Testing for over- and under-dispersion in physics degree outcomes

PRESENTER:

Astra Sword (she/her)

astra.sword@open.ac.uk

BACKGROUND:

With larger datasets available than ever, it is important—and possible!—to check if standard statistical assumptions apply in PER contexts.

METHODS

I used **population-level data** [1] to test if the **binomial assumption** holds for the **academic attainment** of UK physics students.

I achieved this by comparing **the year-on-year variance** with **the expected variance** for each physics degree programme in the data.

43 Institutions; 7 Years
26,960 Physics Graduates
From all accredited UK physics degree programmes that ran 2012/13--2018/19.

Split by Programme Type
Most institutions offer two main types of physics programme.

3-Year "Non-enhanced" (BSc)

4-Year "Enhanced" (MSc)

Calculate the Good Degree Rate for each year of each programme.
The % of students graduating with the top grades: a 1st class or a 2:1. (approx. GPA > 3.3)

Dispersion Analysis

Compare the observed year-on-year variance for each programme to the expected variance under the binomial assumption.

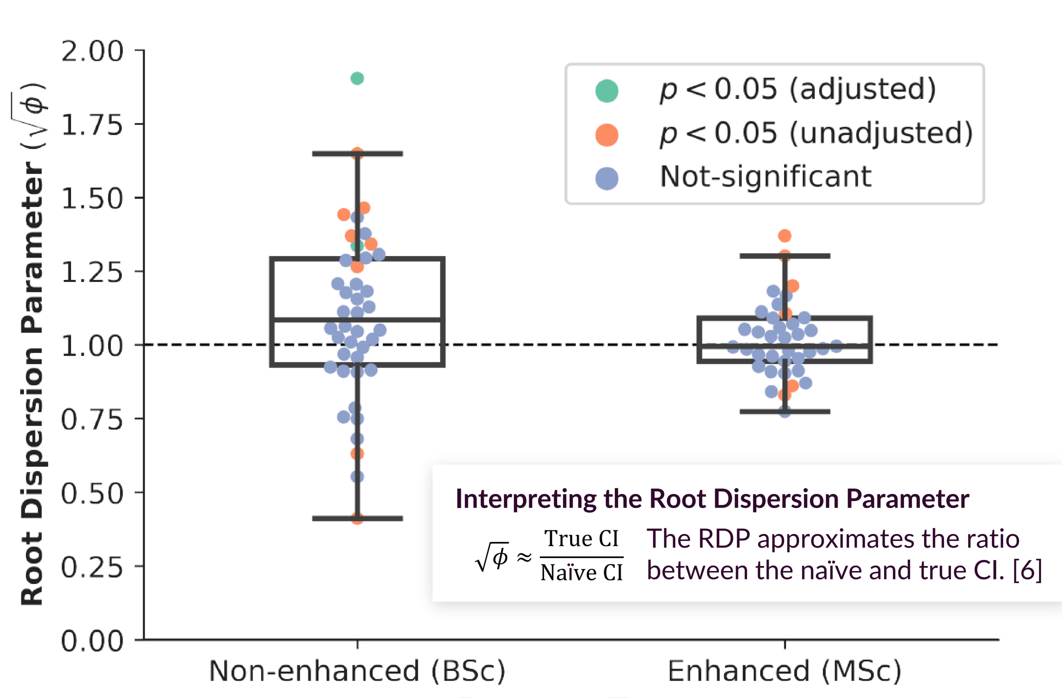
$$\text{Root Dispersion Parameter} = \sqrt{\phi} = \sqrt{\frac{\text{Observed Variance}}{\text{Expected Variance}}}$$

Calculate significance using the simulation-based approach implemented in the R package DHARMa [2] with $H_0: \sqrt{\phi} = 1; \alpha = 0.05$.

Post-hoc adjustment for false discovery:
Benjamin-Hochberg [3]
(FDR=0.05)

RESULTS

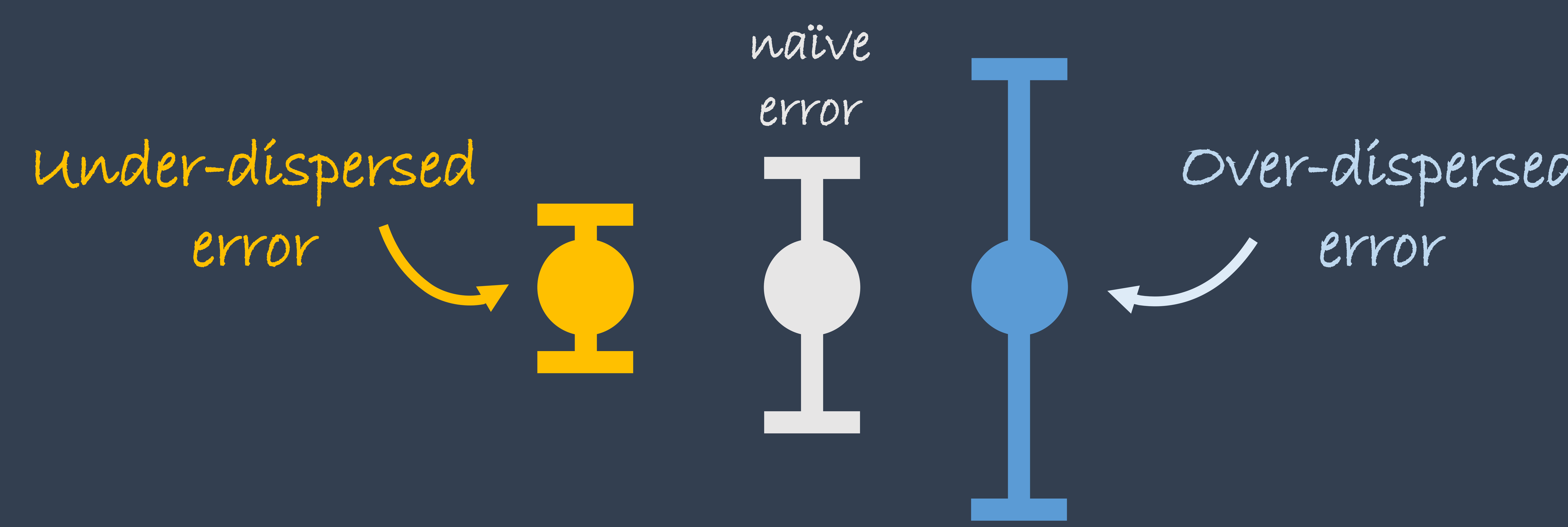
10 non-enhanced, and 6 enhanced degree programmes showed significant non-binomial dispersion ($p < 0.05$).



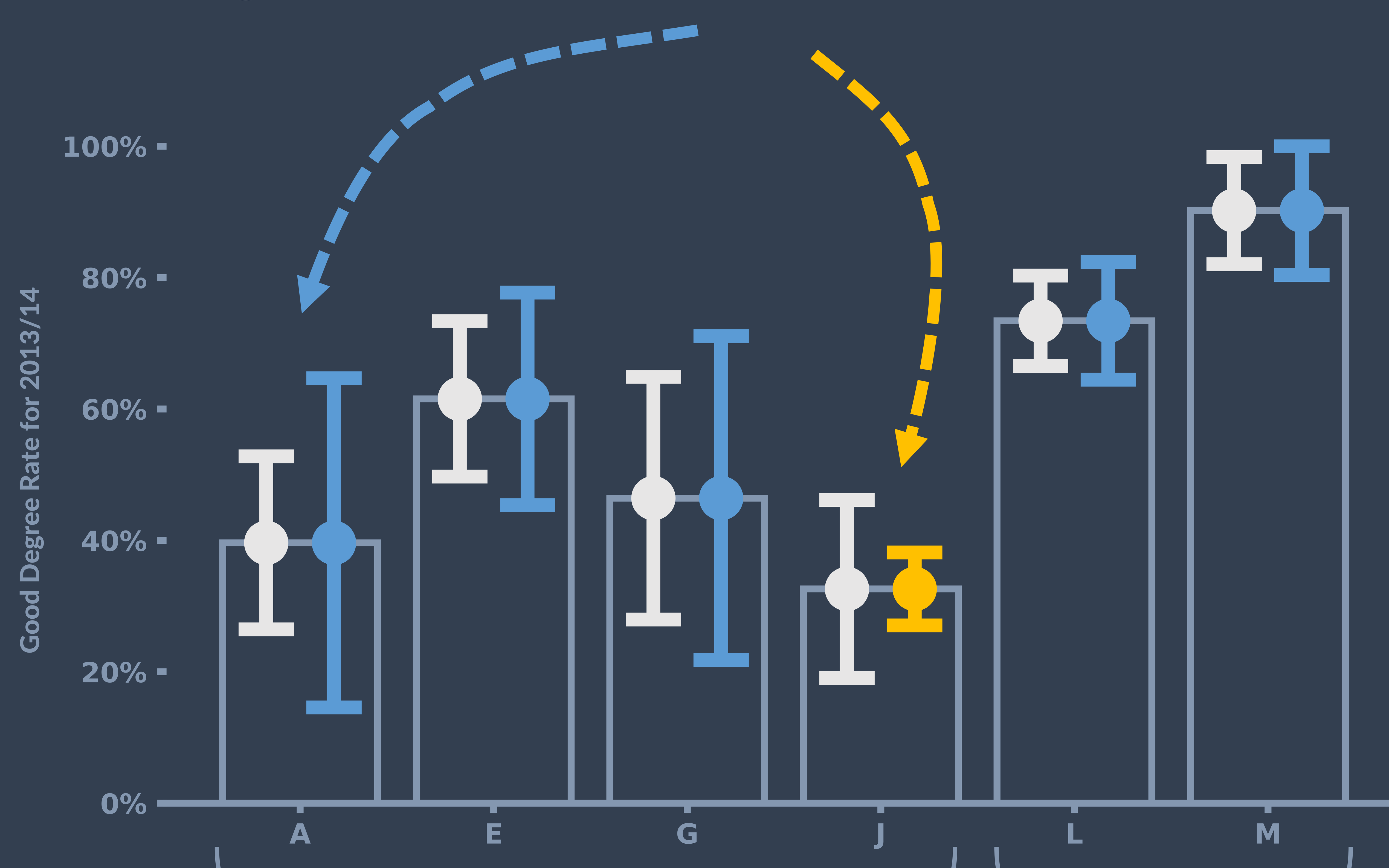
Interpreting the Root Dispersion Parameter
$\sqrt{\phi} \approx \frac{\text{True CI}}{\text{Naïve CI}}$ The RDP approximates the ratio between the naïve and true CI. [6]

After controlling the False Discovery Rate (FDR) at 5% this number dropped to 2 programmes with a further 3 being "near misses," significant at $p < 0.051$.

To illustrate, 2013/14 good degree rates for these programmes are shown on the right!



Error bars for **proportion data** using standard statistical assumptions can be **wrong by a factor of 2** for **real PER data.**

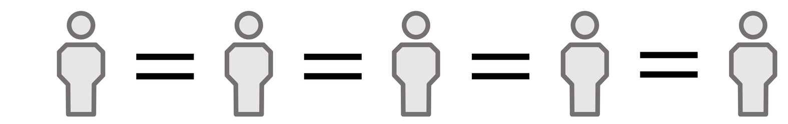


Take a picture to **download** a **digital copy**
PERC Proceedings Paper Coming Soon!
astrasword.github.io/posters/dispersion

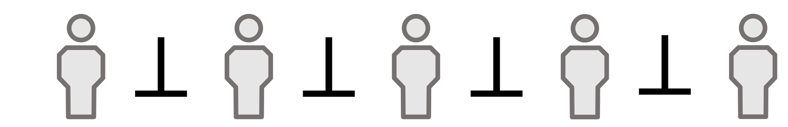
WHAT IS THE BINOMIAL ASSUMPTION?

In this context, that the chance of a student succeeding is:

1. Identical for all students.

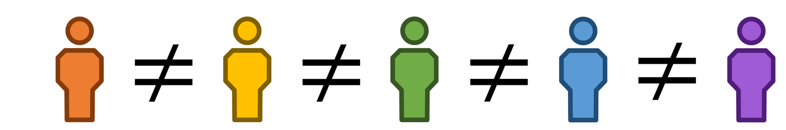


2. Statistically independent.



HOW CAN IT FAIL?

1. Diverse chances of success will lead to **overdispersion**. [4]



In the classroom:
Students are usually diverse, so this effect will be in play most of the time!

2. Correlated outcomes within cohorts also leads to **overdispersion**. [4]

In the classroom:
Happens when one student's success supports other students to succeed, e.g., collaborating on homework.

3. Anti-correlated outcomes within cohorts leads to **under-dispersion**. [4]

In the classroom:
Occurs whenever students must compete for access to learning resources or marks, e.g., grading on a curve.

Physics education is a complex social process so, in the real world, it's likely all three of these factors will be in play at once.

Tug of war:
How these forces balance out in practice is an empirical question.

WHAT CAN WE DO ABOUT IT?

Is it really a problem?: Non-binomial dispersion only matters for most statistical tests if it is present in the dependent variable *after conditioning* on the independent variables used in the model! [5]

Interpret Binomial Tests with Caution

If a statistical models makes the binomial/multinomial assumption, e.g., the χ^2-test or logistic regression, consider how much non-binomial dispersion would be needed to alter interpretation.

Use Dispersion-Aware Statistical Models

Quasi-likelihood models account for both over- and under-dispersion with some downsides. If dealing with overdispersion only, a parametric or mixed effects approach may be better. [5][6]

Measure It!

Non-binomial dispersion is a feature, not a bug! Its size and direction is a clue about what is happening in the classroom. The R package DHARMa has a simulation-based approach for this [2].

REFERENCES

[1] https://www.hesa.ac.uk/services/custom/data
[2] Hartig, Florian. *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*, 2022. http://florianhartig.github.io/DHARMa/.
[3] Benjamini, Yoav, and Yosef Hochberg. 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing'. *Journal of the Royal Statistical Society: Series B (Methodological)* 57, no. 1 (January 1995): 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
[4] Juarez-Colunga, Elizabeth, and C. B. Dean. 'Analysis of Over-and Underdispersed Data'. In *Methods and Applications of Statistics in Clinical Trials*, edited by N. Balakrishnan, 1–9. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2014. https://doi.org/10.1002/9781118596333.ch1.
[5] Davison, A. C. *Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 2003. https://doi.org/10.1017/CBO9780511815850.
[6] Agresti, Alan. *Categorical Data Analysis*. 3rd ed. Wiley Series in Probability and Statistics 792. Hoboken, NJ: Wiley, 2013.