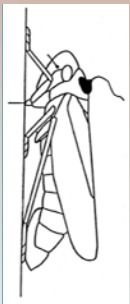


EPG Data Analysis 101

Planning Sample Sizes

- by
- T.A. Ebert
- M.E. Rogers

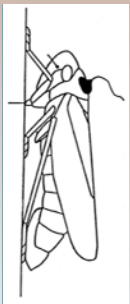
EPG
Work-
shop



The Short Answer

- At least 20 insects in each cohort
- Each insect recorded for sufficient time that every insect performs the behavior of interest.
- The rest of lecture 9 expands on this recommendation.

EPG
Work-
shop



Replication

- The insect is the experimental unit.
- In the hierarchical classification (cohort-insect-probe-event), any variable to the right of insect may have pseudoreplication. (See file “Hierarchy 02|4|7”)
- Pseudoreplication occurs where there are multiple observations from the same experimental unit in a model that does not account for the correlations between the repeated measures and where such correlations are significant.

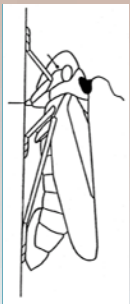
EPG
Work-
shop



What is a publishable value for n ?

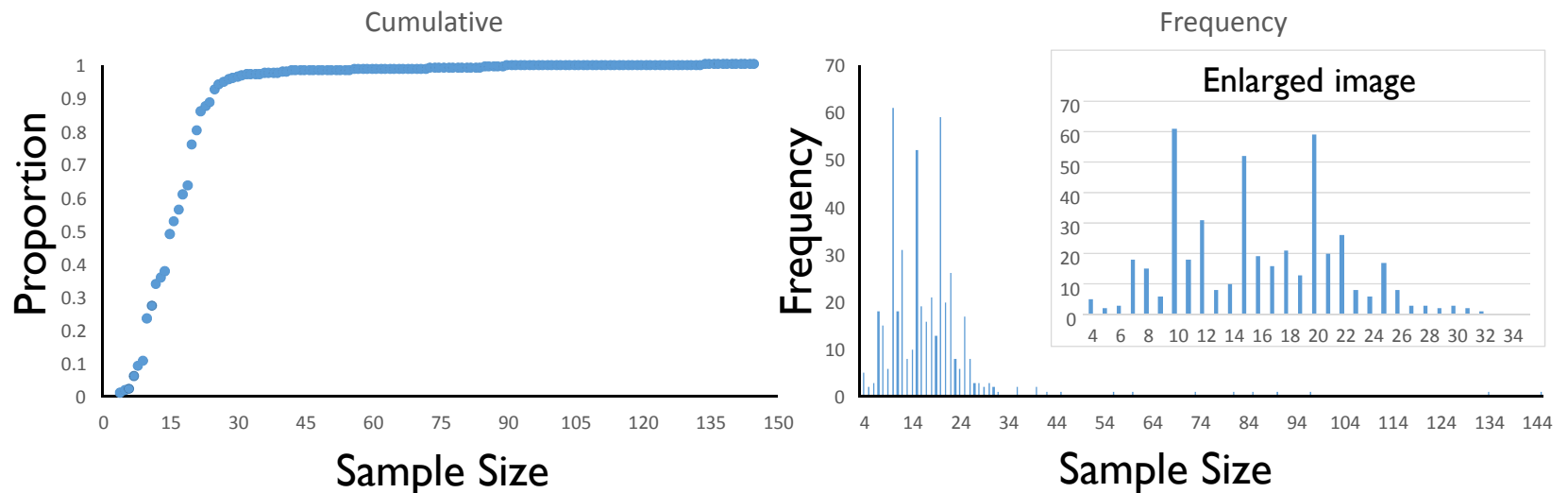
- The answer requires a distinction between planned sample size and unplanned sample size.
- The insect is the experimental unit.
- Planned sample size is the number of insects tested in each cohort.
- Unplanned sample size is the planned sample size less the number of insects that do not exhibit a specific behavior.

EPG
Work-
shop



Planned sample size distribution

- 110 articles with numbers of insects



The cumulative frequency reaches 50% at $n=15$.

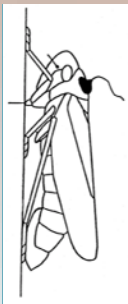
That is, half the published sample sizes are less than 15 and half are greater.

The range is $n=4$ to $n=145$

The enlarged portion of the frequency graph shows three peaks: $n=10$, $n=15$, and $n=20$.

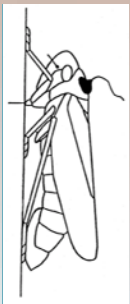
There are 471 data points. Each cohort within a study gives a unique value. So if the plan was to get 20 insects, but the author had to stop when one cohort had 19 and the others had 20, then the 19 was used for that cohort and 20 for the others.

EPG
Work-
shop

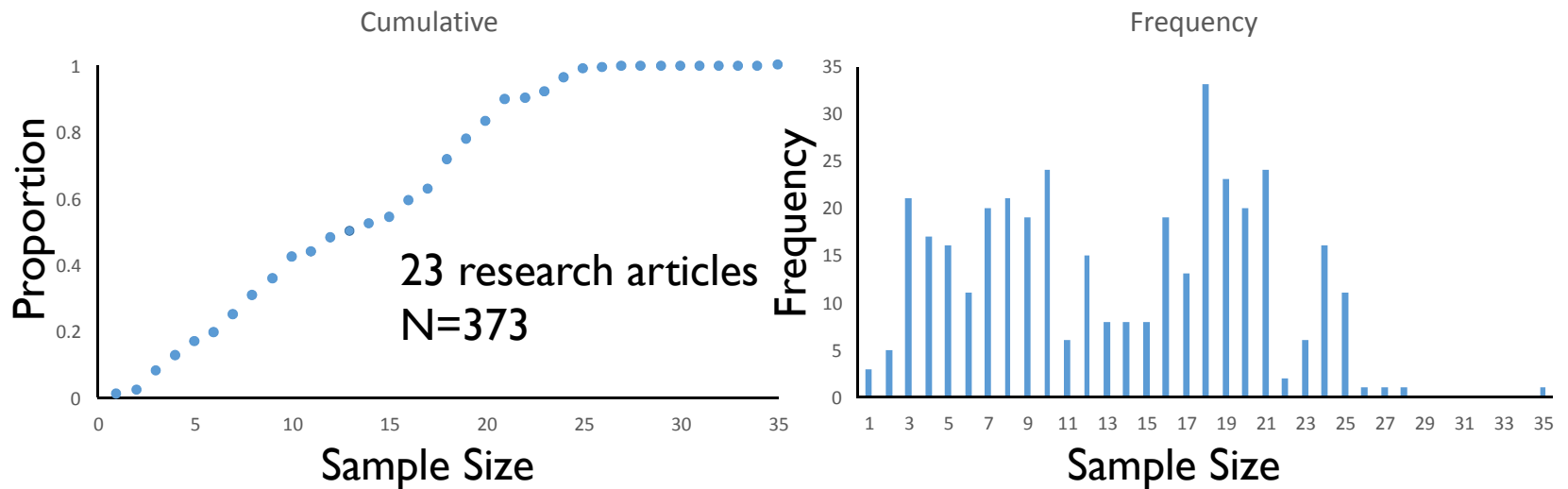


Un-planned sample size distribution

- 23 articles reported sample sizes for all variables. Ten of these also reported the planned sample size.
- Note: A sample size of one was included only if a standard deviation was given or a multiple comparison procedure result was shown.



Un-planned sample size distribution

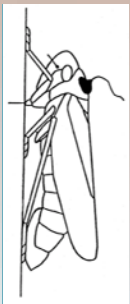


The cumulative frequency reaches 50% at about $n=13$

The range was 1 through 35

There is a peak at $n=18$

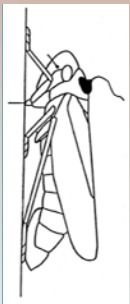
EPG
Work-
shop



The main conclusions

- Planned sample sizes are typically 10, 15, or 20 insects.
- Unplanned sample sizes have
 - a mode of 18.
 - 50% of published results have 13 or fewer replicates
 - 16.6% of published results had five or fewer replicates.

EPG
Work-
shop



Reliability

- It seems like almost anything can get published, though 10, 15, and 20 replicates are often planned.
- In the second part of lecture 9 we rephrase the problem as an issue in reliability.
 - What is the consequence to a sample size of 3?
 - Can I rely on my data, or am I looking at random noise?

EPG
Work-
shop

