	4.6	5.8	6.1
5	71.3	23.3	25.6	68.2	47.1
6	51.2	38.0	36.0	52.6	44.5
Pill type average	38.1	16.5	17.4	31.1	25.8

Figure 2: Fecal Fat dataset

Variance of the subject averages (279.4) is increased by correlation of measurements within individual.

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Calculation of correlations within subjects (ICC)

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

**Correlations
within
subjects
(ICC)**

Random and
fixed effects

What is your estimate of the variability due to subjects, from the 2-way ANOVA?

```
sum(residuals(fit2way)^2) / 15 / 4 #df=15, divided by 4 pilltypes
```

[1] 26.74972

```
279.419 - 26.75 #var(SUBJECT_i)
```

[1] 252.669

Residual variance is:

```
sum(residuals(fit2way)^2) / 15 #df=15
```

[1] 106.9989

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

**Correlations
within
subjects
(ICC)**

Random and
fixed effects

Calculation of correlations within subjects (ICC)

Finally calculate ICC:

$$\begin{aligned} ICC &= \frac{\sigma^2_{subj}}{\sigma^2_{subj} + \sigma^2_{\epsilon}} \\ &= \frac{253}{253 + 107} \quad = 0.70 \end{aligned}$$

This calculation will become easier when we learn to estimate *random coefficients* in directly in the regression model.

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

Random and fixed effects

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

The next step: a mixed effects model

- Two-way ANOVA is a fixed effects model:

$$FECFAT_{ij} = \beta_0 + \beta_{subjecti}SUBJECT_i + \beta_{pilltypej}PILLTYPE_j + \epsilon_{ij}$$

  - Assumption: $\epsilon_i \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$

- Instead of fitting a $\beta_{subjecti}$ to each individual, assume that subject effects are selected from a distribution of possible subject effects:

$$FECFAT_{ij} = \mu + SUBJECT_i + \beta_{pilltypej}PILLTYPE_j + \epsilon_{ij}$$

where $SUBJECT_i \overset{iid}{\sim} N(0, \sigma_{subj}^2)$

- Here subject is a *random* effect, and pill type is a *fixed* effect.

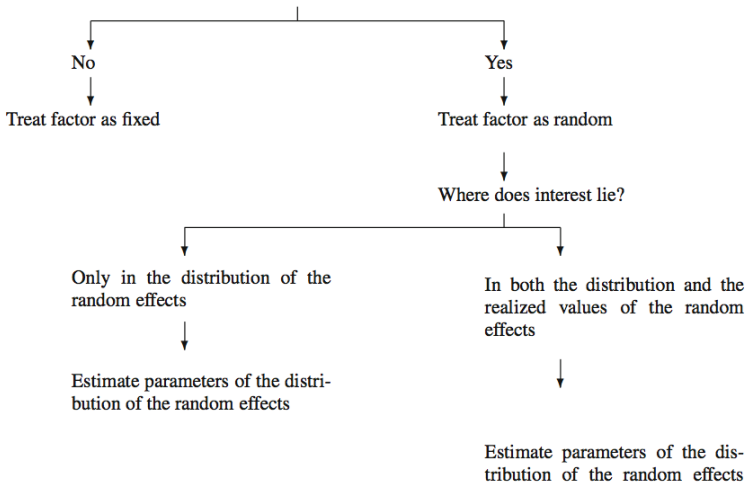- This is also a random intercept model

Session 9:
Repeated
Measures and
Longitudinal
Analysis I

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

Random and
fixed effects

Random and fixed effects

Table 7.14 Decision tree for deciding between fixed and random

Is it reasonable to assume levels of the factor come from a
probability distribution?

No

Treat factor as fixed

Yes

Treat factor as random

Where does interest lie?

Only in the distribution of the
random effects

Estimate parameters of the distri-
bution of the random effects

In both the distribution and the
realized values of the random
effects

Estimate parameters of the dis-
tribution of the random effects

Session 9: Repeated Measures and Longitudinal Analysis I

Levi Waldron

Learning objectives and outline

Intro: hierarchical and longitudinal data

Fecal Fat example

Correlations within subjects (ICC)

Random and fixed effects

Summary: correlations within subjects

- Subject-to-subject variability simultaneously raises or lowers all the observations on a subject
  - induces correlation of within-subject measurements
- Variability of individual measurements can be separated into that due to subjects and that left to residual variance.
  - $var(FECFAT_{ij}) = \sigma_{subj}^2 + \sigma_\epsilon^2$
- 2-way ANOVA does not directly estimate variability due to subjects
  - variance of coefficients for individual is not too far off

**Session 9:
Repeated
Measures and
Longitudinal
Analysis I**

Levi Waldron

Learning
objectives and
outline

Intro:
hierarchical
and
longitudinal
data

Fecal Fat
example

Correlations
within
subjects
(ICC)

**Random and
fixed effects**

Summary: hierarchical data

- Estimates of coefficients (or "effect sizes") are unchanged by hierarchical modeling
- Ignoring within-subject correlations results in incorrect estimates of variance, F statistics, p-values
  - not always "conservative"
- Intraclass Correlation (ICC) provides a measure of correlation induced by grouping
- Should be able to recognize fixed and random effects