Nested designs

STAT 6160 • Experimental Design • Fall 2026

Siddharth Vishwanath

University of Virginia

§1. A new reading curriculum

A new reading curriculum

A district is considering implementing a new reading curriculum in its classrooms. The question they want to answer is:

> Does the new curriculum raise reading scores?

- A teacher teaches one curriculum to the whole class, so a **classroom** receives a curriculum: the current one (C) or the new one (N).
- The district has **eight classrooms** and **40 students** in total
- Response: a student's reading score at the end of the year.

How many classrooms get the new curriculum, and how many students do you test?

Three designs

Design 1: one classroom per curriculum, 20 students per classroom.

Classroom	1	2
Curriculum	C	N

Design 2: four classrooms per curriculum, 5 students tested in each.

Classroom	1	2	3	4	5	6	7	8
Curriculum	N	C	C	N	C	N	N	C

Design 3: (hypothetical) 40 classrooms, 1 students tested in each.

Classroom	1	2	3	4	5	...	...	40
Curriculum	N	C	C	N	C	...	...	C

Questions

- What does each design compare?
- What differs between two students in the same classroom? Between two students in different classrooms?
- In design 1, what else differs between the 20 C students and the 20 N students, apart from the curriculum?
- Design 3 also tests 80 students. What are the disadvantages?

§2. Units

Experimental units

An **experimental unit** is the smallest thing to which a treatment is assigned, independently of the other units.

A **measurement unit** is the thing on which a response is recorded.

Here the classroom receives the curriculum, and the student receives the test.

Examples:

- a cage of rats is given ionized air, and each rat's activity is measured
- a batch of granola is made with one recipe, and each bar is rated
- a furnace heat is set to one temperature, and each bar in it is measured
- a reactor run uses one catalyst, and the product is assayed several times

Pseudo-replication

Pseudo-replication is analysing measurement units as if they were experimental units.

- The number of replicates of a treatment is the number of experimental units that received it.
- Design 1 has one replicate per curriculum. Its t test claims 48 degrees of freedom it did not earn.

How many degrees of freedom does the curriculum comparison have in each design?

A real one

Five determinations on a standard process **A** and five on a modified process **B**:

A	58.3	57.1	59.7	59.0	58.6
B	63.2	64.1	62.4	62.7	63.6

$$t = 8.8, \qquad \mathrm{df} = 8, \qquad p\text{-value} < 0.001$$

The five A values are five tests on **one sample from one batch** of process A. The same for B.

What does $t = 8.8$ rule out? What does it leave open?

BHH, §9.3, p. 348.

Your turn: rats in cages

Eight cages hold four rats each. Each cage, with its four rats, is exposed to either positively or negatively ionized air. The activity level of every rat is recorded.

- What is the experimental unit? What is the measurement unit?
- How many degrees of freedom does the comparison of the two kinds of air have?
- A two-sample t test on the 32 rats reports 30 degrees of freedom. What went wrong?

Owen, §9.3, after Kirk (2013).

Crossed or nested?

	Tasters	Classrooms
Can one receive every treatment?	yes	no
Relation to the treatment	crossed	nested
Role in the design	block	experimental unit
Its effect in a treatment comparison	cancels	does not cancel
How many are needed	one is enough	several each

- Tasters: the blocked taste test, where eight tasters each rated every recipe. A taster effect sits once in every recipe mean, so it cancels from $\bar{y}_{1\bullet} - \bar{y}_{2\bullet}$.
- Classrooms: this design. The N mean carries the N classroom effects and the C mean carries different ones, so nothing cancels.
- “Classroom 1 of N” and “classroom 1 of C” are different classrooms. The label 1 means nothing across curricula.

§3. The model

Notation

$$y_{ijk} = \text{score of student } k \text{ in classroom } j \text{ of curriculum } i$$

$$i = 1, \ldots, g, \qquad j = 1, \ldots, b, \qquad k = 1, \ldots, n$$

$$\bar{y}_{ij\bullet} = \frac{1}{n}\sum_{k} y_{ijk}, \qquad \bar{y}_{i\bullet\bullet} = \frac{1}{bn}\sum_{j,k} y_{ijk}, \qquad \bar{y}_{\bullet\bullet\bullet} = \frac{1}{N}\sum_{i,j,k} y_{ijk}$$

- $N = gbn$ scores, gb classrooms.
- Design 2: $g = 2, b = 4, n = 5, N = 40$.

The nested model

$$y_{ijk} = \mu + \tau_i + \eta_{j(i)} + \epsilon_{ijk}, \qquad \eta_{j(i)} \overset{iid}{\sim} \mathcal{N}(0, \sigma_C^2), \qquad \epsilon_{ijk} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2), \qquad \sum_i \tau_i = 0$$

Read $\eta_{j(i)}$ as classroom j inside curriculum i. The $\eta_{j(i)}$ and the ϵ_{ijk} are independent of each other.

- μ :
- τ_i :
- $\eta_{j(i)}$:
- ϵ_{ijk} :

Fixed or random?

Two readings of the classroom effect $\eta_{j(i)}$:

- **Fixed.** These eight classrooms are the only ones in question. The curriculum difference is a statement about them, and its error is student-to-student: $\mathrm{MS}_{\mathrm{Res}}$.
- **Random.** The classrooms are a sample of the classrooms the district could run. The curriculum difference is a statement about the curricula, averaged over classrooms, and its error includes classroom-to-classroom spread: $\mathrm{MS}_{C(T)}$.

We want the second. Nobody cares about $\eta_{2(1)}$, classroom 2 of C; we care about σ^2_C.

In the blocked taste test, each of the eight tasters rated every recipe. Does calling the taster effect β_j fixed or random change the recipe comparison there?

Assumptions

1. **Additivity.** The expected curriculum difference is the same in every classroom.
2. **Independence.** The classroom effects $\eta_{j(i)}$ and the errors ϵ_{ijk} are mutually independent.
3. **Constant variances.** $\mathrm{Var}\,(\eta_{j(i)}) = \sigma_C^2$ and $\mathrm{Var}\,(\epsilon_{ijk}) = \sigma^2$ for every curriculum and classroom.
4. **Normality.** $\eta_{j(i)}$ and ϵ_{ijk} are normally distributed.

Two students in the same classroom are **not** independent. They share $\eta_{j(i)}$.

§4. The variance of a mean

A classroom mean

$$\bar{y}_{ij\bullet} = \mu + \tau_i + \eta_{j(i)} + \bar{\epsilon}_{ij\bullet}$$

$$\text{Var}\,(\bar{y}_{ij\bullet}) =$$

A curriculum mean

$$\bar{y}_{i\bullet\bullet} = \mu + \tau_i + \bar{\eta}_{\bullet(i)} + \bar{\epsilon}_{i\bullet\bullet}$$

$$\text{Var}\,(\bar{y}_{i\bullet\bullet}) =$$

$$\text{Var}\,(\bar{y}_{2\bullet\bullet} - \bar{y}_{1\bullet\bullet}) =$$

The three designs again

With $\sigma_C = 6$ and $\sigma = 10$:

Design	b	n	$2\left(\frac{\sigma_C^2}{b} + \frac{\sigma^2}{bn}\right)$	SD	Simulated SD
One classroom, 25 tested	1	25	$2(36 + 4) = 80$	8.9	8.9
Four classrooms, 5 tested	4	5	$2(9 + 5) = 28$	5.3	5.3
Four classrooms, 25 tested	4	25	$2(9 + 0.8) = 19.6$	4.4	4.4

- Testing more students shrinks only the second term. The variance cannot fall below $2\sigma_C^2/b$: 72 with one classroom per curriculum, 18 with four.

The district offers two more classrooms, or testing every student in the eight. Which do you take?

§5. The linear model

From rows to vectors

Small case: **2 curricula, 2 classrooms each, 2 students per classroom**, $N = 8$.
Stack the rows of $y_{ijk} = \mu + \tau_i + \eta_{j(i)} + \epsilon_{ijk}$:

$$\underbrace{\begin{pmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \end{pmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}}_{\boldsymbol{X}} \begin{pmatrix} \beta_0 \\ \beta_N \end{pmatrix} + \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\boldsymbol{Z}} \begin{pmatrix} \eta_{1(1)} \\ \eta_{2(1)} \\ \eta_{1(2)} \\ \eta_{2(2)} \end{pmatrix} + \boldsymbol{\epsilon} \qquad \boxed{\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\eta} + \boldsymbol{\epsilon}}$$

- $\boldsymbol{X}$ codes the curriculum (dummy coding): $\beta_0 = \mu + \tau_1$ and $\beta_N = \tau_2 - \tau_1$. Fixed.
- $\boldsymbol{Z}$ has one column per classroom, and $\boldsymbol{Z\eta}$ repeats each classroom effect once per student. Random.

The covariance of y

$$\text{Var}\,(\boldsymbol{y}) = \text{Var}\,(\boldsymbol{Z\eta}) + \text{Var}\,(\boldsymbol{\epsilon}) = \sigma_C^2\,\boldsymbol{ZZ}^\top + \sigma^2\boldsymbol{I}$$

$$\boldsymbol{ZZ}^\top = \begin{pmatrix} \boldsymbol{J}_n & & \\ & \ddots & \\ & & \boldsymbol{J}_n \end{pmatrix} = n\boldsymbol{H}_{\text{classes}}, \qquad \text{which implies that} \qquad \boxed{\text{Var}\,(\boldsymbol{y}) = \sigma^2\boldsymbol{I} + n\sigma_C^2\boldsymbol{H}_{\text{classes}}}$$

where $\boldsymbol{H}_{\text{classes}} = \boldsymbol{Z}(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top$ is block diagonal with blocks $\frac{1}{n}\boldsymbol{J}_n$.

- Two students in the same classroom: covariance σ_C^2. In different classrooms: 0.

With $\sigma_C = 6$ and $\sigma = 10$:

```
X = np.column_stack([np.ones(8), np.repeat([0, 1], 4)]).astype(int)
Z = np.kron(np.eye(4), np.ones((2, 1))).astype(int)
V = sigma**2 * np.eye(8) + sigma_C**2 * Z @ Z.T
```

136	36	0	0	0	0	0	0
36	136	0	0	0	0	0	0
0	0	136	36	0	0	0	0
0	0	36	136	0	0	0	0
0	0	0	0	136	36	0	0
0	0	0	0	36	136	0	0
0	0	0	0	0	0	136	36
0	0	0	0	0	0	36	136

$\text{Var}\,(y_{ijk}) = \sigma_C^2 + \sigma^2 = 136$ on the diagonal; $\sigma_C^2 = 36$ inside each 2×2 classroom block.

Two projections

$$H_{\text{classes}} = Z(Z^\top Z)^{-1}Z^\top, \qquad H_X = X(X^\top X)^{-1}X^\top$$

- $H_{\text{classes}}\, y$ replaces each score by its classroom mean. $H_X\, y$ replaces it by its curriculum mean.
- Each column of X is a sum of columns of Z: a curriculum is a union of classrooms. So $H_X H_{\text{classes}} = H_X$, and the four pieces below are orthogonal projections.

$$I = \underbrace{H_1}_{\text{mean}} + \underbrace{(H_X - H_1)}_{H_T} + \underbrace{(H_{\text{classes}} - H_X)}_{H_{C(T)}} + \underbrace{(I - H_{\text{classes}})}_{\text{within classrooms}}$$

- Traces: $1, g-1, g(b-1)$, and $gb(n-1)$.

```
def hat(M):
    return M @ np.linalg.solve(M.T @ M, M.T)

H_classes, H_X = hat(Z), hat(X)
H1 = np.full((8, 8), 1 / 8)
HT, HC = H_X - H1, H_classes - H_X
parts = (H1, HT, HC, np.eye(8) - H_classes)

print(np.allclose(H_X @ H_classes, H_X))
print(np.allclose(HT @ HC, 0))
print([round(np.trace(M)) for M in parts])

X_S = np.column_stack([np.ones(8),
                       np.repeat([1, -1], 4)])
print(np.allclose(hat(X_S), H_X))
```

```
True
True
[1, 1, 2, 4]
True
```

- The last line: sum-to-zero coding of the curriculum, ± 1 in place of $0, 1$, gives the same $\boldsymbol{H}_X$. $\boldsymbol{H}$ depends on the column space, not on the coding.

§6. Sums of squares

Classrooms within curricula

Design 2 has $gb = 8$ classroom means, four under each curriculum:

	class 1	class 2	class 3	class 4	mean
C	$\bar{y}_{11\bullet}$	$\bar{y}_{12\bullet}$	$\bar{y}_{13\bullet}$	$\bar{y}_{14\bullet}$	$\bar{y}_{1\bullet\bullet}$
N	$\bar{y}_{21\bullet}$	$\bar{y}_{22\bullet}$	$\bar{y}_{23\bullet}$	$\bar{y}_{24\bullet}$	$\bar{y}_{2\bullet\bullet}$

- Curricula were assigned to classrooms, so the four C classrooms are replicates of C, and their spread is the noise of the curriculum comparison.
- **Classrooms within curricula** is that spread, pooled over curricula:

$$\mathrm{SS}_{C(T)} = n \sum_i \sum_j (\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet})^2, \qquad \mathrm{df} = \sum_i (b-1) = g(b-1).$$

- No classroom main effect: reading down a column compares two unrelated classrooms.

Classrooms within curricula

What is the one-way ANOVA of the 8 classroom means, with curriculum as the group?

The ANOVA table

Each line is a quadratic form in one of the projections:

Source	df	SS	
Curriculum	$g-1$	$bn\sum_i(\bar{y}_{i\bullet\bullet}-\bar{y}_{\bullet\bullet\bullet})^2$	$\boldsymbol{y}^\top \boldsymbol{H}_T\,\boldsymbol{y}$
Classrooms within curricula	$g(b-1)$	$n\sum_{i,j}(\bar{y}_{ij\bullet}-\bar{y}_{i\bullet\bullet})^2$	$\boldsymbol{y}^\top \boldsymbol{H}_{C(T)}\,\boldsymbol{y}$
Residual	$gb(n-1)$	$\sum_{i,j,k}(y_{ijk}-\bar{y}_{ij\bullet})^2$	$\boldsymbol{y}^\top (\boldsymbol{I}-\boldsymbol{H}_{\text{classes}})\,\boldsymbol{y}$
Total	$N-1$	$\sum_{i,j,k}(y_{ijk}-\bar{y}_{\bullet\bullet\bullet})^2$	$\boldsymbol{y}^\top (\boldsymbol{I}-\boldsymbol{H}_1)\,\boldsymbol{y}$

- Small case: $1+2+4=7$. Design 2: $1+6+32=39$.
- The first two lines live in the classroom means. The residual is within classrooms.

Which line has the degrees of freedom of the curriculum comparison?

Expected mean squares

$\mathrm{MS}_{C(T)}$ is n times the residual mean square of a one-way ANOVA on the gb classroom means, and a classroom mean has variance $\sigma_C^2 + \sigma^2/n$:

$$\mathbb{E}(\mathrm{MS}_{C(T)}) = n\left(\sigma_C^2 + \frac{\sigma^2}{n}\right) =$$

Source	df	$\mathbb{E}(\mathrm{MS})$
Curriculum	$g-1$	$\sigma^2 + n\sigma_C^2 + \frac{bn}{g-1}\sum_i \tau_i^2$
Classrooms within curricula	$g(b-1)$	$\sigma^2 + n\sigma_C^2$
Residual	$gb(n-1)$	σ^2

- The between-classroom lines are on the scale $\sigma^2 + n\sigma_C^2$. The within-classroom line is on the scale σ^2.

The F test

$$F = \frac{\mathrm{MS}_T}{\mathrm{MS}_{C(T)}} \sim F_{g-1,g(b-1)} \quad \text{under } H_0 : \tau_1 = \cdots = \tau_g = 0$$

Under H_0, from the expected mean squares:

$$\mathbb{E}(\mathrm{MS}_T) = \sigma^2 + n\sigma_C^2, \qquad \mathbb{E}(\mathrm{MS}_{C(T)}) = \sigma^2 + n\sigma_C^2$$

- Numerator and denominator estimate the same quantity under H_0. A curriculum effect inflates only the numerator.
- Curricula were given to classrooms, so the curriculum comparison is judged against classroom-to-classroom spread.

The wrong denominator. Under H_0,

$$\mathbb{E}(\mathrm{MS}_T) = \sigma^2 + n\sigma_C^2, \qquad \mathbb{E}(\mathrm{MS}_{\mathrm{Res}}) = \sigma^2, \qquad \frac{\mathrm{MS}_T}{\mathrm{MS}_{\mathrm{Res}}} \approx \frac{\sigma^2 + n\sigma_C^2}{\sigma^2} = \frac{100 + 5(36)}{100} = 2.8$$

Centred near 2.8, not 1, when the curricula are identical; near 10 with 25 students tested. It rejects too often.

The same test from the classroom means. Treat the gb classroom means as g groups of b numbers:

$$\mathrm{SS}_T = n \cdot \underbrace{b \sum_i (\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet})^2}_{\text{one-way treatment SS}}, \qquad \mathrm{SS}_{C(T)} = n \cdot \underbrace{\sum_{i,j} (\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet})^2}_{\text{one-way residual SS}}$$

Same degrees of freedom, $g - 1$ and $g(b - 1)$, and n cancels in the ratio: F is the one-way F of the classroom means.

§7. The classrooms in Python

Simulating design 2

Two curricula, four classrooms each, five students tested per classroom. Classrooms are labelled 1 to 4 **inside each curriculum**.

```
rng = np.random.default_rng(2026)
curricula, b, n = ["C", "N"], 4, 5
eta = rng.normal(0, sigma_C, 2 * b)                    # one effect per classroom

scores = pd.DataFrame({
    "curriculum": np.repeat(curricula, b * n),
    "classroom": np.tile(np.repeat([1, 2, 3, 4], n), 2),
    "student": np.tile(np.arange(1, n + 1), 2 * b),
})
scores["score"] = (70 + delta * (scores.curriculum == "N") + np.repeat(eta, n)
                   + rng.normal(0, sigma, 2 * b * n)).round(0)
```

The scores

```
wide = scores.pivot(index=["curriculum", "classroom"], columns="student", values="score")
wide["mean"] = wide.mean(axis=1)
print(wide)
```

```
student                   1     2     3     4     5  mean
curriculum classroom
C          1           63.0  63.0  72.0  70.0  65.0  66.6
           2           71.0  73.0  65.0  67.0  77.0  70.6
           3           57.0  45.0  54.0  65.0  56.0  55.4
           4           77.0  85.0  97.0  71.0  92.0  84.4
N          1           66.0  80.0  66.0  91.0  86.0  77.8
           2           84.0  63.0  79.0  67.0  68.0  72.2
           3           77.0  81.0  74.0  66.0  64.0  72.4
           4           90.0  82.0  83.0  98.0  68.0  84.2
```

- Read down the classroom means under C: they run from 55 to 84. Read across a row and the five students scatter around their classroom's level.

Two ANOVA tables*

```
fit = smf.ols("score ~ C(curriculum) + C(curriculum):C(classroom)", data=scores).fit()
table = anova_lm(fit)
table.index = ["curriculum", "classrooms within curricula", "Residual"]
print(table.round(3))
```

```
                            df  sum_sq  mean_sq      F  PR(>F)
curriculum                 1.0   547.6   547.60  7.430    0.01
classrooms within curricula  6.0  2631.9   438.65  5.952    0.00
Residual                  32.0  2358.4    73.70    NaN     NaN
```

```
ms_T, ms_C, ms_res = table["mean_sq"]
F = ms_T / ms_C
print(f"F = {ms_T:.2f} / {ms_C:.2f} = {F:.3f}")
print(f"p-value = {stats.f.sf(F, 1, 6):.3f}   on (1, 6) df")
```

```
F = 547.60 / 438.65 = 1.248
p-value = 0.307   on (1, 6) df
```

- `anova_lm` divides every mean square by $\mathrm{MS}_{\mathrm{Res}}$. The curriculum line's $F = 7.43$ on $(1, 32)$ df is $\mathrm{MS}_T / \mathrm{MS}_{\mathrm{Res}}$: the pseudo-replicated test, $p = 0.01$.
- The right test is $\mathrm{MS}_T / \mathrm{MS}_{C(T)} = 1.25$ on $(1, 6)$ df, with p -value 0.31.
- `C(curriculum):C(classroom)` says that classroom 1 of C and classroom 1 of N are different classrooms. With one label per classroom across both curricula, `C(classroom)` alone would do the same.

The true effect is 4 points. Which analysis is right, and why is it inconclusive?

The same test from the classroom means

```
means = scores.groupby(["curriculum", "classroom"]).score.mean().reset_index()
one_way = smf.ols("score ~ C(curriculum)", data=means).fit()
print(anova_lm(one_way).round(3))
```

```
                df  sum_sq  mean_sq      F  PR(>F)
C(curriculum)  1.0  109.52   109.52  1.248   0.307
Residual       6.0  526.38    87.73    NaN     NaN
```

- Average each classroom, then run the one-way ANOVA of deck 3 on the 8 classroom means.
- Same F, same p-value. The sums of squares are those of the nested table divided by $n = 5$.

Estimating σ^2_C

From the expected mean squares:

$$\hat{\sigma}^2 = \text{MS}_{\text{Res}}, \qquad \hat{\sigma}^2_C = \frac{\text{MS}_{C(T)} - \text{MS}_{\text{Res}}}{n}$$

```
scores["id"] = scores.curriculum + scores.classroom.astype(str)   # one label per class
mixed = smf.mixedlm("score ~ C(curriculum)", data=scores, groups=scores["id"]).fit()

print(f"by hand: {(ms_C - ms_res) / n:.1f}  {ms_res:.1f}")
print(f"mixedlm: {mixed.cov_re.iloc[0, 0]:.1f}  {mixed.scale:.1f}")
```

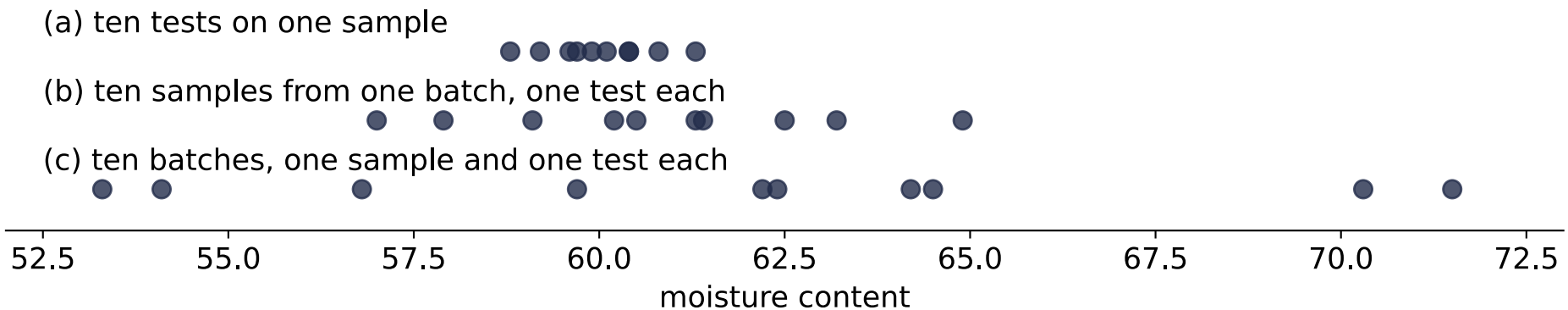
```
by hand: 73.0  73.7
mixedlm: 73.0  73.7
```

- Truth: $\sigma^2_C = 36$, $\sigma^2 = 100$. Six degrees of freedom for σ^2_C is not many.
- $\hat{\sigma}^2_C$ is negative whenever $\text{MS}_{C(T)} < \text{MS}_{\text{Res}}$. Report 0.

§9. Where does the variation come from?

Three sources of variation

The same structure in a factory: batches of pigment paste, sampled and tested for moisture.



	Variance	Estimates
(a) tests	0.55	σ_T^2
(b) samples	5.81	$\sigma_S^2 + \sigma_T^2$
(c) batches	37.82	$\sigma_P^2 + \sigma_S^2 + \sigma_T^2$

BHH, Table 9.7 and Figure 9.5, p. 347. Real data.

A nested design for three variances

15 batches; 2 samples from each batch; 2 tests on each sample. 60 determinations.

```
tests = [40,39,30,30, 26,28,25,26, 29,28,14,15, 30,31,24,24, 19,20,17,17,
         33,32,26,24, 23,24,32,33, 34,34,29,29, 27,27,31,31, 13,16,27,24,
         25,23,25,27, 29,29,31,32, 19,20,29,30, 23,23,25,25, 39,37,26,28]
paste = pd.DataFrame({"batch": np.repeat(np.arange(1, 16), 4),
                      "sample": np.tile(np.repeat([1, 2], 2), 15),
                      "y": tests})
paste_fit = smf.ols("y ~ C(batch) + C(batch):C(sample)", data=paste).fit()
paste_tab = anova_lm(paste_fit)
print(paste_tab[["df", "sum_sq", "mean_sq"]].round(2))
```

```
                     df   sum_sq  mean_sq
C(batch)           14.0  1216.23    86.87
C(batch):C(sample) 15.0   871.50    58.10
Residual           30.0    27.00     0.90
```

BHH, Table 9.8, p. 350. Samples are labelled 1 and 2 inside each batch.

Source	df	MS	$\mathbb{E}(\text{MS})$
Batches	14	86.87	$\sigma_T^2 + 2\sigma_S^2 + 4\sigma_P^2$
Samples within batches	15	58.10	$\sigma_T^2 + 2\sigma_S^2$
Tests	30	0.90	σ_T^2

```
ms_P, ms_S, ms_T2 = paste_tab["mean_sq"]
print(f"tests    {ms_T2:.2f}")
print(f"samples  {(ms_S - ms_T2) / 2:.2f}")
print(f"batches  {(ms_P - ms_S) / 4:.2f}")
```

```
tests    0.90
samples  28.60
batches  7.19
```

- Each multiplier counts the tests behind one mean: 4 per batch, 2 per sample.

Sampling is the largest source. If the next 60 determinations are yours to spend, where do they go?

§10. Close

Summary

	Students as units	Classrooms as units
Model	$\mu + \tau_i + \epsilon_{ijk}$	$\mu + \tau_i + \eta_{j(i)} + \epsilon_{ijk}$
Var $(\boldsymbol{y})$	$\sigma^2 \boldsymbol{I}$	$\sigma^2 \boldsymbol{I} + n\sigma^2_C \boldsymbol{H}_{\text{classes}}$
Error for curricula	MS_{Res}	$\text{MS}_{C(T)}$
Error df	$N - g$	$g(b-1)$
Var $(\overline{y}_{2\bullet\bullet} - \overline{y}_{1\bullet\bullet})$	$2\sigma^2/(bn)$	$2(\sigma^2_C/b + \sigma^2/(bn))$
Design 2	$F = 7.43, p = 0.01$	$F = 1.25, p = 0.31$

Before running the experiment

1. What receives the treatment? That is the experimental unit.
2. Can one unit receive every treatment? Then it is a block. Can it receive only one? Then it is nested, and the treatment error lives there.
3. How many such units does each treatment get? That is the replication.
4. Measurement units within an experimental unit improve precision only down to σ_C^2/b.

Reading

- Owen, *A First Course in Experimental Design*: §5.8, §§9.3 and 9.4, and §9.6 on cluster randomized trials.
- Meier, *ANOVA and Mixed Models*, Chapter 6: §6.1.1 and §6.1.3 (online).
- Box, Hunter, and Hunter, *Statistics for Experimenters*, 2nd ed.: §9.3.
- Dean and Voss, *Design and Analysis of Experiments*, 2nd ed.: §§18.2 to 18.4.

Appendix

The histogram panels

```
lim = (-25, 35)
fig, axes = plt.subplots(3, 1, sharex=True, figsize=(8, 5))
for ax, values, label in zip(axes, estimates, labels):
    ax.hist(values, bins=40, range=lim, color=blue)
    ax.axvline(delta, color=orange, linewidth=2)
    ax.set(title=label, yticks=[], xlim=lim)
    ax.tick_params(labelbottom=True)
axes[-1].set_xlabel("N - C")
fig.tight_layout()
plt.show()
```

The pigment-paste dot plots

```
rows = [("(a) ten tests on one sample", tests_one),
        ("(b) ten samples from one batch, one test each", samples_one),
        ("(c) ten batches, one sample and one test each", batches_ten)]
fig, ax = plt.subplots(figsize=(10, 2.7))
for level, (label, values) in enumerate(rows):
    ax.plot(values, [2 - level] * 10, "o", color=blue, ms=8, alpha=0.8)
    ax.text(52.5, 2 - level + 0.28, label, fontsize=13)
ax.set(yticks=[], xlim=(52, 73), ylim=(-0.6, 2.7), xlabel="moisture content")
ax.spines[["top", "right", "left"]].set_visible(False)
fig.tight_layout()
plt.show()
```