

Module 6: Random Effects & Mixed Models

Subsampling in Mixed Models

Subsampling

Occurs when multiple observations are taken within each experimental unit.

- Introduces additional variability that must be accounted for.
- Solutions:
 - Average over the subsamples for one observation per experimental unit.
 - Use Mixed models to capture the variability both at the experimental and subsample/measurement unit levels.

Example 6.4: Turkeys in Pens

Researchers study the effect of protein % in diet on turkey growth.

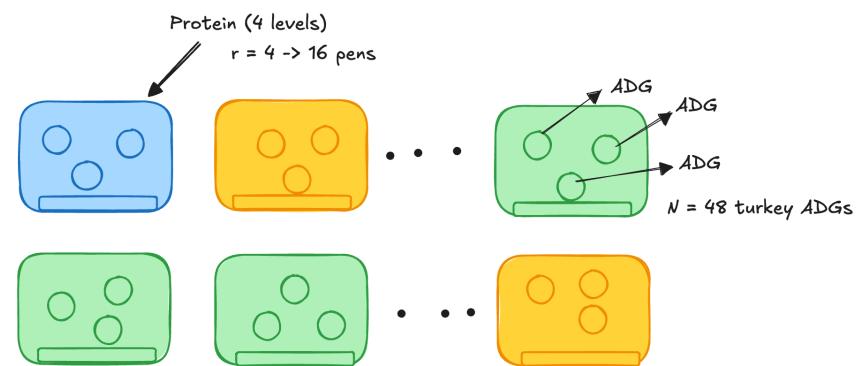
48 turkeys total.

4 diets: 15%, 20%, 25%, 30% applied to each pen.

4 pens per diet

3 turkeys per pen

Response: Average Daily Gain (ADG)



Study Structure

Treatment Structure

A one-way treatment design with 4 levels of protein diet (15%, 20%, 25%, 30%) for $t = 4$.

Design Structure

Protein diet was randomly assigned to pens (e.u.) in a CRD with $r = 4$. However, ADG is measured on each turkey (m.u.) with $n = 3$ turkeys per pen.

```
1 library(tidyverse)
2 turkey_data <- read_csv("data/06_turkey_animal_data.csv") |>
3   mutate(across(Diet:Turkey, as.factor))
4 head(turkey_data)
```

```
# A tibble: 6 × 4
  Diet   Pen   Turkey   ADG
  <fct> <fct> <fct>   <dbl>
1 15%    1     1       1.7
2 15%    1     3       1.9
3 15%    1     2       1.9
4 15%    2     2       1.3
5 15%    2     3       1.3
6 15%    2     1       1.3
```

The problem – inflate Type I error!



Incorrect Analysis - turkey level

```
1 turkey_mod0 <- lm(ADG ~ Diet, data = turkey_data)
2 anova(turkey_mod0)
```

Analysis of Variance Table

Response: ADG
Df Sum Sq Mean Sq F value Pr(>F)
Diet 3 2.2783 0.75944 9.3645 6.648e-05 ***
Residuals 44 3.5683 0.08110

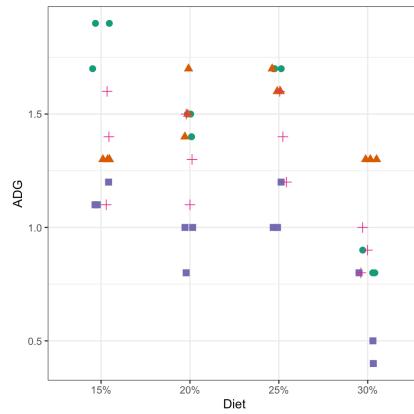
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

SV

DF: 48 turkeys - 1 = 47

Diet $(4 - 1) = 3$

Turkey(Diet) $(12-1)(4) = 44$

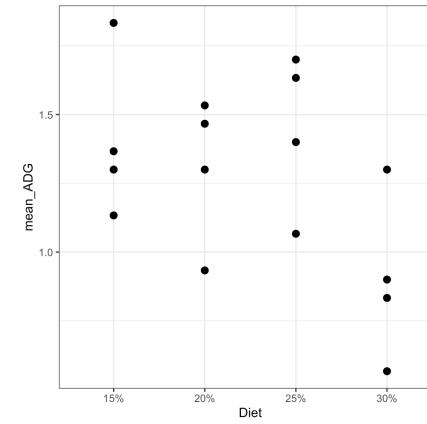


Solution 1: Average over the turkeys

$$y_{ij} = \mu + \tau_i + \epsilon_{ij} \text{ with } \epsilon_{ij} \text{ iid } \sim N(0, \sigma^2)$$

```
1 pen_data <- turkey_data |>
2   group_by(Diet, Pen) |>
3   summarize(mean_ADG = mean(ADG))
4 pen_data
```

```
# A tibble: 16 × 3
# Groups: Diet [4]
  Diet   Pen   mean_ADG
  <fct> <fct>     <dbl>
1 15%    1      1.83
2 15%    2      1.3
3 15%    3      1.13
4 15%    4      1.37
5 20%    1      1.47
6 20%    2      1.53
7 20%    3      0.933
8 20%    4      1.3
9 25%    1      1.7
10 25%   2      1.63
11 25%   3      1.07
12 25%   4      1.4
13 30%   1      0.833
14 30%   2      1.3
15 30%   3      0.567
16 30%   4      0.9
```



Solution 1: Average over the turkeys

```
1 pen_mod <- lm(mean_ADG ~ Diet, data = pen_data)
2 anova(pen_mod)
```

Analysis of Variance Table

```
Response: mean_ADG
          Df  Sum Sq Mean Sq F value Pr(>F)
Diet        3 0.75944 0.253148   3.016 0.07186 .
Residuals 12 1.00722 0.083935
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

SV	DF: 16 pens - 1 = 15
Diet	(4-1) = 3
Pen(Diet)	(4-1)(4) = 12

Solution 2: Linear Mixed Model

$$y_{ijk} = \mu + \tau_i + \epsilon_{ij} + s_{ijk} \text{ with } \epsilon_{ij} \text{ iid } \sim N(0, \sigma_\epsilon^2) \text{ and } s_{ijk} \text{ iid } \sim N(0, \sigma_s^2)$$

for $i = 1, 2, 3, 4; j = 1, 2, 3, 4; k = 1, 2, 3$

where:

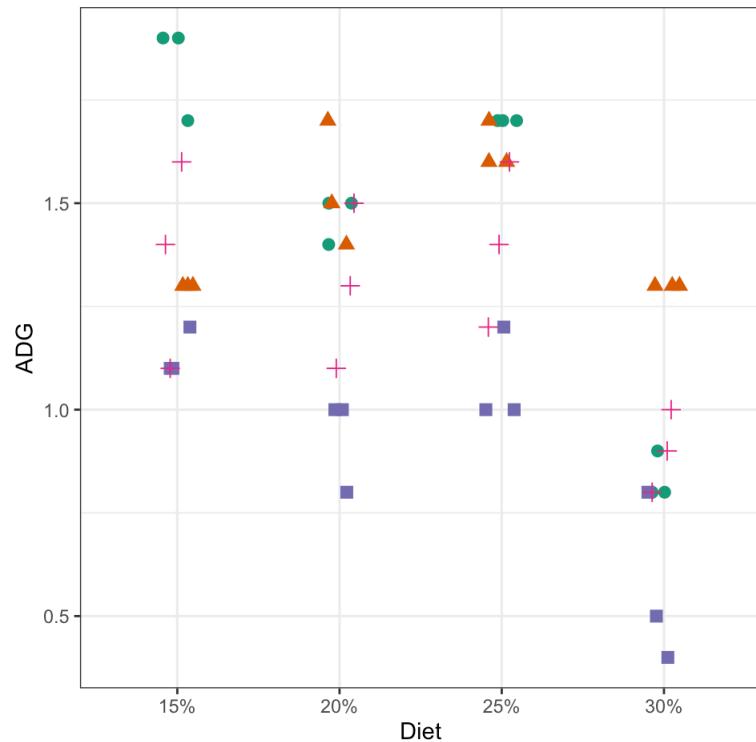
- y_{ijk} is the ADG for the k^{th} turkey in the j^{th} pen receiving the i^{th} diet
- μ is the overall mean ADG
- τ_i is the fixed effect of the i^{th} diet
- ϵ_{ij} is the random error term associated with the j^{th} pen within the i^{th} diet
- s_{ijk} is the random error term associated with the k^{th} turkey within the j^{th} pen within the i^{th} diet

Solution 2: Linear Mixed Model

SV

	DF: N - 1 where N = nrt
Treatment	t - 1
e.u.(Treatment) → experimental error σ_{ϵ}^2	(r - 1)t
m.u.(e.u x Treatment) → subsampling variability σ_s^2	(n - 1)rt

Solution 2: Linear Mixed Model



SV DF: 48 turkeys - 1 = 47

R: Fitting the LMM

```
1 library(lme4)
2 library(lmerTest)
3 turkey_mod <- lmer(ADG ~ Diet + (1 | Diet:Pen),
4                      data = turkey_data)
5 summary(turkey_mod)
```

Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]

Formula: ADG ~ Diet + (1 | Diet:Pen)
Data: turkey_data

REML criterion at convergence: -12

Scaled residuals:

Min	10	Median	30	Max
-2.06187	-0.40530	-0.00216	0.34474	1.76359

Random effects:

Groups	Name	Variance	Std.Dev.
Diet:Pen	(Intercept)	0.07824	0.2797
Residual		0.01708	0.1307

Number of obs: 48, groups: Diet:Pen, 16

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	1.40833	0.14486	12.00000	9.722	4.85e-07 ***
Diet20%	-0.10000	0.20486	12.00000	-0.488	0.6342
Diet25%	0.04167	0.20486	12.00000	0.203	0.8422
Diet30%	-0.50833	0.20486	12.00000	-2.481	0.0289 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
1 library(emmeans)
2 library(multcomp)
3 anova(turkey_mod)
```

Type III Analysis of Variance Table with Satterthwaite's method
Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
Diet 0.15457 0.051523 3 12 3.016 0.07186 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
1 turkey_emmeans <- emmeans(turkey_mod, ~ Diet, infer = c(T, T))
2 turkey_emmeans |>
3 cld(Letters = LETTERS, decreasing = T, adjust = 'tukey')
```

Diet	emmean	SE	df	lower.CL	upper.CL	t.ratio	p.value	group
25%	1.45	0.145	12	1.026	1.87	10.010	<0.0001	A
15%	1.41	0.145	12	0.985	1.83	9.722	<0.0001	A
20%	1.31	0.145	12	0.885	1.73	9.032	<0.0001	A
30%	0.90	0.145	12	0.476	1.32	6.213	0.0002	A

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

Conf-level adjustment: sidak method for 4 estimates

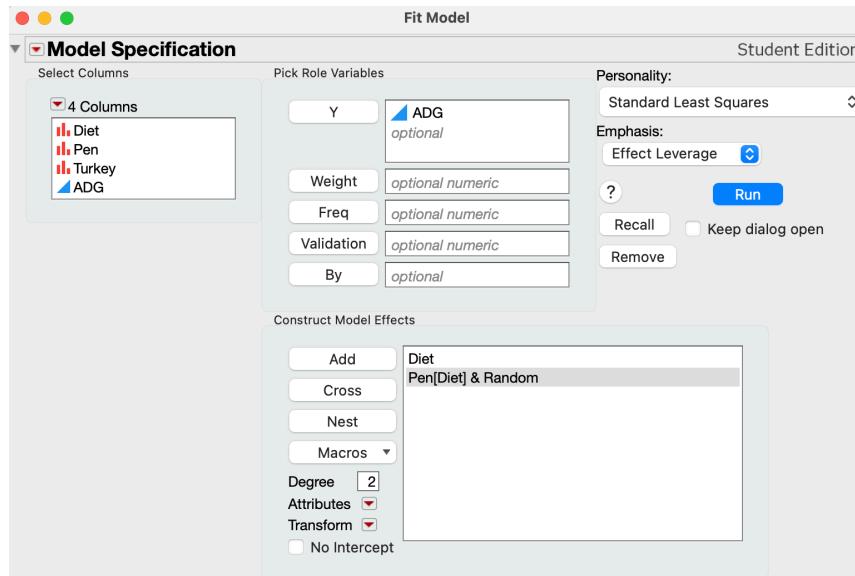
P value adjustment: sidak method for 4 tests

P value adjustment: tukey method for comparing a family of 4 estimates

significance level used: alpha = 0.05

NOTE: If two or more means share the same grouping symbol,
then we cannot show them to be different.
But we also did not show them to be the same.

JMP: Fitting the LMM



Parameter Estimates						
Term	Estimate	Std Error	DFDen	t Ratio	Prob> t	
Intercept	1.2666667	0.072429	12	17.49	<.0001*	
Diet[0.15]	0.1416667	0.125451	12	1.13	0.2809	
Diet[0.2]	0.0416667	0.125451	12	0.33	0.7455	
Diet[0.25]	0.1833333	0.125451	12	1.46	0.1696	

REML Variance Component Estimates							
Random Effect	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	Wald p-Value	Pct of Total
Pen[Diet]	4.5799458	0.0782407	0.034296	0.0110219	0.1454596	0.0225*	82.079
Residual		0.0170833	0.0042708	0.0110481	0.0298876		17.921
Total		0.0953241	0.0343845	0.0523424	0.2254076		100.000

-2 Residual Log Likelihood = -9.199233276
 Note: Total is the sum of the positive variance components.
 Total including negative estimates = 0.0953241

► Covariance Matrix of Variance Component Estimates
 ► Iterations
 ► Random Effect Predictions
 ► Fixed Effect Tests

Source	Nparm	DF	DFDen	F Ratio	Prob > F
Diet	3	3	12	3.0160	0.0719

JMP: Fitting the LMM

