Balancing Data Needs in Pedestrian Dynamics Experiments: Crowd Size, Number of Trials, and Trial Duration

Max Kinateder*[1], Paul Geoerg[2], and Nikolai Bode[3]

[1]National Research Council Canada
[2]Akkon University of Applied Sciences
[3]School of Engineering Mathematics and Technology, University of Bristol

Abstract Determining the required data for Pedestrian Evacuation Dynamics studies is essential but complex. Using an illustrative example from a study of pedestrians moving through a bottleneck, we discuss estimating crowd size, number of trials, trial duration, and how to report results. Critically, these considerations depend on the level of measurement, with different approaches needed for microscopic versus macroscopic measurements.

Keywords Sample size, Reproducibility, Pedestrian crowds, Power analysis

1 Introduction

Planning how many data points are needed to answer a research question reliably is critical to any empirical study. Many fields have established methods to estimate sample sizes, typically rooted in statistical procedures as well as theoretical and practical considerations. However, in Pedestrian Evacuation Dynamics, this answer is not necessarily straightforward, as sample size requirements can apply to the number of people needed in a single trial[1] (i.e., the crowd size) as well as the number and duration of trial repetitions for a single condition (either with the same participants or with a new crowd). Further, microscopic and macroscopic levels of measurement need to be considered. Here, we discuss three topics with relevance to *macroscopic* experimental studies of pedestrian dynamics: (1) crowd size, (2) sample size, i.e., the number of data points from the same conditions needed, and (3) duration of trials as avenues for increasing sample size. We begin with an illustrative example, demonstrate challenges, and provide recommendations for planning and reporting on each of the three topics.

2 Case study: effect of wheelchair users on egress times

The example uses a previously published study on a common experimental paradigm, in which crowds moved repeatedly through a bottleneck (crowd sizes ranged from 21 to 28). In the wheelchair condition, two participants used wheelchairs; in the control condition, all participants moved without any assistive devices [2]. Each condition was repeated more than 30 times. See [2] for details.

We sampled 30 trials per condition from the data and calculated egress times - defined as the last participant clearing the bottleneck in a given trial - from the trajectory data. Figure 1 shows the descriptive statistics of the conditions compared. The egress times were on average 4.3 s longer in the wheelchair condition compared to the control condition (t-test, t(49.00) = 7.59, $p < .001$), implying an effect size of d = 2.04 (Cohen's d - relates the mean difference to the variability).

Crowd size: Crowd size requirements fundamentally depend on the geometry of the study layout and desired microscopic or macroscopic measurements. In the example above, space was needed for the crowd to achieve a steady state of free flow before reaching the bottleneck and enough participants to generate sufficient densities at the bottleneck to impact free movement speeds. Data from pilot empirical or simulation studies can help to inform this decision.

*Email of the corresponding author: `Max.Kinateder@nrc-cnrc.gc.ca`
[1]A trial refers to an instance of data collection in an experiment; sometimes the term 'run' is used synonymously. For more information on terminology in pedestrian dynamics, see [1]
[2]There were other manipulations in the original study that are omitted from this analysis

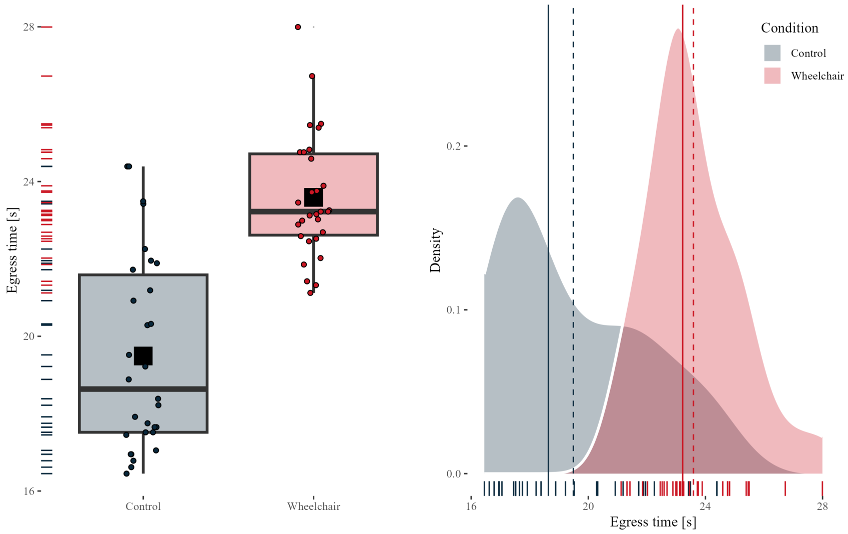

Figure 1: Egress times from multiple trials. Left: box-plot with overlaid scatter plots and marginal rug plot; squares indicate means. Right: density plot; solid lines indicate medians; dashed lines indicate means.

Sample size - number of replicate trials: The range of egress times varied considerably in the example: in 10 % of the trials in the control condition, egress times were longer than the median egress time in the wheelchair condition (Figure 1). That is, had the study only comprised a single trial per condition, it is entirely possible that the inverse results could have been observed.

Cohen introduced the concept of power analysis to estimate sample sizes [3]. Statistical power describes the likelihood of a significance test (e.g., t-test) detecting an effect when there actually is one. Power is sensitive to effect (e.g., Cohen's d) and sample size - the weaker the effect, the more data are needed to detect it reliably. For example, eight independent trials would be needed to detect an effect of $d = 2.04$ (the observed effect size in the example) with a power of 0.9. While effect sizes are often not known *a priori*, estimates can be derived from the literature or pilot studies (e.g., using simulations).

Duration of trials: Another relevant question is the duration of trials. Longer trial durations are a way to increase sample sizes by taking multiple measurements at different time points from one experimental trial. This approach is particularly useful to increase the sample size in macroscopic observations (e.g., flow). However, it assumes the time gaps between consecutive measurements are large enough to ensure independence of data, thus requiring typically relatively large crowd sizes.

3 Conclusion

Crucial in the consideration of sample size is the level of measurement. Statistically robust macroscopic findings require many replicate trials, or long time series for repeated independent measurements. Statistically robust microscopic measurements within crowds require careful accounting for interactions between individuals and the resulting correlations in their behaviour. Careful planning and transparent reporting practices allow researchers in pedestrian dynamics to generate reproducible insights.

References

[1] Juliane Adrian et al. "A Glossary for Research on Human Crowd Dynamics". en. In: *Collective Dynamics* 4 (Mar. 2019), pp. 1–13. ISSN: 2366-8539. DOI: 10.17815/CD.2019.19. URL: https://collective-dynamics.eu/index.php/cod/article/view/A19 (visited on 02/13/2025).

[2] Paul Geoerg et al. "Together Apart: the Influence of Increased Crowd Heterogeneity on Crowd Dynamics at Bottlenecks". en. In: *Collective Dynamics* 9 (June 2024), pp. 1–10. ISSN: 2366-8539. DOI: 10.17815/CD.2024.156. URL: https://collective-dynamics.eu/index.php/cod/article/view/A156 (visited on 02/13/2025).

[3] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. en. Google-Books-ID: 2v9zDAsLvA0C. Routledge, 1988. ISBN: 978-1-134-74270-7. URL: https://doi.org/10.4324/9780203771587.