MS5052 Quantitative Research Methods for Science and Engineering

Assignment

Dara Corr - 22275193

April 26, 2023

1 Part One

(Q1) Name, and provide an example of the type of graphical display you would use to display the following. For each type you can use your own data or data provided as part of the laboratory sessions. The graphs can be produced using software of your own choosing.

i A single continuous variable.

ii Continuous data versus continuous data.

iii A single categorical variable.

i An example of a single continuous variable is the distance (in feet) for a car to stop when travelling at speed. This can be plotted using a histogram. I used data of stopping distances from R's cars dataset to show this.

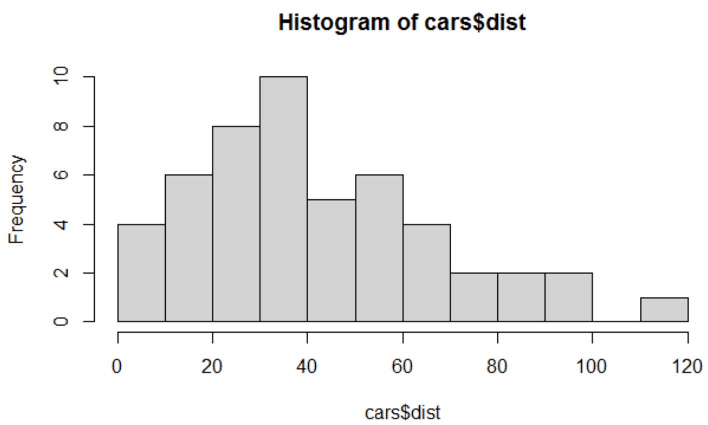

Figure 1: Histogram of Stopping distances (in ft) for different cars from R's cars dataset

ii An example of how to plot continuous data vs continuous data would be to use a scatterplot. An example of this is a plot of stopping distance vs speed from R's cars dataset.

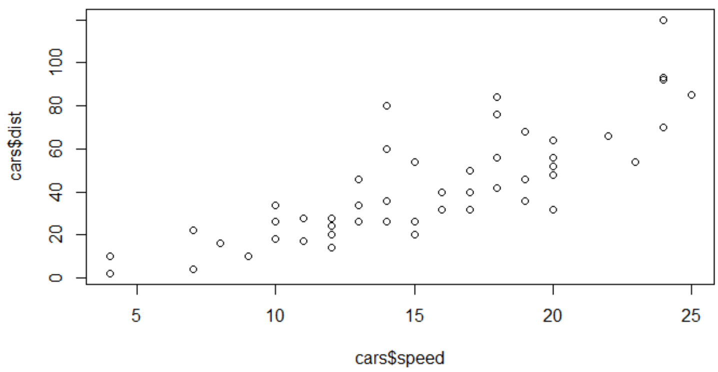

Figure 2: Scatterplot of stopping distance (ft) vs speed (mph) for different cars from the cars dataset in R

iii An example of a graphical display I would use to display a single categorical variable would be a bar chart. I used a bar chart to plot the frequency of eye colours from students in a statistics class from R's HairEyeColour dataset.

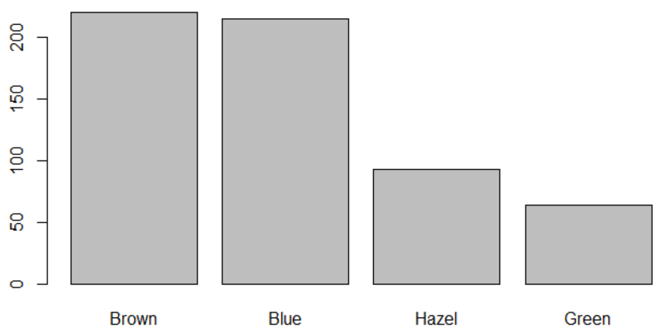

Figure 3: Barchart showing the frequencies of different eyecolours from students in a statistics class from R's HairEyeColour dataset

(Q2) A psychiatrist has recorded data which purports to measure the burden of a patient on a caregiver. Initial data are on a scale of 0-120, but they have reduced this to the categories "None", "Mild", and "Severe". Discuss the two types of data involved, and describe the key issues of concern involved in such a reduction.

The two types of data involved here are continuous numerical data and ordinal categorical data. The scores given are continuous numerical data and they are reduced into categories "None", "Mild" and "Severe" which are ordered in terms of their intensity.

One key concern from this reduction, is that there is a loss of information when the numerical variable is reduced into variables like this. If, for example, the threshold for a patient to be deemed "Severe" was set for scores of 90 and above, then a patient with a score of 89 may be classified as "Mild" when this is a borderline case. This is likely to lead to biased decisions for example, it could lead to scenarios where these patients are not given enough resources and staff to meet their needs as a result of this reduction.

Reducing continuous data into categorised groups reduces the statistical power of the model [1]. This means that because the continuous data has been grouped into categories, the probability that the model will reject the null hypothesis is reduced. This also means that Type II errors are more likely when the continuous data is reduced into categories.

(Q3) Define, and give an explanation, in language understandable by a 'lay' reader of what is meant by each of the following terms: Confidence Interval; P value; Sampling distribution of a statistic; Probability Model; Power (of a test)

Confidence Interval: A confidence interval is a range of values created with a given probability that a true value lies within it.
When measuring a statistic from data there is almost always going to be some error associated with the data measurement, especially if the data is from a sample of a population instead of the whole population. While a point estimate can give us some understanding of the underlying data, it is useful to give a range of possible values for the statistic that we are trying to measure. This range of possible values is called a confidence interval.

A confidence interval is calculated with a parameter alpha α which describes the probability that a value we are estimating will not be included in the interval. Then the probability the value will be included in the interval is $1 - \alpha$. The most common value of alpha used by statisticians is 0.05, for the 95 % confidence interval. A 95 % confidence interval interval can be interpreted as a range of values for the statistic of interest, where there is a 95 % probability that the true value of the statistic is contained in the interval.

P value: A p-value is defined as the probability that a result obtained in a test/model is as extreme (or more extreme) than the null hypothesis result when the null hypothesis is assumed to be true.

When creating a statistical model for a problem, it is of interest to a statistician whether the new model is 'better' than the old model - 'better' meaning, is the new model telling us something new that the old model did not tell us and is it statistically significant. This term 'statistically significant', gives insight as to whether a result from a statistical test or model is due to luck or whether it is actually down to a factor of interest. To measure the statistical significance of a result, statisticians use a term called the 'p-value'.

A p-value is the probability of obtaining results as extreme as the result obtained from the data, if the null hypothesis is taken to be true. So a low p-value means that it is unlikely to find a result as significant as the result found from the data, assuming that the original hypothesis is true. If this p-value is quite low (i.e. $<$ 0.5), then it is deemed that the result from the new data could not be from pure luck, and the original hypothesis is rejected in favour of the result from the new data.

Sampling distribution of a statistic: A sampling distribution of a statistic is a distribution of a statistic in a population found from taking random samples of a population, calculating the statistic on the sample with replacement of the sample back to the original population (e.g. done through bootstrapping techique).

To obtain a sampling distribution of a statistic, first take a sample of the dataset you are working with and calculate the statistic of interest on the sample. Then resample the dataset (the last sample is retained in the dataset) and calculate the statistic on the new sample again. Repeat this a large number of times, and then plotting the results will give us the sampling distibution for that statistic. An interesting example of a sampling distribution of a statistic is the sampling distribution of the sample mean which is always normally distributed for a large enough number of samples.

Probability Model: A probability model is a mathematical model used to describe random events mathematically in terms of events, sample space and probabilities of each event occuring. For example a mathematical model for flipping a fair coin, would have an event space S, with events H (head is flipped) and T (tail is flipped) with probability of event H occuring of 0.5 and probability of event T occuring of

0.5 also.

Power (of a test): The power of a test is defined as the probability that the null hypothesis H_0 is correctly rejected when the null hypothesis H_0 is false and the alternative hypothesis H_1 is true. This is often represented mathematically as $power = Pr(reject\ H_0|H_1\ true) = 1 - \beta$, where β is the probability of a type II error, where the null hypothesis is not rejected even though the null hypothesis is false. The greater the power of a test is, the smaller the chances are of failing to reject the null hypothesis when the alternative hypothesis is true.

(Q4) A medical researcher wished to investigate whether a treatment has a controlling effect on the behaviour of patients suffering from cognitive impairment. Table 2 contains "burden" scores for patients before and after treatment with the drug. A high score is interpreted as larger "burden", and is less desirable than a lower score.

(a) Calculate the following for the difference in scores: arithmetic mean; median; sample standard deviation; range.

(b) A psychologist does a statistical test on the above data, and gets a p value of 0.0083. In the process of discussing the meaning of the p value at a conference, the psychologist makes the following statement "We got a p value of less than one percent. Since this is the probability that the null hypothesis is true, we reject H_0 and conclude that there is a real improvement for the patients under treatment." Explain the error the psychologist has made in interpreting the p value.

(c) Part of the field of statistical inference is concerned with estimation of population parameters using sampled data. Describe the population of interest in this experiment, two population parameters of interest, and explain which sample statistics estimate these parameters.

(d) Calculate a 95% confidence interval for the mean difference in scores in this case. Interpret the result of your calculation.

(a) Mean difference $\bar(X) = 12.4$

Median difference $= 15.5$

Sample standard deviation $s = 7.85$

Range = [0, 23]
calculations done in Rstudio.

(b) The error the scientist made in interpreting the p-value is thinking of it as the probability that the null hypothesis is true, when it is the probability that, given the null hypothesis being true, that a result from a new test/model will have results as extreme (or more extreme) than the null hypothesis. It is the probability that the new result is as extreme or more extreme than the null hypothesis, assuming that the null hypothesis is true, whereas the psychologist's interpretation of the p-value is that it is a probability that the null hypothesis is true.

(c) The population of interest in this study are people with cognitive impairment. Two population parameters of interest are the true mean μ of the difference in burden score from the population and the population standard deviation in burden scores σ. These parameters can be estimated using the sample variance s and the sample sample mean $\bar{X}$ that I calculated in part (a) from this sample.

(d) Using a 95 % confidence interval of the form

$$\begin{aligned} CI &= \bar{X} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} \\ &= \bar{X} \pm 2.262\frac{7.85}{\sqrt{10}} \end{aligned}$$

I obtained a confidence interval of [6.79,18.01] using n = 10 and $\alpha = 0.05$. This means, we would expect 95 % of patients suffering from cognitive impairment to see a decrease in their burden score of between 6.79 and 18.01. It is worth keeping in mind that this test was done on a very small sample size, so there is a larger element of uncertainty associated with this test.

This study implies that this treatment is effective treatment for people suffering from cognitive impairment, but to get a better idea of the drug treatment's effectiveness, a larger sample size would give a more precise measurement of the mean difference in scores. Maybe the researchers should also include a group that is given a placebo drug, to test if the reduction in burder is due to some other random variable not accounted for in the test.

(Q5) (a) The tour operator has totally misinterpreted the t-interval.

(i) Explain in your own words what a confidence interval actually means.

(ii) State any underlying assumptions about what the individual measurements represent that are implicit in using such a method to summarise data.

(iii) Comment on the statement “You will be 95% sure of no more than 1.16mm of rain in any one day.

(Q5) (b) One of the couple had studied statistics, and was not too put off by the elaborate detail the operator had provided. The actual data were requested before any holiday was booked. These are reproduced in Table 1. (overleaf) For these data

(i) Construct the graphical summary you feel is most useful.

(ii) Comment on the main features of these data, which have not been highlighted in the operator’s report.

(iii) In light of the data, comment on the appropriateness, or otherwise, of the confidence interval methodology used by the analysts in their summary.

(a) (i) A confidence interval is a range of values that has a probability $1 - \alpha$ of containing the true value of the parameter that is being estimated. For example, if a test is being done to measure mean heartrate of athletes, using data from a sample of athletes from the University of Limerick, then a 95% confidence interval of the sample means will have a 95 % probability of containing the true mean value for all the athletes who study at the University of Limerick . It does not mean that one is "95 % Confident" that a statistic is correct, but instead that there is a 95 % probability of the true statistic being contained in the interval.

(ii) The underlying measurements are 5 year averages of the rainfall for each day in June, because of this, there is a reasonable chance that the data is biased and unreliable. This is because, some years there may have been almost no rain for weeks in the Summer months and then other years there may have been heavy rainfall most days, then the heavy rainfall will be averaged out with the years that day had no rainfall and the overall result will have a 5-year average of low rainfall on that particular day.

The rainfall is unlikely to have been very similar on each day in June in this 5-year timeframe given the chaotic nature of weather patterns. For example, if there are 4 years the 1st of June has no rain, but one year that there are 5mm of rainfall, then the 5-year average for that day will be 1mm, which is less than the average the operator stated. But this extreme example shows that the 5 year averages can be quite low even if there are frequent heavy rain and storms in the area. Also, consider a day that has had a large amount of average rainfall in the last 5 years, such as the 25th of June from this dataset.

The problem with using the mean as an average in this context is that the arithmetic mean is being used to estimate a central value on the rainfall on this island. But the rainfall data can fall very far away from the central values on rainy and dry days - which ultimately does not make 5-year averages a good predictor for daily rainfall. Median rainfall over the five-years would possibly be a better estimator for the average amount of rainfall one could expect on this island on this month.

Meteorologists and weather forecasters have stated that models used to predict the wether are only effective for predicting weather for up to two weeks in the future [2]. In my opinion, the 5-year average measurements of rainfall for each day in June is not a good prediction of the weather trends - short term weather predictions during the Summer. But the 5-year average may be useful to give climate trends i.e. Do Summers on this island tend to be rainy and wild or dry and mild?

(iii) The statement "You will be 95 % sure of no more than 1.16mm of rain in any one day." is not a correct interpretation of the confidence interval. The travel agent has misinterpreted the meaning of a confidence interval and thinks that it means that there is 95 % certainty that the range of the confidence interval for the 5-year average rainfall will not be exceeded. While in reality, the confidence interval is actually a range of values in which there is a 95 % probability that the true average daily rainfall figure is contained in.

The use of arithmetic mean for the average rainfall used in this confidence interval is not appropriate as the rainfall data is potentially not centred around a central value if there is a mix of very rainy and very dry days for example. It is more beneficial for the travel agent to use an optimistic statistic like average 5-year rainfall using the arithmetic mean rather than a more pessimistic statistic like the median, since the biased nature of the arithmetic mean in this context is in his favour to market this island destination to tourists.

The 5-year average rainfall per day is also not a very intuitive statistic for people understand. Better examples of statistics would be number of rainy days in that month for each year or total rainfall in mm for each year. Number of rainy rays is a very easy statistic for people to understand and can be calculated by observing how many days had rainfall exceeding a certain threshold (e.g. 1mm in a day). total rainfall in mm per month is very easy to compare with rainfall in other tourist destinations and makes comparisons very easy.

(b) (i) I created a plot of rainfall vs day, from the table above containing the 5-year averaged daily rainfall data.

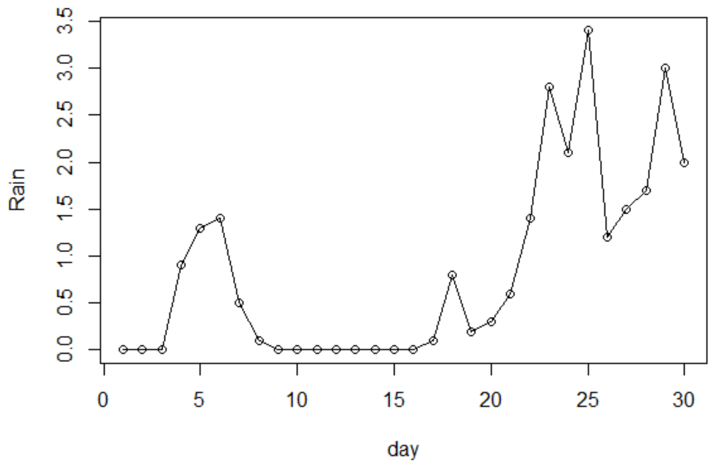

Figure 4: Line plot of 5-year averaged daily rainfall for the month of June

(ii) The plot shows that there are several spikes in the plot, whereas one would expect a distribution that is smoother. The days with large spikes likely indicate that there was large rainfall on a particular day which drove the 5-year average rainfall up.

From this plot, I believe that there is likely to be stormy and unsettled weather with large amounts of rainfall at some stage in the month of June on this island. The last week of June appears to be the week with the most rainfall over the last 5 years. It appears that the last week of June is the week where one is most likely to see heavy rainfall. Based on the data from the last 5 years, there have been 11 days in the Month of June with an 5-year-average rainfall of over 1mm. To me this implies that there is a probability of slightly over 1/3 of experiencing a rainy day on a given day on this island in the month of June.

(iii) From seeing the data in this new table, I can conclude that the confidence interval that the analyst used in their summary was not appropriate and was misleading. They are trying to convince holiday-goers that they are unlikely to see much rain when they are on holidays on this island while in reality it appears that this is not the case.

2 Part Two

The task for Part Two is to write a short essay (up to 5 pages) highlighting "the statistical challenges and limitations associated with generative artificial intelligence."

Over the last few months there has been an explosion of generative artificial intelligence across the world with new technologies like ChatGPT and DALL-E. While generative artificial intelligence (GEN) tools has been around for decades – introduced with chatbots like ELIZA in the 1960s [3] and more recently Siri and Alexa – never before have AI tools been as accessible and as sophisticated as they are now. This recent upsurgeance of AI has left many people concerned about our future with AI. There are many questions being raised about generative AI, particularly regarding the ethics of AI, its limitations and concerns about human jobs being made redundant because of AI. In this essay I will attempt to answer some of these concerns by investigating some of the statistical challenges and limitations associated with generative artificial intelligence.

One major problem with generative AI tools, is the sheer volume of data they need to be trained on – leading to a large amount of bad data in the training sets. The GPT-3 model – which ChatGPT is based upon – is trained on over forty-five terabytes of text data and the model itself has about 175 billion parameters [4]. It becomes very difficult to screen data when working with huge volumes of data like this. OpenAI implement supervised and reinforcement learning techniques to train ChatGPT's model, this helps train the model to avoid giving 'bad' responses to prompts. However, ChatGPT still suffers from 'hallucination' errors where it states facts that do not exist or are not true [6].

Good questions to think of are, where does the training data come from? Is this data reliable? OpenAI have stated that ChatGPT uses a large amount of text sources from the internet including conversation data [7]. It is quite plausible that a lot of the data ChatGPT and similar natural language generating AIs are trained on are unreliable and messy – such as slang and 'memes' on the internet and 'fake news' spread online. Unwanted colloquial language and unusual syntax can be suppressed in the model from users' feedback such as by pressing the "Thumbs Down" button in ChatGPT.

An interesting application of natural language AI is to apply it to differentiating 'fake news' from 'real news'. It is not straightforward to differentiate between real news and fake news online – humans even tend to struggle to differentiate between the two [8]. Interestingly enough, this is a task that AI often outperforms humans at. It is easy to create a supervised statistical model with training data labelled into classes “real” and “fake”. There are a few potential issues with a method like this however. A supervised method like this only works if the training data is classified properly, models built upon bad data create bad models - ‘rubbish in, rubbish out’.

It is worth paying attention to how supervised learning is carried out for modern generative AI models. ChatGPT for example uses a reinforcement learning from human feedback approach to supervised learning. When a user provides ChatGPT with a prompt, a user can click a “Thumbs Up” or “Thumbs Up” button to give feedback to the AI on whether the output was ‘good’ or ‘bad’. The AI’s algorithm adapts based on this feedback and changes weights on its parameters based on this human feedback. The drawback of such a method is that it is slow and costly compared to unsupervised methods. Human feedback can also be inconsistent for certain scenarios – 2 different people may disagree on subjective matters – for example if a digital image-generating AI is provided a prompt to draw ‘good art’, it is likely people may disagree as to whether what the AI outputs is ‘good’ art or not.

Recent leaps in generative AI development have raised many ethical concerns from governments, companies, and individuals. A key focus for developers of these tools is to build them with ethic in mind from the ground up - a notable example of this is Microsoft’s commitment to creating ethical AI [9] . It is desirable that generative AI will be ethical and unbiased. This is especially important in the future as AI continues to be used more and more in critical decision-making scenarios. However, at present there are still issues with this in some generative AI tools. There are documented examples of left-leaning responses from ChatGPT when given prompts on socio-political topics [10] [11]. The left-leaning responses of ChatGPT are likely sourced from text data on websites – such as google or Wikipedia - that lean towards more liberal definitions than conservative ones. This means its likely that the algorithm tries to maintain its neutral stance but it may have a tendency to ‘trust’ some sources more than others which can lead to slightly biased responses if the source data is biased.

Another example of bias in generative AI are stereotypes and assumptions that these tools make. For example, the generative AI tool DALL-E takes natural language inputs as prompts and it uses AI to generate a realistic image to the user. There are reported instances of DALL-E creating images which tend to depict racial and gender stereotypes [12] [13]. Should the algorithm be tweaked to include more photos of women or people from different ethnic backgrounds in its output? One could possibly argue that by doing this it leads to an inflated representation of women and ethnic minorities in images than that is found in reality.

Perhaps another approach is to attempt to include more source images from under-represented regions in the world, to try and remove the western-world bias from AI generation somewhat. There is no clear answer to this question and perhaps it really depends on how precise the output needs to be for a given application or perhaps how 'natural' or realistic the output is for the human user. To some extent – it is a trade-off between how realistic we want the machine's output to be or how we want the machine's output to align with what we feel is 'right' or 'correct'. But of course, how accurate the output is also depends on whether the data we feed into the model is an accurate representation of what the machine is trying to model. The process generating the data that is fed to the generative AI may often be biased by human decision making, whether intentional or unintentional.

Training AI models on data-sets that exclude certain types of people can lead to major problems. An example of this is in medical textbooks where light skin tones tend to be overrepresented [14]. If medical textbooks are used as the training corpus for an AI model used to diagnose diseases, then there is a good chance it may misdiagnose people with darker skin tones because of this design bias. Similar issues have occurred in the past with machine vision models for self-driving cars being less-likely to detect dark-skinned pedestrians compared to white-skinned pedestrians [15].

Fears of being made redundant by AI are perhaps the biggest concern people have with regards to Artificial Intelligence at the moment. A common misconception with artificial intelligence is that AI systems are intelligent, while in reality they are "neither artificial not intelligent" [16]. Artificial Intelligence is still just a tool like a calculator or word processor – AI's real intelligence is from the humans who create the complex set of parameters that the generative AI's of today use to 'think'. The generative AI systems of today are effectively finding and matching patterns, and they are very good at doing it. Exam results from GPT-3 and GPT-4 models [6] show that while these models perform excellently regards writing and regurgitating information, there is a long way to go before most peoples' jobs are significantly at risk from AI systems, because AI lacks the power to actually think beyond the models humans have created for them.

Knowing AI is a tool, how much power and responsibility do we give it? We, as humans , must remember that at the end of the day Artificial Intelligence is a computer, using a complex algorithm to do lots of calculations in order to optimise some parameters to provide the user with the 'optimum' output. A computer can take many factors into account into these calculations but ultimately, most of these decisions are black and white, 1s and 0s. The computer cannot think beyond the parameters it is programmed to think in and it shows no emotion. So while AI is a fantastic idea to reduce workload and classify objects with complex non-linear features, like other calculators and tools, I believe AI should be used as an aid and humans should have the final say in the decision making process.

Some believe that if the corpus of information the AI is built upon is flawed and algorithms are not constantly checking the reliability of the data then these AIs could very much be "disinformation machines" [5]. Generative AI tools that are used in decision making contexts risk losing their effectiveness if the underlying training data is labelled incorrectly in supervised model approaches for example images of tissue samples for patients with a type of skin cancer and control images – where some are mislabelled, or some have missing values .

If AI designers are not careful when designing AI systems, they may forget the unintelligent nature of artificial intelligence. If AI designers forget to implement parameters and algorithms that do 'sanity-checks' on the data, we can end up with models with undesirable side-effects. If AI models do not check the integrity of the data that is being used, then it is very possible that fake data could contaminate the AI model. Then the model may fail to distinguish between fake data and real data in applications.

These sanity checks should factor in the ethical implications of the output – which ChatGPT does for example. Earlier versions of ChatGPT had a tendency to behave incorrectly when given sensitive or disallowed prompts [6]. There are still instances where you can 'trick' the AI into generating inappropriate responses to sensitive or disallowed prompts by rephrasing the prompt in a particular way. These 'tricks' are known as 'jailbreaking' the AI and it is a big limitation of modern AI - these flaws could potentially lead to serious illegal activities such as forgeries or identity theft if the AI's checks for legitimacy and appropriateness fail.

To conclude I will say that ultimately the most complicated AI models are only as good as the data the model is trained on. Bad data – whether mislabelled, omitted or fake data - is a massive problem for all statistical models that work with data and generative AI systems are no exception. Data integrity is also important to keep in mind, is the sample of data used is diverse enough to represent enough to minimise biases in the application of interest? Is the underlying data reliable – does it accurately represent what the AI is trying to model? And does the AI protect against fake data and malicious use of generative AI? These are some of the important statistical questions and limitations that I think people should keep in mind when creating generative AI models.

References

[1] S. Selvin, Two Issues concerning the Analysis of Grouped Data, *European Journal of Epidemiology*, 1987 , 3(3), 284–287, http://www.jstor.org/stable/3521127

[2] P. Voosen, How far out can we forecast the weather? Scientists have a new answer, *Science* , 14th February 2019 https://www.science.org/content/article/how-far-out-can-we-forecast-weather-scientists-have-new-answer

[3] G. Lawton, What is generative AI? Everything you need to know, Tech Target https://www.techtarget.com/searchenterpriseai/definition/generative-AI#:~:text=Generative%20AI%20was%20introduced%20in,and%20audio%20of%20real%20people

[4] K. Cooper, OpenAI GPT-3: Everything You Need to Know, Springboard, November 1st 2021 , https://www.springboard.com/blog/data-science/machine-learning-gpt-3-open-ai/#:~:text=GPT%2D3%20is%20a%20very%20large%20language%20model%20(the%20largest,text%20data%20from%20different%20datasets

[5] J. Bersin, Understanding Chat-GPT, And Why It's Even Bigger Than You Think, January 22nd 2023 , https://joshbersin.com/2023/01/understanding-chat-gpt-and-why-its-even-bigger-than-you-think/

[6] GPT-4 research, OpenAI https://openai.com/research/gpt-4

[7] N. Staudacher, What is ChatGPT?, OpenAI https://help.openai.com/en/articles/6783457-what-is-chatgpt

[8] I. Xenogiannis, Using Machine Learning to Distinguish Between What's Real and What Is Not, Towards Data Science, October 14th 2022, https://towardsdatascience.com/using-machine-learning-to-distinct-whats-real-and-what-is-not-9c1c74f73c8c

[9] Responsible AI, Microsoft, https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1:primaryr6

[10] R. Lownie, ChatGPT is not politically neutral, The Post, 7th December 2022, https://unherd.com/thepost/chatgpt-is-not-politically-neutral/

[11] J. Kemper, ChatGPT has left-wing bias – study, The Decoder, January 22nd 2023, https://the-decoder.com/chatgpt-is-politically-left-wing-study/

[12] K. Johnson, DALL-E 2 Creates Incredible Images and Biased Ones You Don't See, Wired, May 5th 2022, https://www.wired.com/story/dall-e-2-ai-text-image-bias-social-media/

[13] N. Tiku, AI can now create any image in seconds, bringing wonder and danger", The Washington Post, September 28th 2022, https://www.washingtonpost.com/technology/interactive/2022/artificial-intelligence-images-dall-e/

[14] P. Louie, R. Wilkes, Representations of race and skin tone in medical textbook imagery, *Social Science Medicine*, Volume 202, 2018, Pages 38-42, ISSN 0277-9536, https://doi.org/10.1016/j.socscimed.2018.02.023. https://www.sciencedirect.com/science/article/abs/pii/S0277953618300790

[15] B. Wilson, J. Hoffman, J. Morgenstern, Predictive Inequity in Object Detection, 2019, https://arxiv.org/pdf/1902.11097.pdf

[16] E. Morozov, The problem with artificial intelligence? It's neither artificial nor intelligent, The Guardian, March 30th 2023, https://www.theguardian.com/commentisfree/2023/mar/30/artificial-intelligence-chatgpt-human-mind