

# Problem set 1: Odd odometers

S470/670 Fall 2021

**Upload a HTML/PDF/Word document with your graphs and write-up to Canvas by 11:59 pm, Tuesday 7th September.**

## Background

A much-cited 2012 psychology paper<sup>1</sup> claimed that asking people to sign a statement of honesty before providing information, rather than afterwards, could strongly affect behavior. One piece of evidence it gave was data from an experiment conducted by an auto insurance company. According to the author, an insurance company had baseline odometer readings of miles traveled (measured some time ago) by the primary car covered under 13,488 policies (Column C in the Excel spreadsheet.) Each of the policy holders was randomly assigned to one of two conditions: “sign at the top” or “sign at the bottom” and asked to report the current mileage of the car (Column D.) The hypothesis was that signing at the top would make customers more likely to tell the truth, so the “sign at the top” group would on average report more miles traveled than the “sign at the bottom” group. The response variable was the difference between self-reported miles and baseline miles (Column D minus Column C.) The author of that section of the paper found that this difference averaged 26,100 miles for the “sign at the top” group and 23,700 miles for the “sign at the bottom” group. A  $t$ -test gives a tiny  $P$ -value, so this data would be very strong evidence that signing at the top affects behavior.

Unfortunately however, it has been recently (2021) claimed that the data was fraudulent.

## Data

The spreadsheet `DrivingdataAll.xls` contains the data set from the experiment. We will only consider the first four columns:

- **OMR Version:** “Sign Top” or “Sign Bottom”.
- **Policy # (masked):** an ID number from 1 to 13488.
- **Odom Reading 1 (Previous):** the baseline odometer reading for the primary car, in miles.
- **Odom Reading 1 (Update):** the self-reported odometer reading for the primary car, in miles.

---

<sup>1</sup><https://www.pnas.org/content/109/38/15197.short>

## Your task

Using only the first four columns of the data set, make the case that the data is suspicious.

You should submit a report containing two well-labeled graphs: a histogram and a QQ plot (hint: not a normal QQ plot). You should also include a paragraph explaining why your graphs might lead someone to think the data is suspicious. The paragraph does not need to be long, but it must be clear.

## Grading

**Graphs (5 points):** Points will be given for graphs that show unusual patterns in the data. Points will be deducted for graphs that are poorly labeled.

**Write-up (5 points):** Points will be given for clear explanations of why the graphs you showed are suspicious. Points will be deducted if we don't know what you're talking about.