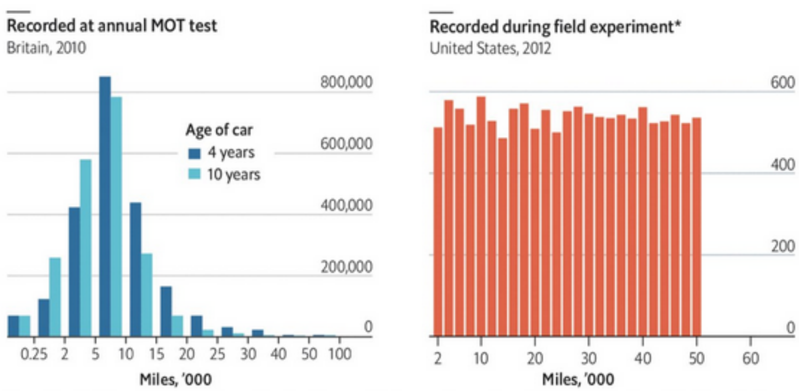
Miles Per Annum Cannot Be uniform

A study on dishonesty was based on fraudulent data

The numbers were clearly faked. No one will admit to faking them

Lies, damned lies and faked statistics

Distribution of miles driven, number of cars

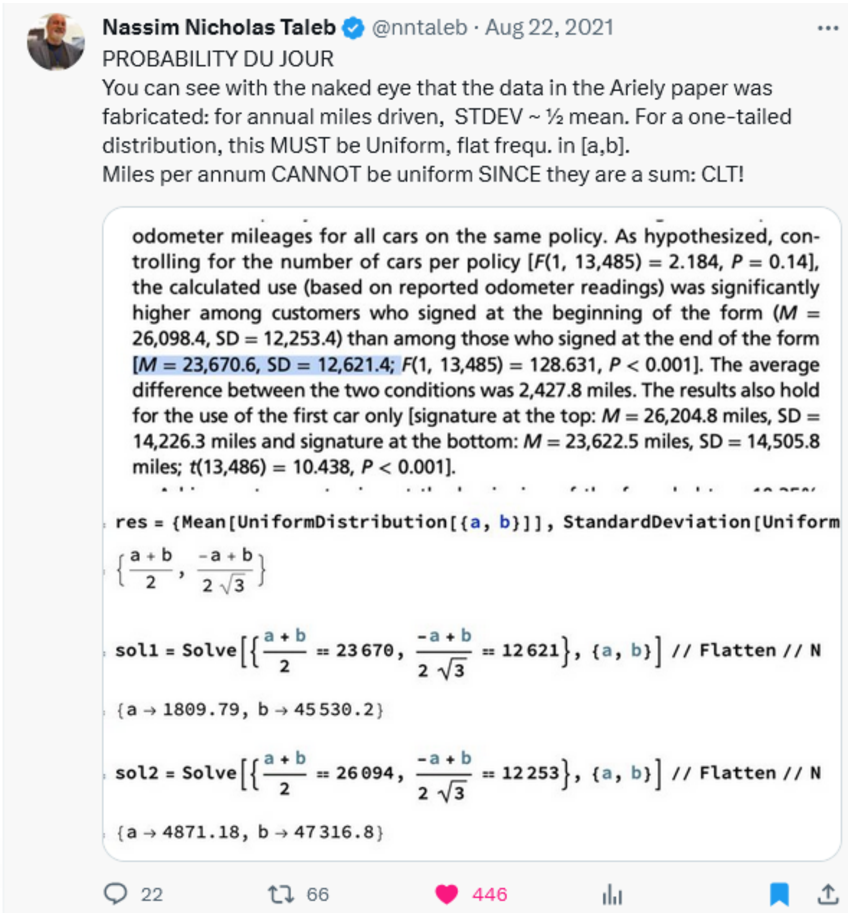


Source: Data Colada

*Mileage allegedly self-reported during experiment minus mileage before experiment. Time interval unknown

The Economist

IF YOU WRITE a book called "The Honest Truth About Dishonesty", the last thing you want to be associated with is fraud. Yet this is where Dan Ariely, a behavioural economist at Duke University, finds himself, along with his four co-authors of an influential study about lying.


Nassim Nicholas Taleb @nntaleb · Aug 22, 2021

PROBABILITY DU JOUR

You can see with the naked eye that the data in the Ariely paper was fabricated: for annual miles driven, STDEV ~ ½ mean. For a one-tailed distribution, this MUST be Uniform, flat frequ. in [a,b].

Miles per annum CANNOT be uniform SINCE they are a sum: CLT!

odometer mileages for all cars on the same policy. As hypothesized, controlling for the number of cars per policy [$F(1, 13{,}485) = 2.184$, $P = 0.14$], the calculated use (based on reported odometer readings) was significantly higher among customers who signed at the beginning of the form (M = 26,098.4, SD = 12,253.4) than among those who signed at the end of the form [M = 23,670.6, SD = 12,621.4; $F(1, 13{,}485) = 128.631$, $P < 0.001$]. The average difference between the two conditions was 2,427.8 miles. The results also hold for the use of the first car only [signature at the top: M = 26,204.8 miles, SD = 14,226.3 miles and signature at the bottom: M = 23,622.5 miles, SD = 14,505.8 miles; $t(13{,}486) = 10.438$, $P < 0.001$].

```
res = {Mean[UniformDistribution[{a, b}]], StandardDeviation[Uniform
```

$$\left\{\frac{a+b}{2}, \frac{-a+b}{2\sqrt{3}}\right\}$$

$$\text{sol1} = \text{Solve}\left[\left\{\frac{a+b}{2} == 23670, \frac{-a+b}{2\sqrt{3}} == 12621\right\}, \{a, b\}\right] \text{ // Flatten // N}$$

{a → 1809.79, b → 45530.2}

$$\text{sol2} = \text{Solve}\left[\left\{\frac{a+b}{2} == 26094, \frac{-a+b}{2\sqrt{3}} == 12253\right\}, \{a, b\}\right] \text{ // Flatten // N}$$

{a → 4871.18, b → 47316.8}

22 66 446

In[∘]:= `Mean[UniformDistribution[{a, b}]]`

Out[∘]=

$$\frac{a+b}{2}$$

In[∘]:= `StandardDeviation[UniformDistribution[{a, b}]]`

Out[∘]=

$$\frac{-a+b}{2\sqrt{3}}$$

In[∘]:=

$$\text{sol1} = \text{Solve}\left[\left\{\frac{a+b}{2} == 23670, \frac{b-a}{2\sqrt{3}} == 12621\right\}, \{a, b\}\right] \text{ // Flatten // N}$$

Out[∘]=

{a → 1809.79, b → 45530.2}

In[∘]:=

$$\text{sol1} = \text{Solve}\left[\left\{\frac{a+b}{2} == 26094, \frac{b-a}{2\sqrt{3}} == 12253\right\}, \{a, b\}\right] \text{ // Flatten // N}$$

Out[∘]=

{a → 4871.18, b → 47316.8}

Why no sum can be uniform (central limit theorem)

https://www.youtube.com/watch?v=QZ6eLXumw98