simon-5502-10-slides

Topics to be covered

- What you will learn
  - Hierarchical models
  - Hypothetical litter weights
  - Cluster randomized trials
  - Within cluster comparisons

Hierarchical data

- Moving beyond the independence assumption
- Correlation within clusters

Speaker notes

Throughout this class, I have discussed the assumptions that you need for the t-test, the chi-square test, the ANOVA test, and so forth. Every single time, I mention the assumption of independence. It’s often one that you can only check qualitatively. I mention special cases where you can’t assume independence. In this lecture, I want to talk about one of those special cases: hierarchical data.

Hierarchical data has some additional grouping factor, often called a cluster. Measurements made within a cluster are correlated with one another, violating the assumption of independence.

Examples of hierarchical data, 1 of 2

- Body parts
  - Left eye/right eye
  - Teeth
  - Skin patches
- Human families
- Animal litters

Speaker notes

A simple example of hierarchical data is when you select a group of patients and then make measurements on two or more parts of their body. You might, for example, put an eye drop medication in the left eye and a placebo drop in the right eye. You might apply different types of sealents on different teeth in a mouth. You might put different food allergens on different parts of a patient’s back.

You might select families from a population and make measurements on two or more members of the same family. Since family members share the same environment and have very similar genetics, any comparison made within a family is likely to be more precise.

Likewise, measurements on the animals from the same litter will be precise because of a shared inter-uterine environment prior to birth and shared feeding from the same mother before weaning.

Examples of hierarchical data, 2 of 2

- Clinics/hospitals
- Communities
- Repeated measurements

Speaker notes

Patients treated at the same clinic or the same hospital will often have similar outcomes. This might be caused by the location of the clinic, which determines the types of patients that come in. it might also be caused by subtle treatment practices that are agreed upon within a clinic but which might vary from one clinic to another.

You might select entire communities and then sample people within each community. You will see some level of similarity within each community because of demographic similarities or because of common dietary or cultural practices.

Often, you take measurements repeatedly on an individual under different experimental conditions.

Longitudinal data (topic for next module)

- Measurements taken at different times
  - Emphasis in changes over time

Speaker notes

A special case that I want to handle separately is longitudinal data. This is similar to repeated measures data. With longitudinal data, often the emphasis is in changes that occur over time. Repeated measurements, in contrast, emphasize different treatments with the hope that the time gaps between the measurements are small enough that you don’t see changes over time other than the changes caused by differences in what you measure and how you measure it.

Between and within cluster comparisons

- Positive correlation
  - Improves precision of within cluster comparisons
  - Hurts precision of between cluster comparisons
- Example with litters
  - Medication administered during pregnancy
  - Medication administered after birth

Basic notation, 1 of 2

- Y_{ij}
  - i defines cluster
    - i=1,…,a
  - j defines individual within cluster
    - j=1,…,n

Basic notation, 2 of 2

- $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$
  - μ unknown constant
  - α_i is normally distributed

    $SD(\alpha_i) = \sigma_{between}$
  - ϵ_{ij} is normally distributed

    $SD(\epsilon_i) = \sigma_{within}$

Some basic results

-
-
- $SD(\bar{Y}_{..}) = \sqrt{\frac{\sigma^2_{between}}{a} + \frac{\sigma^2_{within}}{an}}$
- $Corr(Y_{ij}, Y_{ik}) = \frac{\sigma^2_{between}}{\sigma^2_{between} + \sigma^2_{within}}$
  - Intraclass correlation (ICC)

Expected mean squares, 1 of 2

- $MS(between) = \frac{1}{a-1}\Sigma n(\bar{Y}_{i.} - \bar{Y}_{..})^2$

  $E[MS(between)] = n\sigma^2_{between} + \sigma^2_{within}$

Expected mean squares, 2 of 2

- $MS(within) = \frac{1}{a(n-1)}\Sigma\Sigma(\bar{Y}_{ij} - \bar{Y}_{i.})^2$

  $E[MS(within)] = \sigma^2_{within}$

Variance components estimates

- $\hat{\sigma}^2_{between} = \frac{MS(between) - MS(within)}{n}$
- $\hat{\sigma}^2_{within} = MS(within)$

Break #1

- What you have learned
  - Hierarchical models
- What’s coming next
  - Hypothetical litter weights

Description of litter weights data, 1 of 3

```
data_dictionary: litter-weights.sav

description: |
  Hypothetical data simulated to  illustrate analysis issues  associated with
random litter effects.

source: |
  Sobin, Christina; Golub, Mari (2020),  "Data for: Statistical Modeling with
Litter as a Random Effect in Mixed  Models to Manage Intralitter  Likeness",
Mendeley Data, V1,  doi: 10.17632/bwptvj2cmz.1
```

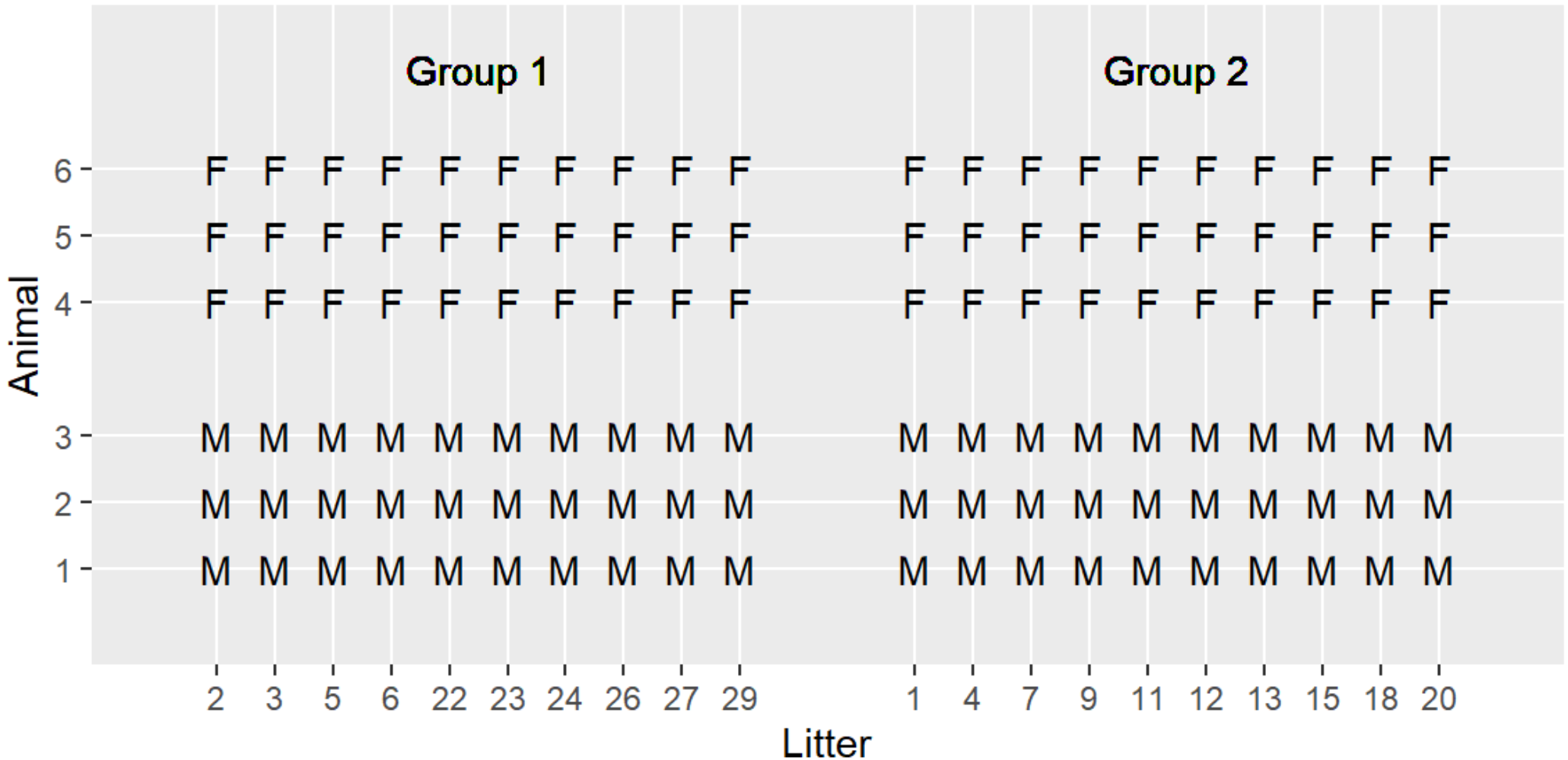
Description of litter weights data, 2 of 3

```
ID:
  label: ID number for each animal
  range: 1 to 180
LITTER:
  label: ID number for each litter
  range: 1 to 30
SEX:
  label: Unspecified sex
  range: 1 to 2
```

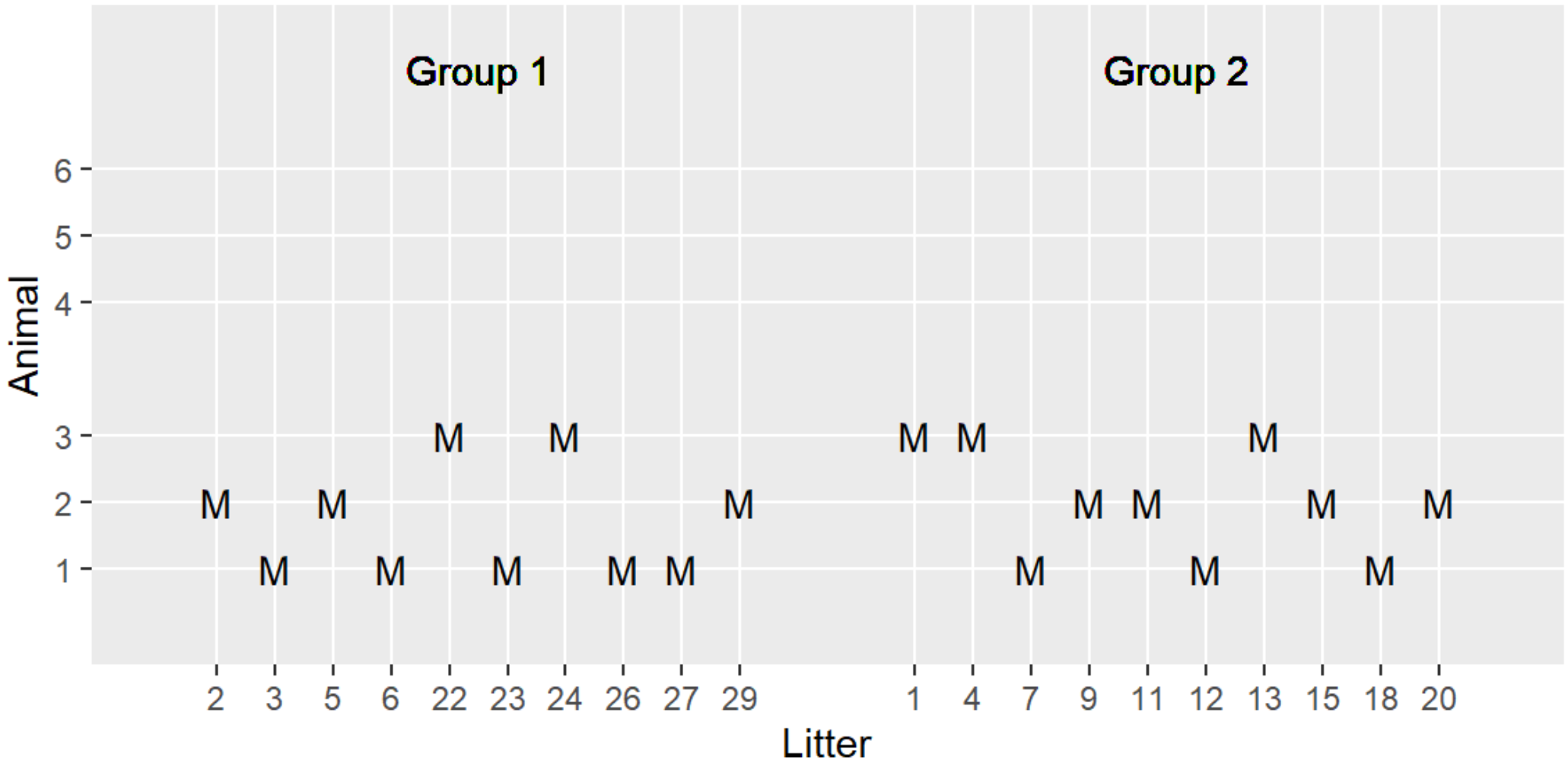
Description of litter weights data, 3 of 3

```
GRP:
  label: Unspecified group
  range: 1 to 3
WGTP21:
  label: Weight
  units: Unspecified
```

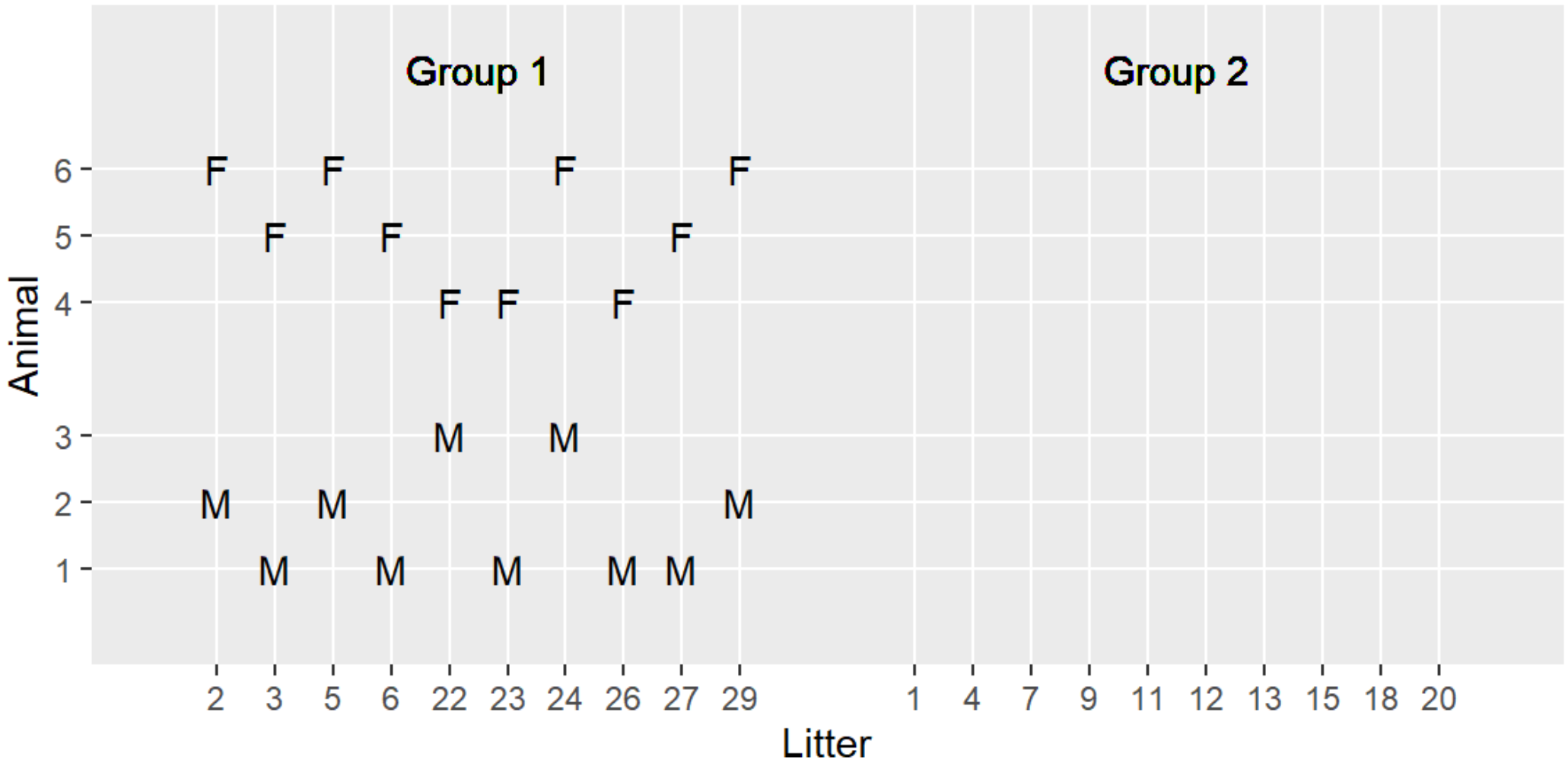
Layout of full dataset



Layout for two sample t-test



Layout for the paired t-test



Why you need to accommodate non-independence

- Increase in power and precision
- Combine multiple tests into single framework
- Handle missing data
- Allows more sophisticated analyses

How much more power/precision?

- Within cluster comparisons standard error
  - $\sqrt{\frac{\sigma^2_{within}}{an}}$
  - $\sqrt{\frac{\sigma^2_{total}}{an}(1-\rho_{ICC})}$
- Between cluster comparisons standard error
  - $\sqrt{\frac{\sigma^2_{within}}{an}+\frac{\sigma^2_{between}}{a}}$
  - $\sqrt{\frac{\sigma^2_{total}}{an}(1+(n-1)\rho_{ICC})}$

Speaker notes

These calculations only work for the simple case where you have the exact same number of observations in each cluster. The formulas illustrate an important issue that also holds when the number of observations does vary from cluster to cluster. The actual formulas, though, become quite tedious in this setting.

Pseudo-replication

- Hypothetical study of water contamination
  - Drill 5 wells
  - Sample 4 times at each well
  - Treat as 20 independent observations
- Standard error is not
  - $\sqrt{\frac{\sigma^2_{total}}{an}}$
- Standard error is
  - $\sqrt{\frac{\sigma^2_{total}}{an}(1 + (n-1)\rho_{ICC})}$

Speaker notes

Stuart H. Hurlbert. (1984), Pseudoreplication and the Design of Ecological Field Experiments. Ecological Monographs, 54: 187-211. https://doi.org/10.2307/1942661

Festing, M.F.W. The “completely randomised” and the “randomised block” are the only experimental designs suitable for widespread use in pre-clinical research. Sci Rep 10, 17577 (2020). https://doi.org/10.1038/s41598-020-74538-3

Design considerations

- Within cluster comparisons always have greater precision and power
  - Not always possible
- Best to increase number of clusters
  - Not always economical

Degrees of freedom

- df = N - k
  - k = number of estimated parameters
  - What is N?
    - N=5? (Number of wells)
    - N=20? (Number of water samples)

Break #2

- What you have learned
  - Hypothetical litter weights
- What’s coming next
  - Cluster randomized trials

Design, 1

- Designate two treatments
  - i=1,2
  - Possible to use three or more treatments

Design, 2

- Randomly assign clusters
  - j=1,…,b
  - Entire cluster gets same treatment
  - Possible to have unequal numbers within each treatment

Design, 3

- Randomly select patients within clusters
  - k=1,…,n
  - Possible to have unequal numbers within each cluster

Arthritis treatment, data dictionary

This data set shows an experiment where ten subjects with osteoarthritis were evaluated after transcutaneous electrical nerve stimulation (TENS), short wave diathermy (SWD), and after no treatment. There is no information on whether the order of the three treatments was randomized. It seems reasonable to assume that the affect of the three treatments was short term with no carry-over effects. Researchers measured pain on a visual analog scale (VAS) and range of motion (ROM).

Glimpse of original data

```
Rows: 10
Columns: 7
$ Subject <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
$ NoROM   <dbl> 35, 110, 101, 99, 126, 118, 117, 73, 95, 110
$ NoVAS   <dbl> 5.3, 2.0, 1.1, 6.3, 4.0, 0.9, 2.0, 6.1, 5.2, 2.2
$ TENSROM <dbl> 50, 90, 110, 103, 137, 89, 70, 68, 38, 87
$ TENSVAS <dbl> 3.8, 7.3, 3.6, 4.0, 1.9, 5.6, 6.6, 4.1, 7.7, 4.8
$ SWDROM  <dbl> 64, 120, 116, 135, 150, 100, 74, 93, 100, 73
$ SWDVAS  <dbl> 7.0, 1.6, 2.4, 0.8, 1.0, 2.0, 8.0, 4.5, 2.3, 4.0
```

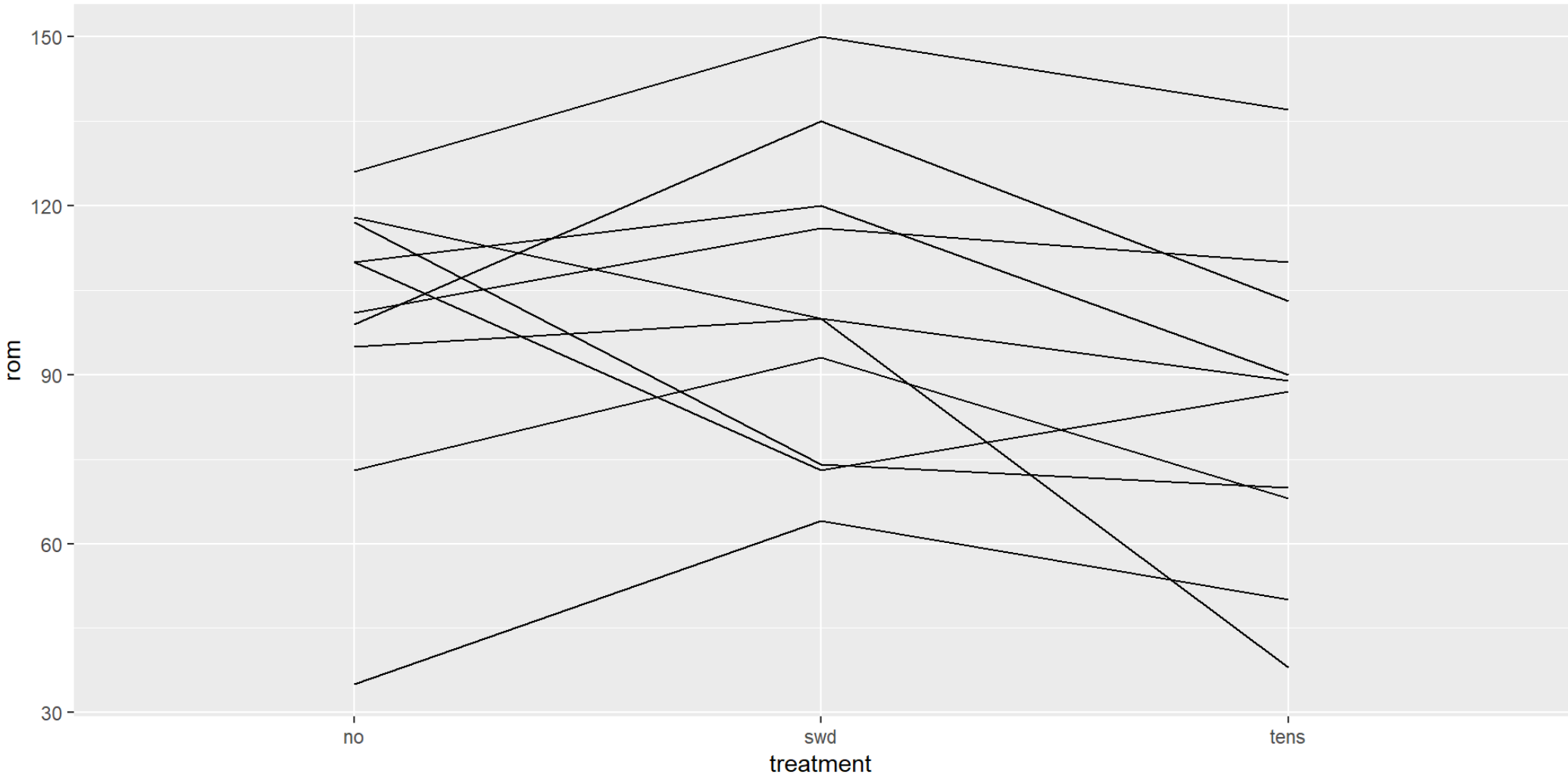
Glimpse of restructured data

```
Rows: 30
Columns: 3
$ Subject   <dbl> 1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 6, 7, 7,
…
$ treatment <chr> "no", "tens", "swd", "no", "tens", "swd", "no", "tens",
"swd…
$ rom       <dbl> 35, 50, 64, 110, 90, 120, 101, 110, 116, 99, 103, 135, 126,
…
```

Descriptive statistics

```
# A tibble: 3 × 3
  treatment rom_mean rom_sd
  <chr>        <dbl>  <dbl>
1 no            98.4   26.8
2 swd          102.    28.0
3 tens          84.2   29.1
```

Plot by subject



Mixed model, 1

```
library(lme4)

m1 <- lmer(
  rom ~ treatment + (1 | Subject),
  data=oa_2)
```

Speaker notes

Information about this code is in simon-5502-11-demo

Mixed model, 2

```
Linear mixed model fit by REML ['lmerMod']
Formula: rom ~ treatment + (1 | Subject)
   Data: oa_2

REML criterion at convergence: 253.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.8326 -0.4790  0.1129  0.6702  1.4533
```

- (You can ignore this part of the output)

Mixed model, 3

```
Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 481.7    21.95
 Residual             301.8    17.37
Number of obs: 30, groups:  Subject, 10
```

- = 21.95
- $\sigma_{between}$ = 17.37
- σ_{within}
- ICC = 481.7/(481.7+301.8)
  - = 0.61

Mixed model, 4

```
Fixed effects:
             Estimate Std. Error t value
(Intercept)    98.400      8.851  11.117
treatmentswd    4.100      7.769   0.528
treatmenttens -14.200      7.769  -1.828
```

- = 4.1 $\bar{X}_{swd} - \bar{X}_{no}$
  - t is close to zero, accept H0
- = -14.2 $\bar{X}_{tens} - \bar{X}_{no}$
  - Goes in the wrong direction!!!
  - t is close to zero, accept H0

Mixed model, 5

```
Correlation of Fixed Effects:
            (Intr) trtmnts
treatmntswd -0.439
treatmnttns -0.439  0.500
```

- (You can ignore this part of the output)

Break #3

- What you have learned
  - Cluster randomized trials
- What’s coming next
  - Within cluster comparisons

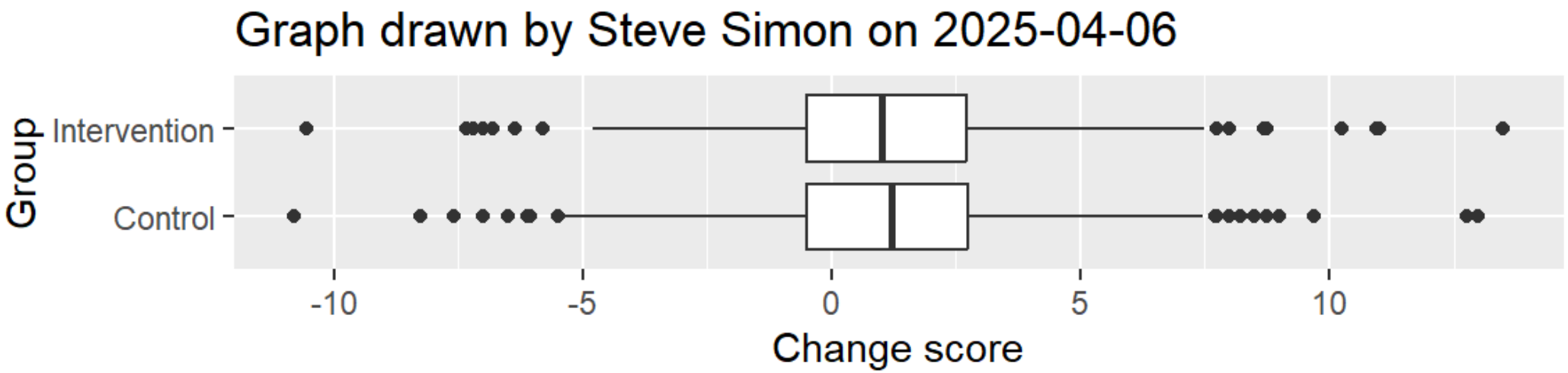
Physical activity study, data dictionary

This data set is from a cluster randomized trial in Norway. Students from 57 schools were recruited for the study. 29 schools were randomly assigned to physical activity intervention and 28 schools served as a control group. The outcomes were changes in various measures of cardiometabolic health.

Counts by school (partial listing)

	School	Group_allocation	n
1	1	Control	11
2	2	Control	10
3	3	Intervention	10
4	4	Intervention	28
5	6	Intervention	19
6	8	Control	9
7	9	Intervention	27
8	10	Control	11
9	12	Control	32
10	13	Control	20

Boxplot of change scores



Mixed model, 1

```
library(lme4)

m2 <- lmer(
  change_score ~ Group_allocation + (1 | School),
  data=pa_1)
```

Speaker notes

Information about this code is in simon-5502-11-demo

Mixed model, 2

```
Linear mixed model fit by REML ['lmerMod']
Formula: change_score ~ Group_allocation + (1 | School)
   Data: pa_1

REML criterion at convergence: 5013.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.6825 -0.5667  0.0021  0.5684  4.5686
```

- (You can ignore this part of the output)

Mixed model, 3

```
Random effects:
 Groups   Name        Variance Std.Dev.
 School   (Intercept) 1.590    1.261
 Residual             6.578    2.565
Number of obs: 1043, groups:  School, 57
```

- = 1.261 $\sigma_{between}$
- = 2.565 σ_{within}
- ICC = 1.590/(1.590+6.578)
  - = 0.19

Mixed model, 4

```
Fixed effects:
                             Estimate Std. Error t value
(Intercept)                   1.34079    0.27058   4.955
Group_allocationIntervention -0.01049    0.38135  -0.028
```

- = -0.01 $\bar{X}_{Intervention} - \bar{X}_{Control}$
  - t is close to zero, accept H0

Mixed model, 5

```
Correlation of Fixed Effects:
            (Intr)
Grp_llctnIn -0.710
```

- (You can ignore this part of the output)

Summary

- What you have learned
  - Hierarchical models
  - Hypothetical litter weights
  - Cluster randomized trials
  - Within cluster comparisons