

Introduction to Statistics

Exercises: Part 1

Sophie Lee

Exercise 1: Missing data

A mobile phone app collects user information. Some of this data are missing as users chose to opt out of location-based services.

Give one scenario where the missing data would lead to **biased** analysis results.

Give another scenario where the missing data leads to a reduced sample but could still be used to produce **unbiased** results.

Exercise 2: Summary statistics

Question 1

The following line graphs are taken from the Criminal Courts statistics quarterly report, showing the average number of days from offense to completion for defendants at the Crown Court. Which graph is the most appropriate to display the average time and why?

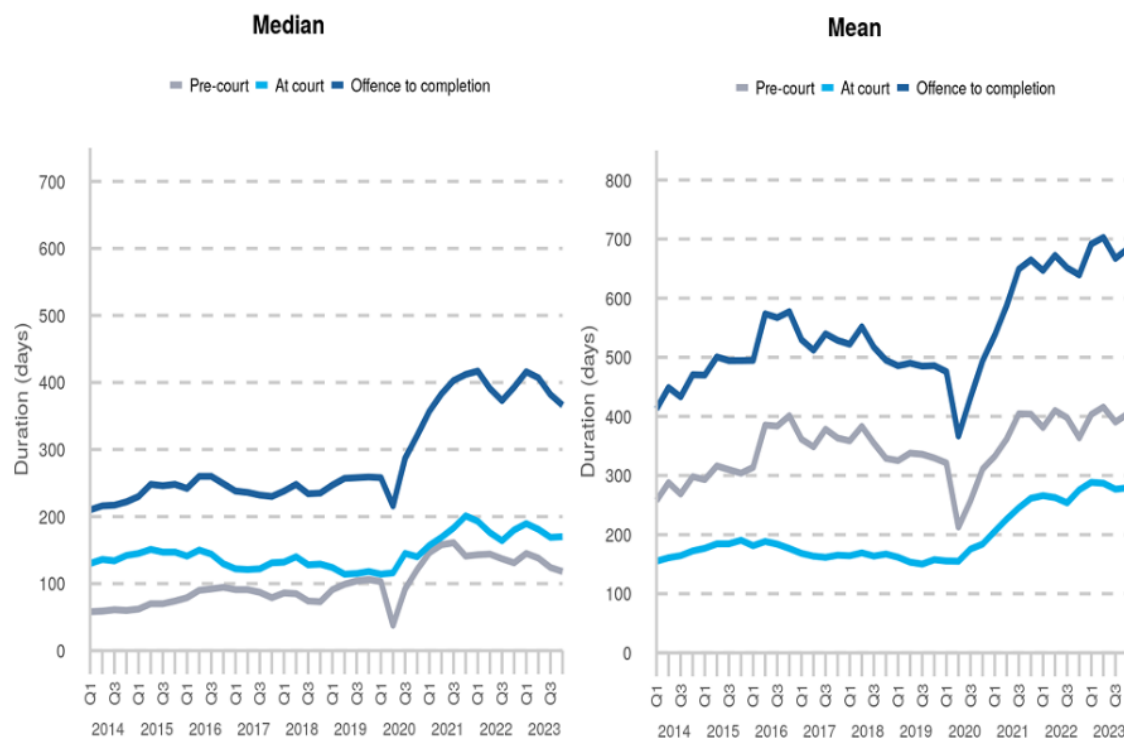


Figure 1: Average number of days from offence to completion for defendants dealt with at the Crown Court, Q1 2014 – Q4 2023

Question 2

The mean and standard deviation of waiting times in weeks for hearings in the Crown Court between 2020 and 2023 are given in the table below. Using this information, what can you tell about the distribution of these times?

| Year | Mean wait (weeks) | SD wait (weeks) |
|------|-------------------|-----------------|
| 2020 | 13.5 | 10.2 |
| 2021 | 19.1 | 15.6 |
| 2022 | 20.9 | 13.9 |
| 2023 | 22.8 | 14.7 |

Exercise 3: Inferential statistics

The output below is taken from an analysis that compared reoffending behaviour of 249 men participating with the Keyworking programme from Only Connect (OC) with those receiving standard support. More information about the findings of this report and the intervention itself can be found on the [report webpage](#).

Comment on the way in which these results have been presented. Are the results clear? What has been done well? Is there anything you think could be improved? Do the results shown appear to be valid (from the information we are given)?

Table 1: Proportion of men who committed a proven reoffence in a one-year period (reoffending rate) after support from Only Connect compared with a matched comparison group

| Number in treatment group | Number in comparison group | Treatment group rate (%) | Comparison group rate (%) | Estimated difference (% points) | Significant difference? | p-value |
|---------------------------|----------------------------|--------------------------|---------------------------|---------------------------------|-------------------------|---------|
| 249 | 31,254 | 27 | 28 | -7 to 4 | No | 0.66 |

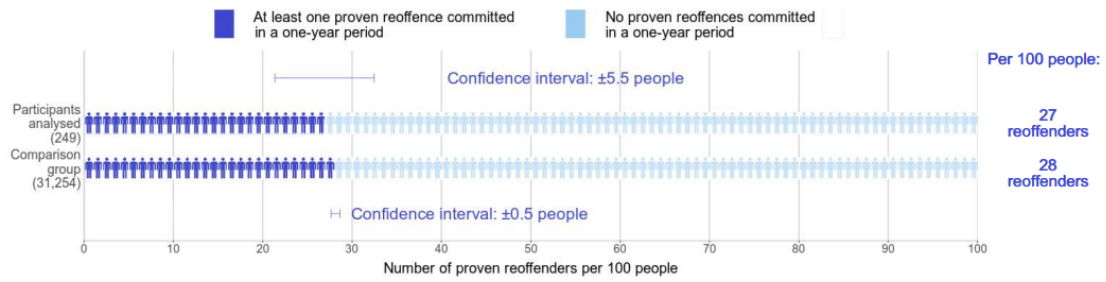
Table 2: Number of proven reoffences committed in a one-year period (reoffending frequency - offences per person) by men who received support from Only Connect compared with a matched comparison group

| Number in treatment group | Number in comparison group | Treatment group frequency | Comparison group frequency | Estimated difference | Significant difference? | p-value |
|---------------------------|----------------------------|---------------------------|----------------------------|----------------------|-------------------------|---------|
| 249 | 31,254 | 0.64 | 0.73 | -0.29 to 0.12 | No | 0.40 |

Table 3: Average time (days) to first proven reoffence in a one-year period for men who received support from Only Connect, compared with a matched comparison group (reoffenders only)

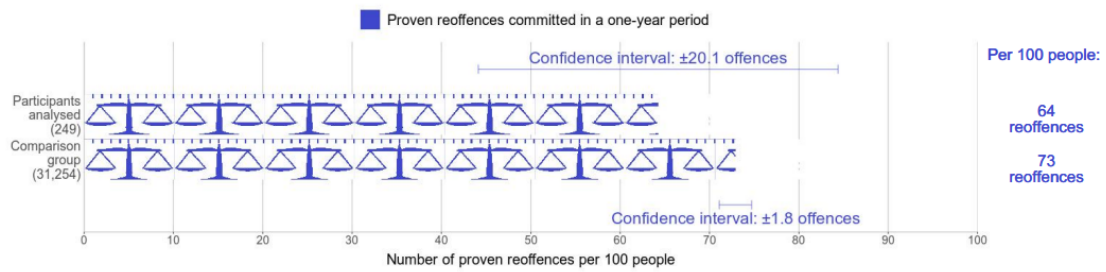
| Number in treatment group | Number in comparison group | Treatment group time (days) | Comparison group time (days) | Estimated difference | Significant difference? | p-value |
|---------------------------|----------------------------|-----------------------------|------------------------------|----------------------|-------------------------|---------|
| 67 | 7,503 | 184 | 166 | -6 to 43 | No | 0.14 |

One-year proven reoffending rate after support from Only Connect



Non-significant difference between groups

One-year proven reoffending frequency after support from Only Connect



Non-significant difference between groups