Summary statistics, uncertainty, and variation in calls

Eleutherodactylus Research Studio · Fall 2026

Tom Devitt

2026-10-12

Where we are

- **Today is led by your peer mentors.** Tom is away; questions about the lesson go to the mentors in the room, or to Tom on Canvas
- **Wed Oct 14:** AmphibiaWeb full draft due; mid-semester evaluations
- **Wed Oct 21:** Milestone 2 — clean data + QC notebook
- Your team notebook already has Step 5, *summarize by male, then by species*. Today is **why** it's built that way, and what to write once it runs

Three questions for today:

1. What is your *n*?
2. How variable are your frogs, and in which parameters?
3. How well do you know the mean?

Same frog, two answers

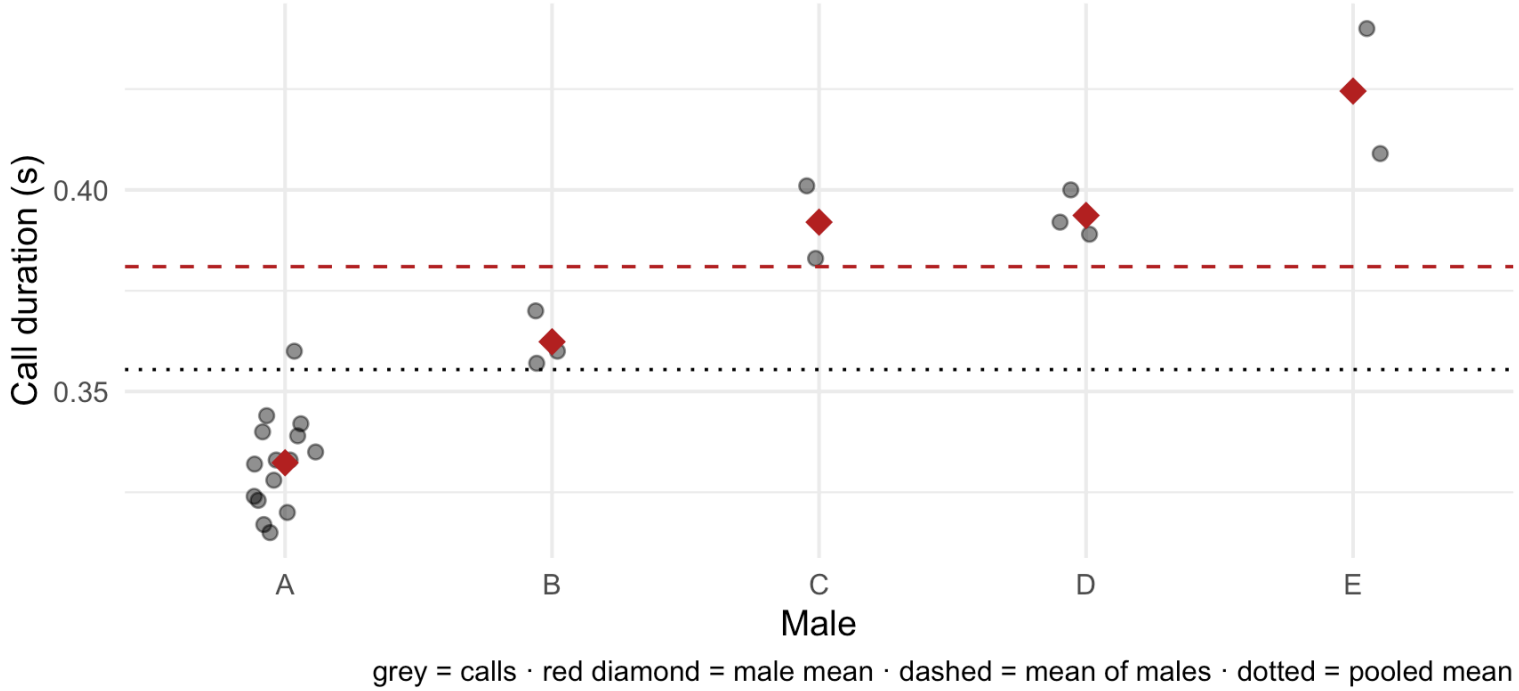
```r
set.seed(1012)
males <- tibble(
  male     = c("A", "B", "C", "D", "E"),
  n_calls  = c(15, 3, 2, 3, 2),        # male A was right by the microphone
  dur_mean = c(0.33, 0.37, 0.41, 0.39, 0.43),
  dom_mean = c(2.95, 3.20, 3.05, 3.35, 3.10),
  pul_mean = c(4, 5, 4, 6, 5)
)
calls <- males |>
  uncount(n_calls) |>
  mutate(duration_s = round(dur_mean + rnorm(n(), 0, 0.012), 3),
         dom_khz    = round(dom_mean + rnorm(n(), 0, 0.04), 2),
         n_pulses   = pmax(1, round(pul_mean + rnorm(n(), 0, 1.2)))) |>
  select(male, duration_s, dom_khz, n_pulses)
```

Simulated: one species, five males, 25 calls.

- **(a) Pool every call:** duration 0.355 ± 0.034 s, *n* = 25
- **(b) Average each male first:** duration 0.381 ± 0.035 s, *n* = 5 males

Which one belongs in a description of the species?

Calls are nested in males



- Male A's 15 calls tell you a lot about **male A.** They drag the pooled mean toward him, and they make *n* look five times bigger than it is
- Calls from one male aren't independent: counting them as separate frogs is **pseudoreplication**

The rule is already in your notebook

Two-step summary: mean of each male, then mean ± SD (range) of those male means. Report both *n*s.

```
by_male <- measurements |>
  summarise(n_calls = n(), across(c(duration_s, dom_khz), list(mean = mean)),
            .by = c(species, male, format))

by_species <- by_male |>
  summarise(n_males = n(), n_calls = sum(n_calls),
            across(ends_with("_mean"), list(mean = mean, sd = sd, min = min, max = max)),
            .by = species)
```

That's Step 5 of `team-measurement-template.qmd`, trimmed. You already have it. The SD in `by_species` is **among males**, which is the number a reader wants.

How few males is too few?

Males	Report	Don’t report
1	that male’s mean and range of calls; “*n* = 1 male”	any among-male SD
2	both male means, or the range	SD (two points don’t make one)
3+	mean ± SD (range), *n* calls from *n* males	—

- Small *n* is normal in this genus. Say it plainly: a reader can work with “*n* = 2 males,” not with a hidden one
- **Team 1:** nights on one AudioMoth are your “calls”; recorders are your “males”

Three ways to describe spread

- **SD:** typical distance from the mean, in the parameter's units
- **Range:** min–max. Always report it with small *n*; it shows the reader your actual extremes
- **CV = SD / mean:** unitless, so you can compare how variable duration is against how variable frequency is

```
cv <- \(x) 100 * sd(x) / mean(x)

within <- calls |>
  filter(n() >= 3, .by = male) |>            # need a few calls per male
  summarise(across(c(duration_s, dom_khz, n_pulses), cv), .by = male) |>
  summarise(across(-male, mean)) |> mutate(level = "within males")
among  <- by_male |>
  summarise(across(c(duration_s, dom_khz, n_pulses), cv)) |>
  mutate(level = "among males")
```

Static and dynamic properties

CV	duration	dominant freq.	pulses per call
within males	2.3%	1.1%	35.3%
among males	9.2%	4.3%	25.5%

- **Static** properties barely change within a male from call to call (Gerhardt's rough cutoff: within-male CV under ~5%). Dominant frequency is the classic one: it tracks body size
- **Dynamic** properties vary a lot within a male (CV over ~12%): rates, pulse counts. They shift with temperature and with what the neighbors are doing
- Low within-male CV is what makes a static property useful for telling species apart: a few calls from one male are a fair sample of him

Gerhardt, H. C. 1991. Female mate choice in treefrogs: static and dynamic acoustic criteria. *Animal Behaviour* 42:615–635.

Spread vs. precision

- **SD** is about the frogs: how different are the males from each other?
- **SE = SD / √n** and the **95% CI** are about you: how well do you know the species mean?

Mean call duration, 95% CI:

- 5 males: 0.338–0.424 s
- 3 of those males: 0.287–0.439 s
- all 25 calls pooled: 0.341–0.369 s. Narrow, and it misses the mean of the males

A description reports **SD and range**. Save CIs for comparisons (is species X's call longer than species Y's?). A wide interval is a result, so report it rather than hide it.

Frog + observer

Every number in your table is **the frog's call plus an observer's error**. Last week's calibration measured the second part.

- **Random error adds in quadrature.** A within-male SD of 12 ms plus an observer SD of 6 ms gives a measured SD of 13.4 ms: 12% wider than the frog. Male means barely move, because random errors average out over ten calls
- **Habits don't average out.** If one observer starts every call 10 ms early, every male that observer measured gets 10 ms longer. If that observer did all of one species, the habit **becomes a species difference**
- That's why the rules are: calibrate until no observer's habit is bigger than the observer SD, and spread each species' males across observers

Your calibration numbers

One person per team, one minute. Three numbers for the board:

Team	Duration repeatability	Observer SD (ms)	Among-male SD (ms), from `by_species`
1–6			

- Is the observer SD **small next to the differences between your frogs**? If it's half the among-male SD or more, say so in your methods
- Which parameter was most sensitive to the observer, and why? (Pulse counts and faint first pulses are the usual answer)
- Repeatability is the within/among split from the CV slide again: how much of the variation is the thing you care about

Before you summarize: two caveats

Temperature

- Rates and durations change with temperature; a call rate without a temperature is hard to compare
- Only a handful of our vouchers have a recorded temperature. **Don't fit a regression to four points.** Report the temperature where you have one, and state the gap in your caveats

Outliers

- Step 4b flags calls more than 2 SD from the species mean. **Look at the spectrogram** before you touch the number
- A male whose calls all sit away from the others is a **male**, not an outlier. Check whether the flagged calls come from one male before you call them errors

Reading a published call description

For your AmphibiaWeb account, you summarize **published** call descriptions only. Read them with today's questions:

- Is the SD among males, or among calls? Can you tell?
- How many males? How many calls per male?
- Was temperature reported?
- Was the source recording compressed (mp3) or not?

If the paper doesn't say, you can't either. Write what it reports, with its *n*.

Your turn (in your team notebook)

1. Render your notebook through Step 5
2. Add a CV column to `by_species` for dominant frequency and one dynamic parameter (rate or pulses). Is dominant frequency the static one in your species?
3. Write your team's **results sentence** under Step 5:

> Call duration 0.38 ± 0.03 s (0.33–0.42 s; *n* = 25 calls from 5 males); dominant frequency 3.12 ± 0.13 kHz. Temperature not recorded.

(Today's simulated frog. The range is of male means.)

Mp3 sources: durations, rates, counts, and dominant frequency only.

Exit ticket

On Canvas, before you leave: **Assignments → Exit ticket: Oct 12**. Everyone submits their own.

1. What is your team's *n* (males, and calls)?
2. What caveat does your results sentence have to carry?
3. Is your observer error small next to your among-male differences?

A sentence or two each is enough. Graded complete/incomplete.

Wednesday: work session + stand-ups. AmphibiaWeb draft due; mid-semester evaluations open.

Credits

- *Prepared with Claude Code (Claude Opus 5.5, Anthropic): Oct 5 additions only (observer-error slides, mentor-led notes)*