* How does the presence of outliers affect the average completion time?
  + The effect of outlier:
  + sample means jumping from the 0.998 Clean to 5.900 contaminated.
  + Median barely increase a bit from 0.693 to 0.808
  + Trimmed Mean: 0.830 to 1.080
  + The effect of the outlier causes the average completion time to increase overall, with the mean increase noticeably higher than the rest of the average estimator.
* How does the mean, median, and trimmed mean perform in the clean versus contaminated settings?
  + In the clean settings, the mean performs very well with a very small bias, and it also has the smallest MSE overall as well.
  + The Median and Trimmed Mean doesn’t change much even in contaminated data sets, but their bias is very high compared to the mean, which means that their variance is very high as well.
* Based on your findings, which estimator appears most robust to the presence of outliers?
  + the mean fluctuated a lot while the median and trimmed mean is much smaller, make the latter two more robust.
  + Thus, in contaminated dataset the mean is a good estimator for the true mixture with the outlier and it’s very efficient too with a very small MSE 1.02. But the median and trimmed means are more robust to outliers and changes, which is shown in their high bias and high MSE, but very small variance overall at around 0.076.
* How do your simulation results support your conclusions?
  + The mean has a much higher variance than the median and trimmed mean, and even itself within the clean dataset, meaning that its values vary a lot in the face of outlier.
  + However, the median and trimmed mean still retain very small variance MCSD, which showed their robustness.
  + The reason why the MSE is so high for the median is because of the bias when taking the contaminated models as the accurate truth.

2. Which estimator appears to be the most efficient in the clean data

* The most efficient is the clean with its small variance is the mean.

Yes, my answer would change, they would the trimmed mean or the median with much smaller variance than the rest.

- trimmed mean

- median, but very close to trimmed mean,

Both are very close to each other

3. real world implications

- They should consider whether their data contains very large values that could have been a outlier or not,

- whether they want that error in their data or not. Because for true users, that error should be ignored. or more robust estimator, communication of what the typical value is.

* recommended treated the 50 days or so entries as error and then.
* Thus, the average time for employees to complete phishing awareness modules after receiving an email prompt is 0.808, using the median or trimmed mean instead of the mean.

interpretation

“ensure healthy lives and promote well-being for all at all ages.” A key indicator of a population’s health is its **life expectancy**, which is often influenced by economic conditions such as **GDP per capita**.

How would you explain your findings to a policymaker or a non-technical audience?

* Using the concept of odds ratios to explain. Using the odd ratios between yes higher than life expectancy and no below life expectancy, to get a comparison between the odds of having high life expectancy. Then input the number in 10% increase in gdp per capita, you are 25% more like.

What do your results suggest about the relationship between GDPs per capita and life expectancy?

* richer countries have more GDP per capita longer life expectancy, while poor countries have been less likely to have longer life expectancy.

What are the real-world implications of your findings for achieving UNSDG Goal 3? How could this analysis help guide global health investments?

* helping to increase GDP per capita is a crucial factor in helping to achieve UNSDG goal faster. We can choose to focus on
* the threshold at 5865.739, countries that are below the median are more less likely to get higher life expectancy. While the country that is above the threshold is more likely to get higher life expectancy. Thus, global health investment should focus on investing in countries that are below the threshold. where the 10% increase in GDP per capita is 25% more likely to be higher life expectancy.