

Master's Programme in Engineering Physics and Mathematics

On the quality of mathematical writing produced by ChatGPT and Gemini

Meri Aho

© 2024

This work is licensed under a [Creative Commons](https://creativecommons.org/licenses/by-nc-sa/4.0/) “Attribution-NonCommercial-ShareAlike 4.0 International” license.



Author Meri Aho

Title On the quality of mathematical writing produced by ChatGPT and Gemini

Degree programme Engineering Physics and Mathematics

Major Mathematics

Supervisor Assoc. Prof. Pauliina Ilmonen

Advisor Assoc. Prof. Pauliina Ilmonen

Date **Number of pages** 48+73 **Language** English

Abstract

Large Language Models (LLMs) have gained popularity in recent years, particularly after the release of ChatGPT, which is an LLM created by OpenAI. LLMs generate human-like text according to a given prompt. LLMs use Natural Language Processing (NLP) to generate the text.

In this thesis, we look into two LLMs, ChatGPT and Google Gemini, and investigate the quality of mathematical text generated by both of the models. The research questions of this thesis are the following three questions: 1) Can ChatGPT and Gemini write good mathematical text? 2) Can they write sensible references? and 3) Which LLM is better for these two tasks? This is done by asking both ChatGPT and Gemini to write abstracts on 25 statistical terms and give references to the texts. A student and a professor from Aalto University then assess the quality of writing in terms of mathematical accuracy. The results are further analysed by performing sign tests. Additionally, the LLMs are asked to give sources in the texts, which are then checked for accuracy. The two LLMs are then given a normalized score based on the amount of mistakes made in the references. A sign test and a t-test for paired observations are then conducted for assessing the difference of these scores.

We discover that both ChatGPT and Gemini generate quite poor quality of mathematical text. Although the text is well written, it lacks accuracy and mathematical correctness and also sometimes chooses quite goofy words, considering the type of text being written. The references are often riddled with mistakes or they flat out do not exist. Although both ChatGPT and Gemini were found out to be bad in both producing mathematical text and producing references, ChatGPT was found out to be better than Gemini in producing references, but Gemini was found out to be slightly better in producing mathematical text according to both the student and the professor.

Keywords Large Language Models , ChatGPT , Google Gemini , Statistical testing

Tekijä Meri Aho

Työn nimi ChatGPT:n ja Geminin tuottaman matemaattisen tekstin laadusta

Koulutusohjelma Teknillinen Fysiikka ja Matematiikka

Pääaine Matematiikka

Työn valvoja Prof. Pauliina Ilmonen

Työn ohjaaja Prof. Pauliina Ilmonen

Päivämäärä

Sivumäärä 48+73

Kieli englanti

Tiivistelmä

Suuret kielimallit ovat tulleet suosituiksi viime vuosina, erityisesti ChatGPT:n julkaisemisen jälkeen. Suuret kielimallit tuottavat tekstiä annetun kehotteen mukaan, ja ne käyttävät luonnollisen kielen prosessointia tekstin tuottamiseen.

Tässä diplomityössä tarkastellaan kahta suuriin kielimalleihin perustuvaa tekstigeneraattoria, ChatGPT:tä ja Google Geminia, sekä niiden tuottaman matemaattisen tekstin laatua. Tutkimuskysymykset ovat seuraavat: 1) Osaavatko ChatGPT ja Gemini kirjoittaa hyvää matemaattista tekstiä? 2) Osaavatko ne kirjoittaa järkeviä viitteitä? ja 3) Kumpi menetelmä on parempi näissä kahdessa tehtävässä? Tutkimus tehdään pyytämällä kielimalleja kirjoittamaan tiivistelmät 25:stä tilastollisesta termistä ja antamaan viitteet teksteihin. Tämän jälkeen yksi Aalto-yliopiston oppilas sekä yksi Aalto-yliopiston professori erikseen valitsevat mielestään paremman tekstin, ja näille arvioinneille tehdään merkkitestit. Lisäksi teksteille annetaan normalisoitu pistemäärä niiden tekemien viitevirheiden perusteella. Tälle pistemäärädatalle tehdään merkkitesti, sekä parillisten otosten t-testi.

Tekstejä lukiessa huomaa, että sekä ChatGPT että Gemini tuottavat melko huonolaatuista matemaattista tekstiä. Vaikka teksti on hyvin kirjoitettu, se on matemaattisesti virheellistä ja valitsee joskus varsin hassuja sanoja kyseiseen tekstityyppiin. Viitteissä on usein virheitä tai ne eivät yksinkertaisesti ole olemassa. Vaikka sekä ChatGPT:n että Geminin havaittiin olevan huonoja sekä matemaattisen tekstin että viitteiden tuottamisessa, sekä opiskelijan että professorin mukaan ChatGPT oli parempi viitteiden tuottamisessa, mutta Gemini oli hieman parempi matemaattisen tekstin sisällön tuottamisessa.

Avainsanat Suuret kielimallit , ChatGPT , Google Gemini , Tilastollinen testaus

Contents

Abstract	3
Abstract (in Finnish)	4
Contents	5
Abbreviations	6
1 Introduction	7
2 Background	8
2.1 Other studies	8
3 Methods	10
3.1 Sign test	11
3.2 t-test for paired observations	13
3.3 Comparison of sign test and t-test for paired observations	14
3.4 Possible issues	15
4 Results	16
4.1 Examples	24
4.1.1 Sign test	24
4.1.2 t-test	29
4.1.3 Derivative	33
4.1.4 Gaussian processes	37
4.2 Other observations	42
5 Discussion	44
References	45
A Appendix	49

Abbreviations

LLM	Large Language Model
GPT	Generative Pre-trained Transformer
NLP	Natural Language Processing
AI	Artificial Intelligence

1 Introduction

Artificial Intelligence (AI) has had major developments in the past years, and a major breakthrough has been seen in the creation of chatbot like text generators, which utilize Large Language Models (LLMs). In this thesis we analyse text generated by two LLMs, ChatGPT and Gemini, to assess the quality of the content of the produced text.

This thesis is intended to answer the following three research questions: 1) Can ChatGPT and Gemini write good mathematical text? 2) Can they write sensible references? and 3) Which LLM is better for these two tasks?

Section 2 gives an introduction into LLMs, and provides information on other studies similar to the study conducted in this thesis. It serves as a brief literature review on LLMs.

Section 3 describes the methods used for the comparison and analysis of texts on mathematical topics generated by the two LLMs. These methods will aid with answering the research questions. Particularly, sign test and t-test for paired observations are utilized. They are introduced and discussed in Subsections 3.1 and 3.2, respectively. The methods are also compared to one another in Subsection 3.3. Finally, some possible concerns about the study are addressed in Subsection 3.4.

The results of the sign tests and t-test for paired observations are displayed and discussed in Section 4. In addition to the results of statistical testing, insight is given into what kind of texts the LLMs have produced. Moreover, general observations and characteristics of the generated texts are discussed. A few concrete examples of the generated texts are given in Subsection 4.1. This gives the reader a brief overview of how the assessments are made and what qualities are deemed important or necessary for good mathematical writing produced by an LLM.

Finally, concluding remarks, discussion, and ideas for future research are given in Section 5.

2 Background

LLMs are a type of artificial intelligence, which are trained with a vast amount of text-based data. The LLM can then produce its own text, answer questions, categorize text, and perform other tasks related to language [1]. The process of generating responses to such tasks is called Natural Language Processing (NLP) [2]. Other common NLP problems include generative text (such as a search engine suggesting words), spam detection and translation.

ChatGPT, which stands for Chat Generative Pre-trained Transformer, is an LLM powered by the GPT-3.5 architecture. It was first introduced to the public in late November of 2022 by OpenAI. [3] Since its release, it has quickly become an immensely popular tool for solving NLP tasks [4]. There is also premium version of ChatGPT available for a monthly fee. ChatGPT Plus uses the GPT-4 architecture, which is a newer and upgraded version to the GPT-3.5 architecture. [5]

Soon after the release of ChatGPT, Google introduced its own chatbot, which is powered by Language Model for Dialogue Applications, or LaMDA [6]. The name of the chatbot was originally Bard, but has recently (since the 8th of February 2024) rebranded to Google Gemini [7]. For the sake of this thesis, we will refer to the LLM as Gemini, even though some of the content was made before the rebrand.

Research suggests that LLMs do not understand the logic behind math equations presented to it. The LLM mimics how human would respond without understanding the logic behind the correct answer. [8], [9] For example, when generating an answer for a math related NLP problem, the LLM would often calculate incorrectly, but the reasoning is very human-like [8]. The LLM solely understands logic on a superficial level, but fails to synthesize and apply mathematics rules. For instance, LLMs are notoriously bad at applying arithmetic rules [10]–[12], particularly when the amount of digits in the input data is high [10], [11].

This study will focus more on the generated mathematical text and the LLMs' abilities to explain mathematical concepts, rather than its ability to complete mathematical operations.

2.1 Other studies

Several studies have explored the performance of LLMs by examining the accuracy of the generated responses. These responses have been compared to other LLMs and even human responses to the same questions. Additionally, the level of difficulty of the NLP problems presented to the LLMs have differed in these studies. For instance, several studies have been made comparing test performances of students to that of ChatGPT.

One study tested how well ChatGPT would perform in the British SAT exams. The results were then compared to the results of students. Notably, over 90% of students performed better than ChatGPT in two exams for the Test of Mathematics for University Admission (TMUA) in 2020 and 2021. One of the two TMUA exam in 2019 also had this result, whilst ChatGPT performed better in the other exam. Still, the result was sub par and 82% of students scored better than the LLM. However, for

the first section of BioMedical Admissions Test, ChatGPT scored closer to the average student, even surpassing the average student. Section one of this exam focused on thinking skills, which more closely resemble the NLP tasks ChatGPT has been trained on and thus would likely succeed better in. [13]

Another similar study was done for three high school level national English comprehension exams in the Netherlands. Students averaged 6.99 in a scale from 1-10, whereas GPT-3.5 had a worse score of 6.5. However, after reprompting the wrong answers, the score was bumped up to a 7.3. GPT-4 scored significantly higher with a score of 8.3. Additionally, GPT-4 was more consistent in giving a similar answer when the test was repeated. [14]

However, when raising the difficulty of exams to a university level, the LLMs tend to perform weaker than students studying the subject. For example, another study comparing student results to ChatGPT results was done in Poland for 3 medical final exams. The study also compared the LLMs performances in two languages, Polish and English. GPT-4 was found to give better responses than GPT-3.5 in all 3 examinations. Additionally, the responses were similar enough in both languages that there was zero difference in the points given for the exams in different languages. GPT-3.5 did not pass any of the exams in Polish whilst barely passing 2 of the 3 exams in English. However, both GPT-3.5 and GPT-4 still scored lower than the average medical student in Poland. [15]

Interestingly, when a similar study was done for biomedical course exams at the University of Florida, ChatGPT outperformed the students in 7 of the 9 courses [16], having an entirely different result to that of the previous example.

The arithmetic performance of various LLMs decreases as the number of digits in the input data increases [10], [11]. For multiplication, the success rate of accurately calculating the product plummets when just increasing the number of digits in the input values from 2 to 3, and when there are 4 digits, the success rate is practically zero. The same happens with addition and subtraction, however the rate of change is much smaller. [10]

In addition to NLPs with objectively correct and incorrect answers, there are also some with subjective answers. These include, for example humor or sarcasm detection from text. When a study was done comparing human answers and ChatGPT generated answers to such NLP problems, ChatGPT, for example, did not detect aggression as often as humans did in input data, but did detect unhealthy comments more often than humans. This study gave ChatGPT tasks categorized into, for instance spam detection, humor (sarcasm) recognition, offensiveness detection, emotion recognition and answering questions related to the input dataset. It was found that ChatGPTs performance was worse than State-of-The-Art models in all categories. [17]

3 Methods

In this chapter, we present the method applied for evaluating the responses generated by ChatGPT and Gemini and how the quality of the responses will be quantified. The purpose is to evaluate, whether the LLMs are a reliable source of information, particularly in the field of mathematics and statistics.

In order to compare the two LLMs, one Aalto University student and one Aalto University professor will separately evaluate the responses of ChatGPT and Gemini, and give a score based on the quality and accuracy of the generated texts.

ChatGPT and Gemini are given the same prompts in the form

'Write a one page abstract on [topic] with references and in-text citations'.

The topics are related to statistics and vary in level of difficulty. There will be a total of 25 topics, 19 of which are simple statistical terms and the last 6 being more advanced topics. The quality of the responses are then evaluated and compared by assigning the better response 1 point and 0 to the other.

The better candidate for each topic is determined by the mathematical accuracy and thoroughness of the writing alongside the quality of presentation. In cases where the responses are similar in quality, the better option was chosen based on subjective opinion. This data is then analysed by performing a sign tests.

Let $a = [a_1, a_2, \dots, a_{25}]$ be a $(0, 1)$ -vector representing the points given to ChatGPT and $b = [b_1, b_2, \dots, b_{25}]$ another $(0, 1)$ -vector representing the points given to Gemini. Each index i represents a topic. The values of a_i are assumed to be independent to each other, i.e. the outcome of a_1 does not affect a_2 etc. Here,

$$a_i = 1 - b_i \quad \forall i \in \{1, \dots, 25\}.$$

Terms a_i and b_i are evaluated based solely on the content written for topic i and the accuracy of the references are not considered in the evaluation, unless they significantly interfere with the quality of presentation. The references are however evaluated separately.

Finally, the responses are then given to a mathematics professor at Aalto University and they will evaluate each response and choose the better response of the pair, similarly to as before. However, the professor will not be told which LLM generated which text. Thus the evaluation will remain anonymous and unbiased.

In addition to the quality of the generated responses, the references will be manually checked and given a separate score. The purpose of this is to validate the existence of the sources, and to further understand whether the LLM continues generating random text from the sources, or if it actually knows how to create accurate sources.

For each topic i , the LLMs are given a weighted score from 0 to 1 based on how many mistakes there were in the references in relation to the total amount of references given. Let $a' = [a'_1, \dots, a'_{25}]$ and $b' = [b'_1, \dots, b'_{25}]$ be vectors representing the mistakes made in the references by ChatGPT and Gemini, respectively. We define a'_i as the

following

$$a'_i = \frac{\sum_{j=1}^{n_{A_i}} r_{jA_i}}{n_{A_i}},$$

where n_{A_i} is the number of references given by ChatGPT on the topic i and $r_{jA_i} = 0$, if there are no mistakes in reference j , and $r_{jA_i} = 1$ if there is a mistake in reference j , unless it is a small mistake. Then $r_{jA_i} = 1/2$. The different types of mistakes will be defined in more detail in Chapter 4. The severity of the mistakes will be defined as they occur, so that the chart can then also serve as documentation of the different types of mistakes made by the LLMs. Meaning that although a type of mistake might be thought of, if it does not occur in the generated texts, it will not be listed in the chart.

Similarly, the mistakes in references made by Gemini on the topic i is defined as the follows:

$$b'_i = \frac{\sum_{j=1}^{n_{B_i}} r_{jB_i}}{n_{B_i}}.$$

In-text citations are not reviewed further, as some of the references might not exist or are not easily available.

Mistakes in references are evaluated by performing a sign test and a t-test for paired observations.

3.1 Sign test

The contents of this subsection are based on Hogg, McKean, and Craig [18].

Let $X = \{X_1, \dots, X_n\}$ be a random sample, where the elements are distributed around the median θ of the distribution, such that

$$X_i = \theta + \varepsilon_i \quad \forall i \in \{1, \dots, n\},$$

where $\varepsilon_1, \dots, \varepsilon_n$ are independent and identically distributed around 0, which is the median of their distribution. Then, for hypothesis testing on θ , the one-sided hypothesis can be given as follows:

$$H_0 : \theta \leq \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0,$$

where H_0 is the null hypothesis and H_1 is the alternative hypothesis. Similarly, the two-sided hypothesis can be expressed as

$$H_0 : \theta_0 = \theta \quad \text{vs.} \quad H_1 : \theta_0 \neq \theta. \tag{1}$$

The sign test, denoted s or $s(\theta_0)$, calculates the number of values that are larger than θ_0 . A mathematical expression for this is

$$s = s(\theta_0) = \#\{X_i > \theta_0\}. \tag{2}$$

An alternative expression for this can be given as

$$S = S(\theta_0) = \sum_{i=1}^n I(X_i > \theta_0),$$

where $I(X_i > \theta_0) = 1$, when $X_i > \theta_0$ and otherwise $I(X_i > \theta_0) = 0$.

The sign test merely considers the sign of the difference in measurements, and not, for example, the size of the difference. It is a very simple distribution-free statistical tool. It makes no assumptions of the underlying distribution of X . It only assumes that the observations are independent and identically distributed.

For the one-sided hypothesis, the null hypothesis H_0 is rejected in favor of the alternative hypothesis H_1 if S exceeds some chosen c . The value c should be chosen to be the upper critical value of a significance level α test. The upper critical value c_α is defined as the following.

$$P_{H_0}(S \geq c_\alpha) = \alpha$$

For the two-sided hypothesis, H_0 is rejected in favor of H_1 , if $S \leq c$ or if $S \geq n - c$. Now for a level α test, c should be chosen as the following.

$$P_{H_0}(S \leq c) = \frac{\alpha}{2} \quad (3)$$

For determining the better LLM of ChatGPT and Gemini, the one-sided hypothesis is useful for stating that ChatGPT is better than Gemini or vice versa. However, the two sided hypothesis is more appropriate if there is no preference over ChatGPT or Gemini, stating that ChatGPT and Gemini are equally as good. For this thesis, we will use two-sided hypothesis as we have no preference for ChatGPT or Gemini.

A sign test is performed for comparing the quality of ChatGPT and Gemini. For the sign test

$$S = \sum_{i=1}^{25} a_i \quad \text{or} \quad S = \sum_{i=1}^{25} b_i.$$

Also, the correctness of the references are compared using the sign test. For that,

$$S = \sum_{i=1}^{25} I(a'_i \geq b'_i) \quad \text{or} \quad S = \sum_{i=1}^{25} I(a'_i \leq b'_i)$$

For both sign tests, assign $\theta_0 = 0.5$. Now, under the null hypothesis, about half of the observations are expected to be in favor of ChatGPT. Now c is solved for $\alpha = 0.05$ using the formula in (3):

$$P_{H_0}(S \leq c) = \frac{\alpha}{2} = 0.025.$$

From the CDF of the binomial distribution $b(25, 1/2)$, $c = 7$ is the value of $P_{H_0}(S \leq 7) \approx 0.021$ is closest to 0.025 of all integer values for c . Thus, H_0 is rejected in favor of H_1 if $S \leq 7$ or if $S \geq 18$.

3.2 t-test for paired observations

The contents of this subsection are based on the work of Kim, 2015 [19].

Let $X_A = \{X_{A1}, \dots, X_{An}\}$ and $X_B = \{X_{B1}, \dots, X_{Bn}\}$ be some random paired samples. Let $X_B - X_A = \{X_{B1} - X_{A1}, \dots, X_{Bn} - X_{An}\}$. Assume that the observations in $X_B - X_A$ are independent and identically distributed from some normal distribution. The sample mean of $X_B - X_A$ is denoted by \bar{X}_{B-A} . Let the corresponding population mean (expected value) be μ_{B-A} .

The point of the t-test for paired observation is to determine whether the sample mean of $X_B - X_A$ differs significantly from 0. Thus, define the null hypothesis and the alternative hypothesis as

$$H_0 : \mu_{B-A} = 0 \quad \text{vs.} \quad H_1 : \mu_{B-A} \neq 0. \quad (4)$$

One-sided hypothesis can be given as follows:

$$H_0 : \mu_{B-A} \leq 0 \quad \text{vs.} \quad H_1 : \mu_{B-A} > 0.$$

In what follows, we consider the two-sided alternative hypothesis.

The t-statistic for the paired data is as follows:

$$t = \frac{\bar{X}_{B-A}}{\sqrt{\frac{s_{B-A}^2}{n}}}, \quad (5)$$

where s_{B-A}^2 is the sample variance of the set $X_B - X_A$. A sample variance s^2 is defined as

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}, \quad n > 1. \quad (6)$$

Under H_0 , the t-test follows t-distribution, with $n-1$ degrees from freedom. Let t_v represent the probability density function of Student's t-distribution with $v = n-1$ being the degrees of freedom. Student's t-distribution is similar to the normal distribution in it's symmetry around it's mean and bell-curve shape, but has heavier tails. Notice from Figure 1, that the t-distribution approaches the standard normal distribution $N(0, 1)$ as the degrees of freedom v grow.

Even though t-test requires that the observations are normally distributed, it can be applied for non-normal data as well. However, this requires that the sample size is large and that the assumptions of Central Limit Theorem are fulfilled.

Now, under the distribution t-distribution t_{24} , for a significance level α , find the critical value c such that

$$P(x \geq c) = \frac{\alpha}{2}.$$

For the task at hand, the critical value is 2.391 for an $\alpha = 5\%$ level test. This means that the null hypothesis is rejected in favor of the alternative hypothesis, if the obtained t-value satisfies the inequality $|t| \geq 2.391$.

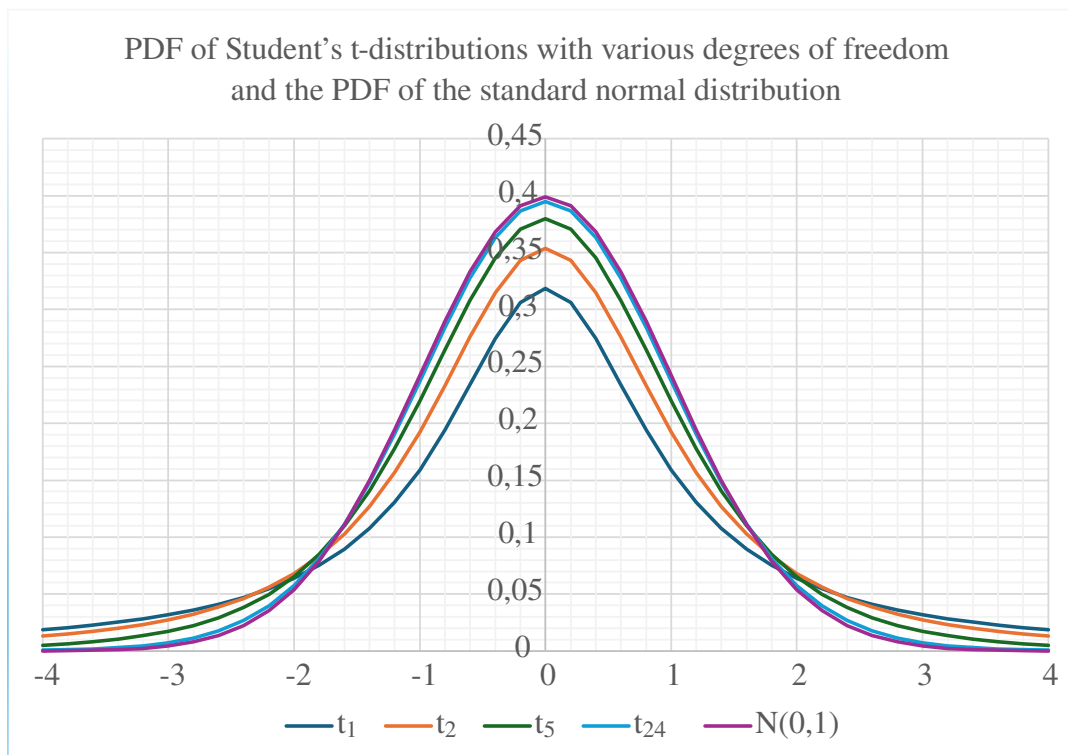


Figure 1: Probability density functions of various t-distributions and the standard normal distribution.

3.3 Comparison of sign test and t-test for paired observations

As the t-test and sign test are methods focused on different measures of location, the same issues that are associated with those locations are also associated with these tests. For instance, the mean of a dataset is heavily influenced by outliers, whereas the median is not as affected. Similarly, outliers in data do not significantly affect the result of sign tests, whereas even a single outlier, if large enough, has a large effect on the result of a t-test.

Since Student's t-test is a parametric test, it makes more assumptions about the data than the non-parametric sign test. The t-test assumes that the underlying distribution of the dataset is a normal distribution. However, if the sample size is large and the underlying distribution has finite expected value and finite variance, t-test can be applied. The reason for that, is that, based on the Central Limit Theorem, the sample mean of independent and identically distributed random variables with finite mean and variance is asymptotically normally distributed. When these assumptions are met, the t-test yields results that are easily comprehensible and generally make sense. The sign test makes less assumptions. Thus its power isn't as strong as that of the t-test, but it can be used more broadly.

3.4 Possible issues

AI models, such as LLMs learn from making mistakes [8]. However, we do not want to encourage the LLMs to learn from previous mistakes in this study. This is because we do not want the quality of a response depend on the quality of a previous response. In other words, we do not want the texts that are generated in later stages to be better than those generated at the start of the research. To ensure that the LLM does not learn from its previous mistakes, it will not be given any feedback or comments. In addition to this, new chats will be started with both ChatGPT and Gemini, instead of giving each prompt in the same chat. This will hopefully ensure that the LLM will not continue to provide a similar style of text moving forward, try to add to previous texts written, etc. Although ChatGPT does store the memory of previous conversations [20], and Google collects data from Gemini chats to improve Gemini [21], we hope that not commenting on the texts will minimize any information leakage from previous chats.

The best way to avoid information leakage from previous chats would be to create separate accounts for each prompt. However, this would take a tremendous amount of effort for very minuscule results, if any. Additionally, all Google accounts by one person are linked, and thus the work would be pointless, since Google could anyway see the other chats if it needed to.

There is only about an 8% chance of ChatGPT accurately correcting it's math related mistakes [8]. This means that even if the mistakes were commented on, there is a high chance of ChatGPT not correcting it's mistakes. However, LLMs correcting mistakes is not in the scope of this thesis, so we will consider the first text generated, even if there was a possibility of the LLM producing a better text.

The texts have been generated within a scope of a few weeks, and some updates have occurred during that time. For example, Gemini was known as Bard at the start of the study. These updates can have an effect on the quality of generated texts at the start of study versus at the end of the study. This means that they are not entirely independent, but we assume that the updates will not have a huge impact on the quality of texts.

4 Results

Table 1 depicts the data of vectors a and b for the different topics. There is also a column presenting the sign for each topic. The signs +/- are chosen such that + represents ChatGPT being the better candidate, or ChatGPT having value 1.

Table 1: Data for ChatGPT and Gemini, sign test

Prompt	ChatGPT	Gemini	Sign
sign test	0	1	-
p-value	0	1	-
Linear regression	0	1	-
Normal distribution	1	0	+
Correlation coefficient	1	0	+
Random variables	1	0	+
t-test	0	1	-
Standard deviation	1	0	+
Causation	0	1	-
Time series analysis	1	0	+
Variance	0	1	-
Median absolute deviation	0	1	-
Skewness	0	1	-
Kurtosis	1	0	+
Cauchy distribution	0	1	-
Nonlinear dependence	0	1	-
Derivative	0	1	-
Eigenvalue	0	1	-
Statistical inference	1	0	+
Functional data analysis	1	0	+
Trend series	0	1	-
Extreme value index	1	0	+
Gaussian processes	0	1	-
Statistical depth	0	1	-
Influence function	0	1	-

From Table 1 and its visualization in Figure 2, it is clear that Gemini is preferred over ChatGPT, with 9 of the 25 prompts giving a better result with ChatGPT and 16 with Gemini. That equates to 36% and 64%, respectively. This is further visualized in Figure 2. The sign test (2) gives a result of merely $S = 9$, which is higher than the upper critical value $c = 7$. Thus we cannot reject the null hypothesis, and the outcome is not statistically significant.

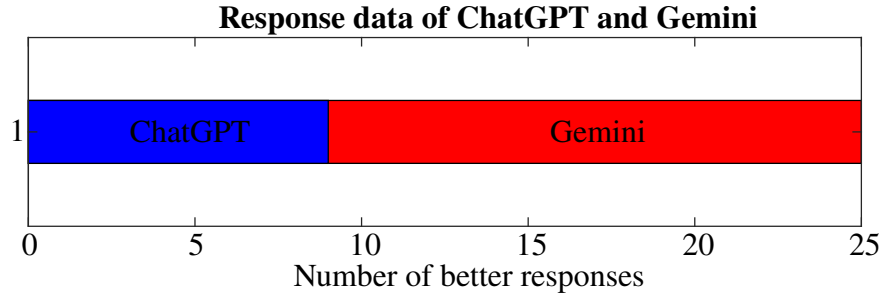


Figure 2: Visualization of the data in Table 1

The preferences of the professor are listed below in Tables 2.

Table 2: Professors preferences of ChatGPT and Gemini

Prompt	ChatGPT	Gemini	Sign
sign test	0	1	-
p-value	0	1	-
Linear regression	0	1	-
Normal distribution	1	0	+
Correlation coefficient	1	0	+
Random variables	0	1	-
t-test	0	1	-
Standard deviation	0	1	-
Causation	0	1	-
Time series analysis	0	1	-
Variance	1	0	+
Median absolute deviation	0	1	-
Skewness	1	0	+
Kurtosis	0	1	-
Cauchy distribution	0	1	-
Nonlinear dependence	0	1	-
Derivative	1/2	1/2	0
Eigenvalue	0	1	-
Statistical inference	0	1	-
Functional data analysis	1	0	+
Trend series	1	0	+
Extreme value index	1	0	+
Gaussian processes	1	0	+
Statistical depth	0	1	-
Influence function	1	0	+

This sign test gives $S = 9,5$. This sign test result is very similar to the result the student had gotten. However, for 10.5 of the 25 topics, the student and professor disagreed on the better option. Still, overall both the student and the professor agreed that Gemini is the clear better option. Figures 3 and 4 illustrates this further.

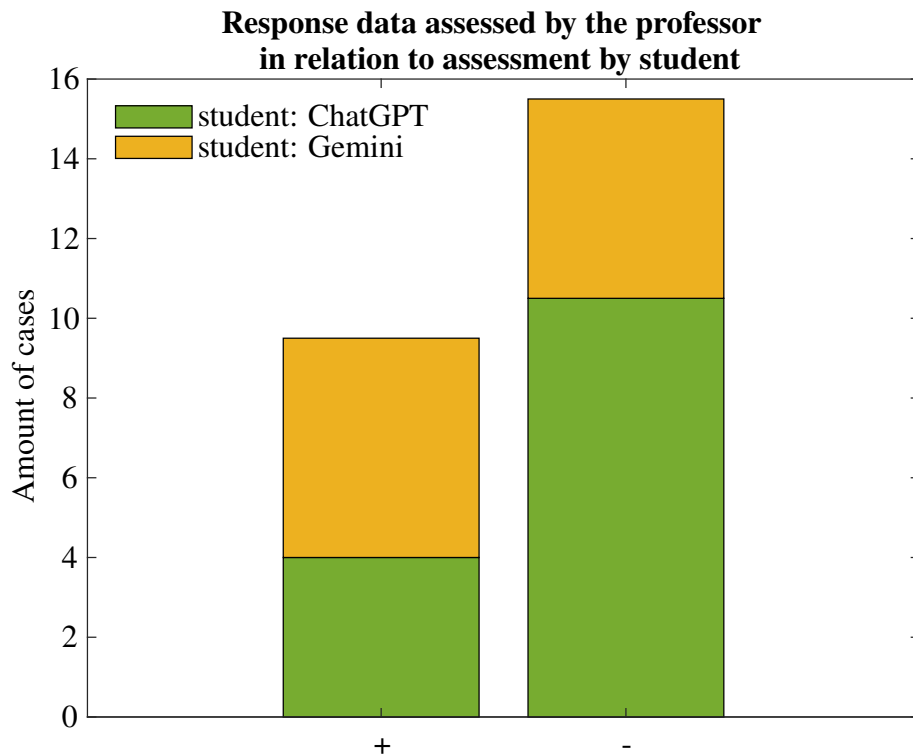


Figure 3: The responses assessed by the professor and it's relation to the student's assessment. The green section of the bar depicts the cases where the student agrees with the professor and the yellow when they disagree.

From these figures, we can see that when it came to the preference of ChatGPT (or choosing + as the sign), they didn't choose the same options often. In fact, for less than half of the cases where the professor preferred ChatGPT, the student agreed.

The reason for this could be due to very similar responses given by ChatGPT and Gemini, making it difficult to choose the better option. A common example, that illustrates the difficulty of choosing the better option, was deciding whether an "empty content" response was more desired to a response with more correct information but also several factual errors. In these cases, both responses could be argued to be the better response, depending on what the reader is expecting and subjective opinion.

When it came to Gemini, it was more common for the student and professor to agree about Gemini giving the better response. Still, for all the cases where at least one of the two assessors preferred Gemini, the other also preferred Gemini in only about half of the cases.

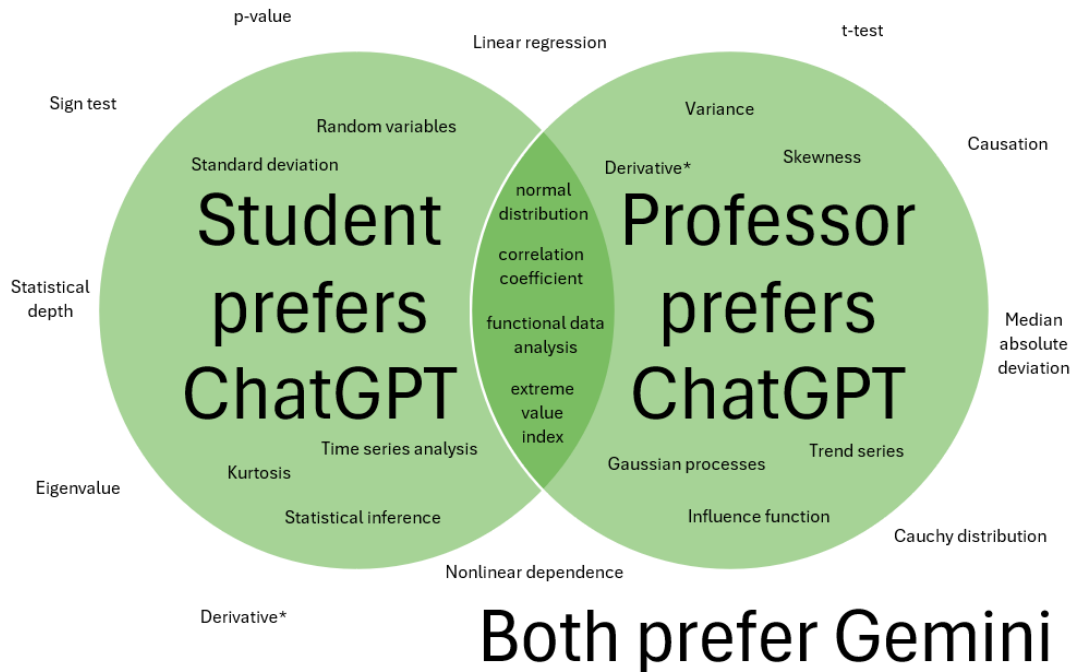


Figure 4: A Venn diagram of the response preferences. *For the topic Derivative, the professor has no preference, or rather, prefers both.

Table 3 lists the mistakes ChatGPT and Gemini made in references normalized by the total amount of references given for the topic, or the vectors of a' and b' . Mistakes are categorized into 1-point and 1/2-point mistakes and are listed in greater detail in Table 4.

Table 3: Mistakes in the references made by ChatGPT and Gemini for a given prompt

Prompt	ChatGPT (X_A)	Gemini (X_B)	$X_B - X_A$	Sign
sign test	$1/2 = 0.5$	$2/9 \approx 0.22$	-0.28	-
p-value	$0/3 = 0$	$4/7 \approx 0.57$	0.57	+
Linear regression	$1/6 \approx 0.17$	$1/3 \approx 0.33$	0.17	+
Normal distribution	$1/3 \approx 0.33$	$2/3 \approx 0.66$	0.33	+
Correlation coefficient	$0/5 = 0$	$3.5/7 = 0.5$	0.5	+
Random variables	$4.5/5 = 0.9$	$3.5/5 = 0.7$	-0.2	-
t-test	$1/5 = 0.2$	$0.5/2 = 0.25$	0.05	+
Standard deviation	$0/4 = 0$	$3/5 = 0.6$	0.6	+
Causation	$1/6 \approx 0.17$	$4/6 \approx 0.33$	0.5	+
Time series analysis	$0.5/4 \approx 0.13$	$4/10 = 0.4$	0.28	+
Variance	$1/5 = 0.2$	$3.5/8 \approx 0.44$	0.24	+
Median absolute deviation	$0/2 = 0$	$1/3 \approx 0.33$	0.33	+
Skewness	$1/3 \approx 0.33$	$3/3 = 1$	0.67	+
Kurtosis	$2/7 \approx 0.29$	$3/3 = 1$	0.71	+
Cauchy distribution	$1/3 \approx 0.33$	$5.5/6 \approx 0.92$	0.58	+
Nonlinear dependence	$2/5 = 0.4$	$4/4 = 1$	0.6	+
Derivative	$4/10 = 0.4$	$3/3 = 1$	0.6	+
Eigenvalue	$0.5/4 \approx 0.13$	$6.5/8 \approx 0.81$	0.69	+
Statistical inference	$0/5 = 0$	$3.5/8 \approx 0.44$	0.44	+
Functional data analysis	$2/3 \approx 0.66$	$3/5 = 0.6$	-0.07	-
Trend series	$0/3 = 0$	$5.5/9 \approx 0.61$	0.61	+
Extreme value index	$0.5/4 \approx 0.13$	$4.5/5 = 0.9$	0.78	+
Gaussian processes	$3/7 \approx 0.43$	$2.5/5 = 0.5$	0.07	+
Statistical depth	$5/5 = 1$	$5/6 \approx 0.83$	-0.17	-
Influence function	$1.5/5 = 0.3$	$3/4 = 0.75$	0.45	+

Out of the 25 sets of sources, ChatGPT is clearly preferred to Gemini in making references. ChatGPT is preferred to Gemini 21 times out of the 25, giving the sign test a result of $S = 21$. This result is statistically significant, as it is larger than the limit 18. We can therefore reject the null hypothesis given in (1), which states that the population medians of the two distributions are equal.

The sample mean values of vectors a' and b' are $\bar{X}_A = 0.27957$, $\bar{X}_B = 0.64169$, respectively. The standard deviations are about $\sigma_A \approx 0,27$ and $\sigma_B \approx 0,25$ for a' and b' , respectively. The standard deviations of the datasets are quite similar in size, however the sample means are contrasting, suggesting that the means are not equal. To accept or deny this theory, the t-test will be performed.

The sample mean, calculated for the t-test $\bar{X}_{B-A} = 0.36212$. The sample variance s_{B-A}^2 , calculated with the formula given in (6), is

$$s_{B-A}^2 = \frac{(0.36212)^2}{24} \approx 0.095953.$$

The value of the t-statistic, given in Equation (5), is now

$$t = \frac{0.36212}{\sqrt{\frac{0.095953}{25}}} \approx 5.845.$$

This t-value is high, and much higher than the upper limit 2.391. The t-statistic $t = 5.845$ gives a p-value of about $5.0 \cdot 10^{-6}$. This result is clearly statistically significant, and we can confidently reject the null hypothesis (4), that states that the population means of the datasets are equal.

These results are further illustrated in Figures 5 and 6, where ChatGPT and Gemini form clusters. In the charts, the data points of ChatGPT have a tendency of placing lower than Gemini, although outliers exist for both datasets.

Since both the sign test and the t-test gave statistically significant results, it is likely that ChatGPT is better at creating references than Gemini. This doesn't, however, mean that ChatGPT is good at making them. These results only show how incredibly bad Gemini is at it in comparison to ChatGPT. In fact, ChatGPT clearly isn't that skilled at making references, either, with the sample mean of ChatGPT suggesting that on average there is a mistake in every fourth reference. There are, however, quite a few cases, where ChatGPT doesn't make any mistakes in the references. The total amount of such cases, in this dataset, is 6. The same cannot be said for Gemini, as there are no cases where Gemini makes 0 mistakes.

Table 4: Categorized mistakes

Points	Mistake
1	No sources were given No references section References given in a sentence (not as a list) Writing doesn't exist Writing exists, but with different authors Authors have made a text, but it is not the one being referred to An extra author mentioned Author's name spelled wrong (including wrong initials) Gives url to a different article Article not found in any volume or issue of given journal Multiple mistakes (regardless of point value)
1/2	Incorrect year published Incorrect page numbers (of journal) Incorrect edition of book or incorrect volume or issue of journal Incorrect publisher Has the text "<invalid link removed>" in place of a url An incorrect author, who however has worked on a previous or later edition Subtitle missing No volume given for journal Gives link to a subsection of referenced site (but does not refer to the subsection itself)
0	Missing edition of book Inconsistent in referencing style Title in lower case Odd placement of References paragraph (not at the end) Bad quality of source (such as wikipedia)

Table 4 lists the mistakes found in the generated sources and categorizes them into 1 and 1/2 point mistakes, based on the severity of the mistake. There is also a 0 point-category, mainly for the appearance of the sources section, or for cases where there is minor information missing, that is not necessarily needed officially. These are listed, as they could affect the readability of the entirety of the text, but are not considered mistakes in this context. The quality of the appearance of the sources is thus not considered in the evaluation, but rather the content and the actual existence of the source.

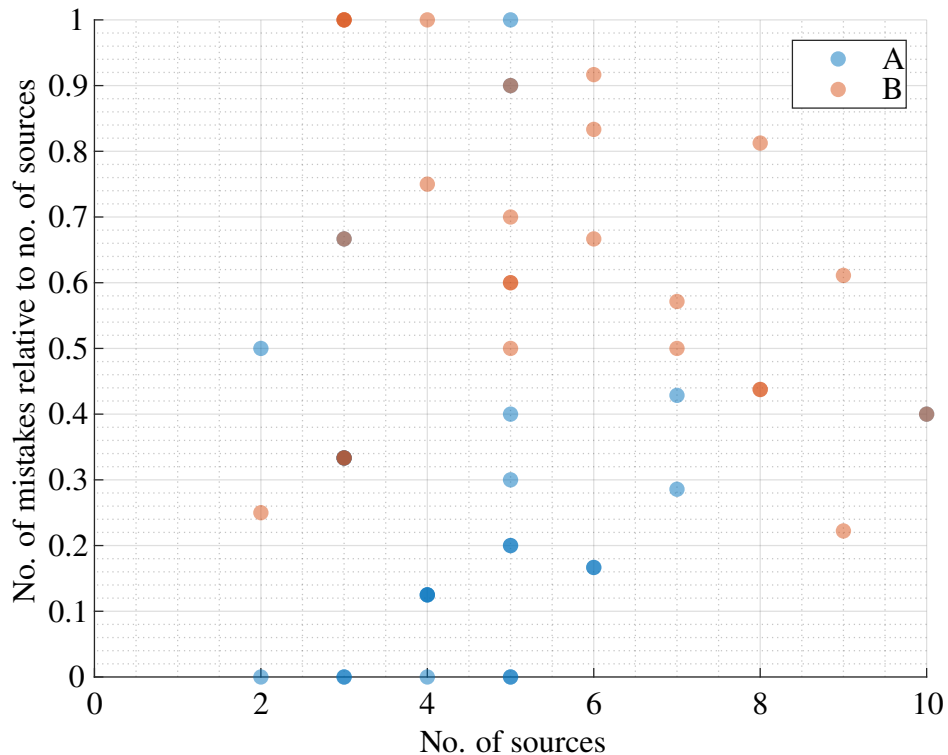


Figure 5: The total mistakes made by ChatGPT and Gemini. Here A represents ChatGPT and B represents Gemini.

Figures 5 and 6 depict the absolute and relative mistakes made in the generation of sources by ChatGPT and Gemini. Evident from the diagrams, all the cases of no mistakes are by ChatGPT. Gemini is more likely to have 50% or more mistakes, and ChatGPT is more likely to have 50% or less mistakes.

Although both ChatGPT and Gemini has cases, where each reference had a mistake, only ChatGPT also had cases where no mistakes were made, whereas Gemini did not.

Interestingly, from Figure 6 it would seem that the more sources ChatGPT generates, the more mistakes it makes proportionally to the number of sources, while the reverse is true for Gemini. However, there are still far too many outliers for both ChatGPT and Gemini. Additionally, the sample size is very small, so each outlier has an impact. To support or reject this claim, the sample size should be much larger. It is impossible to say if the effect could be minimal or nonexistent given a bigger sample size or if the same trend would appear given a larger sample size.

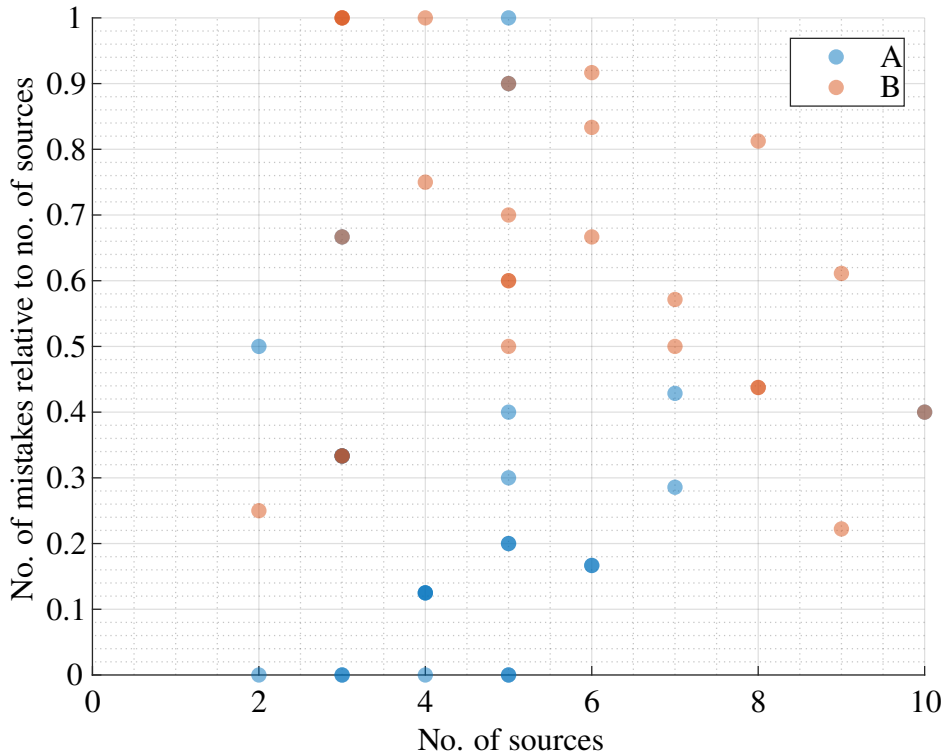


Figure 6: Visual representation of Table 3. Here A represents ChatGPT and B represents Gemini.

4.1 Examples

Next we will highlight a few of the topics and analyse the responses in greater detail. The topics chosen for this analysis are sign test, t-test, derivative and Gaussian processes. sign test and t-test are chosen, because Sections 3.1 and 3.2 describe the concepts and give a rough idea of what to be expected from the generated texts. Note, however, that Section 3.2 is of a specific variation of the t-test, and does not define or explain t-tests in general. The generated text is expected to have other forms of t-tests too. Derivatives were chosen as the response of ChatGPT had a special case of misunderstanding. Finally, Gaussian processes was chosen to illustrate the differences in responses when the topic is more difficult. Note also, that abstracts typically don't include sources [22], but we have specifically asked for them.

All of the generated texts are available in Appendix A.

4.1.1 Sign test

The response given by ChatGPT is the following.

Example 1: Text generated by ChatGPT on sign test

Abstract

The sign test, a non-parametric statistical method, is a valuable tool for analyzing data when the assumptions of parametric tests are not met or when dealing with small sample sizes. This test is particularly useful when dealing with ordinal or skewed data, or when the underlying distribution is unknown or non-normal. The sign test evaluates whether the median of a dataset differs significantly from a hypothesized value, making it robust against outliers and resistant to the effects of non-normality.

This paper provides an overview of the sign test, including its assumptions, calculation, and interpretation. Additionally, it discusses its application in various fields such as medicine, environmental science, and social sciences where parametric assumptions may be violated or where small sample sizes are common. Through a review of relevant literature and case studies, this paper demonstrates the versatility and utility of the sign test in practical research settings.

The sign test has been extensively studied and applied in research literature. In a study by Wilcox (2016), the sign test was utilized to analyze the efficacy of a new treatment for chronic pain, demonstrating its applicability in medical research. Furthermore, in a study by Ertefaie et al. (2018), the sign test was employed to assess the impact of air pollution on respiratory health, highlighting its use in environmental science.

In conclusion, the sign test is a valuable statistical method for analyzing data in situations where parametric assumptions are not met or when dealing with small sample sizes. Its robustness, simplicity, and versatility make it a valuable tool for researchers across various disciplines.

References

1. Ertefaie, A., Small, D. S., & Flory, J. H. (2018). sign tests for the matched-pair design: assessing the effect of air pollution on health outcomes. *Journal of the American Statistical Association*, 113(523), 1415-1427.
2. Wilcox, R. R. (2016). *Introduction to robust estimation and hypothesis testing*. Academic Press.

The second paragraph is irrelevant and acts as just filler text. This is likely an unfortunate consequence of the prompt asking for an abstract and the nature of abstracts. The response also falsely states it will discuss applications in social sciences.

It does, however, give examples of applications in medicine and environmental science as promised, although the environmental science reference ends up being falsely cited.

This abstract also did not give a formula or tell how to calculate the sign test in further detail. By stating that "The sign test evaluates whether the median of a dataset differs significantly from a hypothesized value", it alludes to the formula but does not explicitly tell how the sign test is performed. This significantly affects the quality of the content.

The first article in the references does not exist. Interestingly, two of the three authors, A. Ertefaie and D. S. Small, have published in the mentioned volume and issue of the Journal of the American Statistical Association, but not with the provided title. The title instead is "Quantitative Evaluation of the Trade-Off of Strengthened Instruments and Sample Size in Observational Studies" [23]. This article also has a third author who is not J. H. Flory.

The second reference exists and ChatGPT made no mistakes in referencing it. However, some minor improvements could be done, namely stating that the book is the 4th edition. Based on these two references, $a' = 0.5$.

The response given by Gemini is the following.

Example 2: Text generated by Gemini on sign test

The sign test: A Powerful Non-Parametric Tool for Statistical Analysis

The sign test, a versatile non-parametric statistical method, stands out for its simplicity and broad applicability. Unlike parametric tests requiring specific assumptions about data distribution, the sign test relies solely on the directional difference between paired observations (Field, 2020). This characteristic makes it particularly valuable when:

- Data distribution is unknown or non-normal: Many real-world datasets deviate from normality, rendering parametric tests unsuitable (Siegel, 1956). The sign test bypasses this issue by focusing on the sign of change (+/-) rather than the magnitude of differences.
- Sample sizes are small: With limited data, the robustness of the sign test becomes advantageous (Cramèr, 1946). While parametric tests lose power with small samples, the sign test maintains its reliability.
- Matched pairs are present: The sign test thrives when observations exist as matched pairs, allowing comparisons within individuals before and after treatment, for example (McDonald, 2014).

Applications and Variations:

The sign test boasts diverse applications across various fields, including psychology, medical research, and economics. It can be used to:

continued on next page

continued from next page

- Compare two populations: Assess if a new medication truly impacts pain levels compared to a placebo (Sprent, 1999).
- Test for a median difference: Examine if customer satisfaction scores differ significantly after website revamp (Siegel & Castellan, 1988).
- Identify outliers: Detect observations deviating drastically from the paired trends within groups (Conover, 1999).

While the basic sign test analyzes paired data, variations cater to specific scenarios:

- McNemar's test: Compares two dependent nominal categories, suitable for binary outcomes like pass/fail (McNemar, 1947).
- Wilcoxon signed-rank test: Assigns ranks to differences within pairs, offering more information than just sign (+/-) (Wilcoxon, 1945).

Strengths and Limitations:

The sign test's main strengths lie in its simplicity, flexibility, and robustness. However, limitations exist:

- Loss of information: Focusing only on sign discards valuable information about the magnitude of differences.
- Lower power than parametric tests: When data meets parametric assumptions, parametric tests might be more powerful.

Conclusion:

The sign test, with its ease of use and minimal assumptions, stands as a valuable tool for researchers and analysts. Understanding its applications, variations, and limitations empowers informed decision-making in various statistical analyses.

Sources:

1. Conover, W. J. (1999). Practical nonparametric statistics (3rd ed.). Wiley.
2. Cramèr, H. (1946). Mathematical methods of statistics. Princeton University Press.
3. Field, A. (2020). Discovering statistics using IBM SPSS Statistics (5th ed.). Sage.
4. McDonald, J. H. (2014). Handbook of parametric and nonparametric statistical procedures. Routledge.

continued on next page

continued from next page

5. McNemar, Q. (1947). Psychological statistics. John Wiley & Sons.
6. Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. McGraw-Hill.
7. Siegel, S., & Castellan, N. J. (1988). Nonparametric statistics for the behavioral sciences (2nd ed.). McGraw-Hill.
8. Sprent, P. (1999). Applied nonparametric statistical methods (2nd ed.). Chapman and Hall.
9. Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics Bulletin, 1(6), 80-83.

Similarly to the abstract provided by ChatGPT, this abstract could have also gone into more detail about the calculation of sign test, as nowhere does it explicitly say how the sign test is calculated. A formula, for example, would have been a good addition. It does, however, allude to the calculation by stating that "the sign test relies solely on the directional difference between paired observations".

There is also a mistake in the Application and Variations section, where the text claims that sign test can be used to identify outliers. This is very counter-intuitive, as sign test is a method based on the median of the considered distribution, and therefore doesn't consider the magnitude of the differences between observations. In fact, it is extremely difficult to identify outliers with the sign test alone.

Gemini's response is closer to two pages of writing, even though the prompt only asked for one page. If references are excluded, the length of the response is just over one page. Thus there is one page of actual content, so it could be argued that this would not affect the quality of presentation. However, it could also be argued that the long length does affect the quality. We have chosen to ignore it, but decided that the quality would be negatively affected if a response exceeded two pages with sources.

The quality of the presentation is negatively affected by the bullet points, as they are to be avoided in abstracts. It does represent the content in a easily readable format, however that was not the prompt. However, it should be noted that presentation is secondary to the content, and does not always affect the final score, depending on what the reader values. In this case, we always value the content more than the presentation.

References 1, 2, 3, 6, 7 and 9 exist and there are no problems other than the minor formatting issues that ChatGPT also had. Note, however, that 6 and 7 are references to the same book, only 7 has a newer edition. The fourth reference has the incorrect author, so it is considered a mistake worth 1 point. References 5 and 8 both have the wrong publishing year, and are both worth 0.5 points. Thus b' has a weighted score of $2/9 \approx 0.22$.

In comparison to this response, the text generated by ChatGPT lacks detail. Gemini gives a more in-depth look on sign test with more examples and applications, however also making a mistake. Both skipped relevant aspects, such as the formula. Any

explanation, whether in the form of a formula or just text, on how to calculate the sign test could have benefited both ChatGPT and Gemini. That would be expected from mathematical writing. ChatGPT stayed within the one-page limit, unlike Gemini. Thus it is not clear which response is the better response, but Gemini has been chosen. However, it would also make sense if ChatGPT would be chosen. This solely depends on what the reader is looking for and eventually the result is up to subjective opinion. ChatGPT has better presentation, considering that it stayed within the one-page limit and the style resembles an abstract more, but the content of Gemini is slightly better.

Both the student and the professor chose Gemini as the better option.

4.1.2 t-test

Section 3.2 isn't a perfect example of what is expected, as it covers only t-tests for paired observations, but it gives a rough idea of what to expect.

Example 3: Text generated by ChatGPT on t-test

Abstract:

The t-test is a fundamental statistical method used to determine if there is a significant difference between the means of two groups. Originating from the work of William Sealy Gosset in 1908, the t-test has since become a cornerstone in various fields including psychology, medicine, economics, and more. This abstract provides an overview of the t-test, its applications, assumptions, and variations.

The t-test is commonly employed when dealing with small sample sizes (typically less than 30) and assumes that the data is normally distributed. It is particularly useful in experimental research settings where researchers aim to compare the means of two groups, such as comparing the effectiveness of two different treatments or interventions. The t-test calculates a t-value, which represents the difference between the means of the two groups relative to the variance within the groups.

There are several variations of the t-test, each suited for different scenarios. The independent samples t-test is used when comparing the means of two independent groups, while the paired samples t-test is used when the same group is measured at two different points in time or under two different conditions. Additionally, the one-sample t-test is utilized when comparing the mean of a sample to a known value or population mean.

Assumptions of the t-test include the independence of observations, normality of data distribution, and homogeneity of variances between groups. Violations of these assumptions can affect the validity of the t-test results.

continued on next page

continued from next page

In conclusion, the t-test is a versatile statistical tool for comparing means between two groups in various research contexts. Its simplicity, applicability to small sample sizes, and ability to provide meaningful insights make it a widely used method in scientific inquiry.

Sources:

1. Gosset, W. S. (1908). The probable error of a mean. *Biometrika*, 6(1), 1-25.
2. Field, A. (2009). *Discovering statistics using SPSS*. Sage.
3. Howell, D. C. (2012). *Statistical methods for psychology*. Cengage Learning.
4. Kim, H. Y. (2015). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, 40(4), 328-331.
5. Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: the Bonferroni method. *BMJ: British Medical Journal*, 310(6973), 170.

In the text, it is stated that t-test required that the data is normally distributed. However, the text does not mention that normality assumption can be relaxed in case the sample size is large and the underlying distribution fulfills the assumptions of the central limit theorem.

Secondly, the sentence "The t-test calculates a t-value, which represents the difference between the means of the two groups relative to the variance within the groups" is completely false. The formula of the t-value depends on if the test is one or two-sampled. Regardless, the written formula is still wrong in any case. The given formula is missing the sample size(s) of the data, and fails to specify the use of the sample metrics, which aren't the same as those of the underlying distribution. Finally, the (sample) variance should be under a square root.

The third paragraph mentions the different variations on the t-test, which is good, but despite this, the calculation earlier doesn't specify which variation it is attempting to be. Furthermore, the explanation for each type of t-test is not great.

The assumptions paragraph is also not without flaws. Again, normality assumption is mentioned but possibilities for relaxing this assumption when the dataset is large is not discussed. The assumption about the variance is somewhat imprecise. The independence of the observations is however a correct assumption [19].

The given references are mostly correct, only the fourth reference included two mistakes. The first one being that the article was released in 2013, not 2015, and the volume, issue and page numbers of the journal are not correct. Instead of 40, 4, and 328-331, the correct volume, issue and page numbers are 38, 1 and 52-54, respectively

[24]. However, the same author has realised an article in the cited issue of the journal in the cited pages. The year this article was published is also 2015, however the article has a different title, being "Statistical notes for clinical researchers: effect size" [25]. Interestingly, the two articles seem to be a part of a series of articles by the author, which could have made the LLM think it is okay to apply the two articles in one reference. It is unclear which of the two articles the LLM was actually attempting to refer to, but regardless, these mistakes added up to one mistake and a total of $1/5 = 0.2$ points.

Example 4: Text generated by Gemini on t-test

The T-Test: A Workhorse of Statistical Inference

The t-test, also known as Student's t-test, stands as one of the most fundamental and widely used statistical methods in diverse fields, from psychology and medicine to economics and social sciences (Field, 2009). Its versatility stems from its ability to assess differences between group means under different scenarios, making it crucial for drawing comparisons and testing hypotheses. This abstract delves into the core principles of the t-test, exploring its variations, assumptions, and applications.

Core Concepts and Variations:

At its heart, the t-test assumes normally distributed data and utilizes the t-statistic, a measure of how many standard deviations the observed mean difference between groups falls from a hypothesized value (Cohen et al., 2014). Depending on the research question, two main variants exist:

- One-sample t-test: Compares the mean of a single group to a hypothesized value (e.g., testing if exam scores differ significantly from the average of 70)
- Two-sample t-test: Compares the means of two independent groups (e.g., evaluating if new teaching methods impact learning outcomes compared to traditional ones).

Further variations include paired t-tests, where data from the same subjects are compared across conditions, and Welch's t-test, suitable for unequal group variances (Field, 2009).

Assumptions and Considerations:

While powerful, the t-test relies on crucial assumptions: normality of data distribution, homogeneity of variances (except for Welch's test), and independence of observations (Cohen et al., 2014). Violation of these

continued on next page

continued from next page

assumptions can lead to inaccurate results. Therefore, researchers need to assess and, if necessary, transform data or employ alternative non-parametric tests.

Applications and Significance:

The t-test finds application in numerous areas. In clinical trials, it compares treatment efficacy across groups. In psychology, it assesses differences in cognitive performance or attitudes based on interventions. In economics, it evaluates the impact of policy changes on income levels. Its flexibility caters to a wide range of research questions, making it a cornerstone of statistical analysis.

Sources:

1. Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2014). Applied multiple regression/correlation analysis for the behavioral sciences. Routledge.
2. Field, A. (2009). Discovering statistics using SPSS (3rd ed.). SAGE Publications Ltd.

Limitations:

While providing valuable insights, the t-test has limitations. It is sensitive to outliers and violations of assumptions, and its interpretations require careful consideration of effect sizes and confidence intervals. For more complex comparisons or non-normal data, alternative statistical methods are necessary.

In conclusion, the t-test represents a powerful and versatile tool for researchers across disciplines. Understanding its core concepts, variations, assumptions, and limitations is essential for conducting and interpreting meaningful statistical analyses.

Note: This abstract is approximately one page long and includes in-text citations for the mentioned sources. It provides a basic overview of the t-test; further details and specific applications can be explored based on your specific research area.

Similarly to ChatGPT, Gemini also states that t-test requires normally distributed data, without mentioning that the assumption can be relaxed if the dataset is large and the underlying distribution fulfills the assumption of the central limit theorem. Apart from this slight misunderstanding, the Core Concepts and Variations section gives good explanations of the different variations and gives relevant examples. The section could benefit from formulae for the one and two-sampled t-tests.

The Assumptions and Considerations section is quite similar to the fourth section in ChatGPT's response. Once again it repeats the statement of assuming normal data distribution. Again, the independence of the observations is assumed correctly. The homogeneity of variance assumption is more precise in Gemini's response. Unlike

ChatGPT's response, this response includes Welch's t-test and mentions that this variation of the t-test doesn't require the assumption of same variances, which is correct [26].

The first thing one might notice about the references is their odd placement within the text. However, we have decided to not subtract any points from this, as it is mainly just an aesthetic factor, and doesn't affect the content of the references.

The two references in this text are mostly correct, only the year referenced on the first book is not correct. The book was, however, released as an eBook in 2014. The newest edition of the book seems to have been released in 2002 [27].

Both the student and the teacher agreed, that Gemini produced better text. Although both were imprecise in the required assumptions, ChatGPT also made severe mistakes in describing the calculation of a t-test. Gemini did not include the calculation at all. In this case, it seems that the professor and student agreed that including an incorrect calculation is worse than leaving it out.

Although Gemini is considered the better option, it should be noted that the appearance of ChatGPT is a better representation of an abstract. However, since these qualities are secondary to the content, it is not enough to make ChatGPT the better response.

4.1.3 Derivative

Example 5: Text generated by ChatGPT on Derivative

Title: Exploring the World of Derivatives: A Comprehensive Overview

Abstract:

Derivatives have become integral tools in modern finance, playing a pivotal role in risk management, speculation, and hedging strategies. This paper provides a comprehensive overview of derivatives, encompassing their types, pricing models, applications, and associated risks.

The study begins by delineating the foundational concepts of derivatives, elucidating their definition as financial contracts whose value derives from an underlying asset. It further examines the various types of derivatives, including forwards, futures, options, and swaps, highlighting their distinctive features and applications in different market contexts.

Furthermore, this paper delves into the methodologies employed in pricing derivatives, discussing classical approaches such as the Black-Scholes model and more recent advancements like the Monte Carlo simulation method. Through empirical analysis and real-world examples, it evaluates the efficacy of these models in valuing derivatives accurately, considering factors such as

continued on next page

continued from next page

volatility, interest rates, and time to maturity.

In addition, the research explores the multifaceted applications of derivatives in financial markets, ranging from risk mitigation to speculative trading strategies. It examines how derivatives facilitate risk management by providing mechanisms for investors to hedge against adverse price movements and diversify their portfolios effectively. Furthermore, it elucidates how derivatives enable market participants to capitalize on price fluctuations and leverage their investment positions through speculation.

Moreover, this paper scrutinizes the inherent risks associated with derivatives, emphasizing their potential for amplifying market volatility and systemic risks. It evaluates the regulatory frameworks governing derivatives markets globally and assesses their effectiveness in safeguarding market stability and investor interests.

Finally, drawing upon a synthesis of empirical evidence and theoretical frameworks, this paper offers insights into the future prospects of derivatives markets, anticipating trends such as increased innovation in derivative products, heightened regulatory scrutiny, and evolving risk management practices.

Through a comprehensive analysis of derivatives, this paper aims to provide practitioners, academics, and policymakers with a nuanced understanding of these financial instruments, elucidating their role in contemporary financial markets and their implications for market dynamics and stability.

Sources:

1. Hull, John C. *Options, Futures, and Other Derivatives*. Pearson, 2017.
2. McDonald, Robert L. *Derivatives Markets*. Pearson Education, 2013.
3. Wilmott, Paul. *Paul Wilmott Introduces Quantitative Finance*. John Wiley & Sons, 2007.
4. Black, Fischer, and Myron Scholes. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy*, vol. 81, no. 3, 1973, pp. 637–654.
5. Cox, John C., et al. "Option Pricing: A Simplified Approach." *Journal of Financial Economics*, vol. 7, no. 3, 1979, pp. 229–263.
6. Merton, Robert C. "Theory of Rational Option Pricing." *The Bell Journal of Economics and Management Science*, vol. 4, no. 1, 1973, pp. 141–183.

continued on next page

continued from next page

7. Bakshi, Gurdip S., et al. "Valuing European Options by Numerical Integration: The Gauss–Hermite Quadrature Method." *Review of Financial Studies*, vol. 9, no. 2, 1996, pp. 277–300.
8. Basel Committee on Banking Supervision. "The Basel Committee's Responses to Frequently Asked Questions." Basel Committee on Banking Supervision Publications, 2020, www.bis.org.
9. European Securities and Markets Authority (ESMA). "Derivatives." ESMA Publications, 2022, www.esma.europa.eu.
10. U.S. Commodity Futures Trading Commission (CFTC). "Regulations." CFTC Regulations, 2023, www.cftc.gov.

Since this thesis is in mathematics and not in finance, the correctness of the content of this response will not be analysed further. The word derivative was understood as the term used in finance, not as the term used in calculus.

However, this response is a good example of when the generated response isn't what is expected, but is not an invalid response either. The LLM did not misunderstand the assignment, but rather used a different definition of the prompted word.

Despite the content of the response not being analysed, the references can still be verified. The first six references are existing articles and books, and they are cited correctly. However the last four are all incorrect in some way, each having mistakes worth one point each.

The first of the four, or the seventh overall, simply does not exist. The author has no such article [28] and the referenced journal does not include this article [29].

The last three references all have links to websites aren't exactly what is referred to. All of them are on the right track, as they link to the homepage of the correct website, but not the specific section of the websites where the actual content is found. Ultimately, the links are wrong and thus result in a one-point mistake each.

Thus the total score for the references is $4/10 = 0.4$.

Example 6: Text generated by Gemini on Derivative

Derivatives: Understanding Change

The derivative, a cornerstone of calculus, unveils the intricate relationship between change and function behavior. In essence, it quantifies the instantaneous rate of change of a function's output (dependent variable) with respect to its input (independent variable) ([Math is Fun, n.d.]). Imagine a car speeding down a highway; the derivative, at any given moment, captures the car's instantaneous speed, not just its average speed over a certain distance.

continued on next page

continued from next page

Geometrically, the derivative is visualized as the slope of the tangent line to a function's graph at a specific point ([Britannica, 2024]). This tangent line provides a linear approximation of the function's behavior in the immediate vicinity of that point. For instance, the derivative of the function representing the position of a falling object over time would yield its instantaneous velocity at any given time instance.

Finding the derivative, a process known as differentiation, involves applying a set of well-defined rules. These rules, like the product rule, quotient rule, and chain rule, provide a systematic approach to calculating derivatives of various functions ([Derivative Calculator, n.d.]). Notably, not all functions are differentiable at all points, and understanding these limitations is crucial for accurate interpretation.

The significance of derivatives extends far beyond theoretical understanding. They underpin numerous applications across diverse fields, from physics and engineering (modeling motion, optimizing designs) to economics and finance (analyzing market trends, managing risk). By harnessing the power of derivatives, we gain a deeper understanding of change and its impact, empowering us to make informed decisions and solve complex problems in the real world.

Sources:

1. Britannica. (2024, January 24). Derivative. <https://www.britannica.com/money/derivatives>
2. Math is Fun. (n.d.). Introduction to Derivatives. <https://www.mathsisfun.com/definitions/derivative.html>
3. Derivative Calculator. (n.d.). Derivative Calculator • With Steps! <https://www.derivative-calculator.net/>

In the first chapter, an example of a car speeding down a highway is given. Although it is a simple and easily understandable example, it doesn't specify what function is being differentiated. For better clarity, one could fix this by writing "the derivative of **the car's position** is the car's instantaneous speed."

The text doesn't give the definition of a derivative as a limit. Some differentiation rules are mentioned by name, but they all stem from this definition.

The text also briefly touches on functions that are not differentiable, but could further elaborate on what kind of qualities make a function or a point non-differentiable, for example non-continuous functions, or when the graph of a function has a corner point.

In comparison to other responses, this is one of the shortest responses. This unfortunately seems to have a negative effect on the quality, having it fall short and

lack in detail.

The references cited here are all incorrect in some way, all being severe enough mistakes resulting in the score being the maximum $3/3 = 1$ point.

The first of the three links to a Britannica website with the definition of derivatives. The article is titled "derivatives" in plural, not singular. This alone is worth one point, but additionally the date is incorrect, as at the time the latest edition of the article was published in 2023, not 2024. The article does not give a specific author or authors, however it does state that it was "Written and fact-checked by the editors of Encyclopaedia Britannica", and they should be credited accordingly. [30]

The second reference links to another definition of derivative, this time by a website called mathisfun. The written reference refers to the "Introduction to Derivatives" section of the website, which is different to the "Definition of Derivative" page, which is the linked page. It also doesn't give credit to the author, which is Rod Pierce. It also doesn't give a year, which would be 2024 for the Introduction page and 2023 for the Definition page. [31], [32]

Finally, the last reference links to an online derivative calculator by David Scherfgen [33], but no credit is given to the creator.

The professor didn't have a preference on the two, because they covered different topics. The student preferred Gemini as it at least covered the desired definition of derivative. Both of these opinions are valid and make sense. While one did not see the need to compare the two, given that they are on two completely different topics, the other automatically gave the point to the one that is on the mathematical definition.

4.1.4 Gaussian processes

For the last example, one of the more difficult topics is chosen for analysis. Additionally, an example where the student and professor had different preferences is chosen. One such topic is Gaussian processes, which is the chosen topic.

Example 7: Text generated by ChatGPT on Gaussian processes

Abstract

Gaussian processes (GPs) are powerful probabilistic models used in various fields, from machine learning to Bayesian optimization, due to their flexibility and ability to model complex data distributions. In recent years, there has been growing interest in extending GPs to incorporate information from multiple heterogeneous data sources, leading to the development of Gaussian processes with sources (GPS).

This abstract presents an overview of Gaussian processes with sources, highlighting their key concepts, methodologies, and applications. First, we discuss the basic framework of Gaussian processes, emphasizing their probabilistic nature and their use as non-parametric models for regression

continued on next page

continued from next page

and classification tasks. We then delve into the extension of GPs to Gaussian processes with sources, which enables the incorporation of diverse data modalities or information from multiple domains into a unified modeling framework.

We explore various approaches for integrating multiple data sources within the GPS framework, including concatenation-based methods, feature augmentation techniques, and multi-output Gaussian processes. These methods allow for the seamless fusion of information from different sources while preserving the inherent uncertainty estimates provided by Gaussian processes.

Furthermore, we discuss the applications of Gaussian processes with sources across various domains, such as healthcare, finance, and environmental science. In healthcare, GPS models have been used for personalized medicine, integrating genomic, clinical, and imaging data to predict patient outcomes and recommend treatment strategies. In finance, GPS have been applied to portfolio optimization, combining information from diverse asset classes to improve investment decision-making. In environmental science, GPS have been used for climate modeling, assimilating data from satellites, weather stations, and numerical simulations to enhance predictive accuracy.

In conclusion, Gaussian processes with sources offer a flexible and principled framework for modeling heterogeneous data sources, enabling more accurate predictions and informed decision-making across a wide range of applications. Further research in this area holds great potential for advancing the capabilities of probabilistic modeling in handling complex, multi-source data environments.

In-text citations:

1. Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press.
2. Bonilla, E. V., Chai, K. M. A., & Williams, C. K. I. (2008). Multi-task Gaussian process prediction. In *Advances in neural information processing systems* (pp. 153-160).
3. Vanhatalo, J., Riihimäki, J., Hartikainen, J., Jylänki, P., Tolvanen, V., & Vehtari, A. (2013). GPstuff: Bayesian modeling with Gaussian processes. *Journal of Machine Learning Research*, 14(1), 1175-1179.
4. Dezfouli, A., & Bonilla, E. V. (2015). Gaussian processes with Beta-Sherman-Woodbury kernels: Application to multi-source time-series data. In *Advances in Neural Information Processing Systems* (pp. 1191-1199).

continued on next page

continued from next page

5. Lázaro-Gredilla, M., Quiñonero-Candela, J., Rasmussen, C. E., & Figueiras-Vidal, A. R. (2010). Sparse spectrum Gaussian process regression. *Journal of Machine Learning Research*, 11(Feb), 1865-1881.
6. Wilson, A. G., Dann, C., & Nickisch, H. (2016). STAN: A probabilistic programming language. *Journal of Statistical Software*, 20(1), 22-43.
7. Alaa, A. M., & van der Schaar, M. (2017). Bayesian inference of individualized treatment effects using multi-task Gaussian processes. In *Advances in Neural Information Processing Systems* (pp. 4028-4038).

This abstract provides very little information on what Gaussian processes are, it just continuously makes promises to get into various concepts related to Gaussian processes, but failing to ever do so.

The first, third and fifth references have all the correct information. The second and seventh references both have small mistakes worth 1/2 points each and the fourth and sixth have 1-point mistakes, giving a total of $3/7 \approx 0.43$ points.

The second article is cited correctly, but it is missing the volume of the journal, which would be the 20th volume [34]. The same problem is seen with the fourth and seventh references. All three of these references cite articles in a journal called *Advances in Neural Information Processing Systems*. *Neural Information Processing Systems (NeurIPS)* is a conference in which the proceedings are compiled in the aforementioned journal. It seems that A did not understand how to cite this correctly, as it did not recognize the different volumes of the journal. The seventh reference should have included the volume, which is 30 [35]. The seventh reference also seemingly made up the pages, but this is not confirmed, so no additional points we given as a benefit of doubt. The fourth reference has problems more severe than an incorrect volume, so that is ignored for now.

The fourth reference likely attempted to refer to a paper titled *Scalable Inference for Gaussian Process Models with Black-Box Likelihoods* [36], as it is written by the cited authors and no paper exists with the title "Gaussian processes with Beta-Sherman-Woodbury kernels: Application to multi-source time-series data".

The sixth article exists in the 76th volume of the *Journal of Statistical Software* [37]. The journal is thus correct, but the volume is incorrect. The most severe mistake though, is that the authors are completely wrong. The cited authors have worked together on a different article related to Gaussian processes. This article is titled "Thoughts on Massively Scalable Gaussian Processes" [38]. Neither of these articles were written in 2016, but in 2017 and 2015.

Example 8: Text generated by Gemini on Gaussian processes

Gaussian Processes: A Probabilistic Approach to Function Learning

continued on next page

continued from next page

Gaussian processes (GPs) offer a powerful and versatile framework for machine learning tasks involving function learning. They model data as a collection of random variables with a joint multivariate normal distribution, enabling the representation of uncertainty and the prediction of unseen data points [2]. This one-page abstract delves into the core concepts of GPs, their applications, and their unique advantages in various machine learning scenarios.

Conceptual Foundations:

At its core, a GP defines a distribution over functions, where any finite collection of function values follows a multivariate normal distribution [1]. This probabilistic approach allows GPs to capture the relationship between input and output variables, while also providing uncertainty estimates for predictions [5]. The key component in defining a GP is the kernel function, which determines the similarity between data points and governs the smoothness and behavior of the inferred function [4]. Different kernel functions can be chosen based on the specific problem and prior knowledge about the data.

Applications and Advantages:

GPs have found applications in various machine learning domains, including:

- **Regression:** Predicting continuous-valued outputs based on input features. GPs excel in tasks where the underlying function is complex or non-linear, as they can learn flexible relationships without overfitting the data [3].
- **Classification:** Classifying data points into discrete categories. Gaussian Process Classifiers (GPCs) provide probabilistic outputs, indicating the confidence in the predicted class labels [3].
- **Time series forecasting:** Predicting future values in a sequence of data points. GPs can capture temporal dependencies and uncertainties in time series data, making them suitable for various forecasting tasks.

Several advantages distinguish GPs from other machine learning methods:

- **Non-parametric nature:** GPs do not require explicit assumptions about the underlying functional form, making them adaptable to diverse data distributions.
- **Uncertainty quantification:** GPs inherently provide uncertainty estimates for predictions, allowing for a more nuanced understanding of the model's confidence in its outputs.
- **Bayesian framework:** GPs naturally integrate with Bayesian inference, enabling the incorporation of prior knowledge and the construction of

continued on next page

continued from next page

robust models.

Conclusion:

Gaussian processes offer a powerful and flexible framework for function learning in machine learning. Their ability to model complex relationships, provide uncertainty estimates, and integrate with Bayesian inference makes them valuable tools for various applications. While their computational complexity can be a limitation in certain scenarios, ongoing research continues to explore efficient algorithms and improve the practical applicability of GPs.

Sources:

1. Wikipedia. (2023, November 19). Gaussian process. https://en.wikipedia.org/wiki/Gaussian_process
2. Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian processes for machine learning. The MIT Press.
3. scikit-learn [scikit-learn.org](https://scikit-learn.org/stable/modules/gaussian_process.html). (n.d.). 1.7. Gaussian processes. https://scikit-learn.org/stable/modules/gaussian_process.html
4. Distill. (2019, August 28). A visual exploration of Gaussian processes. [invalid URL removed]
5. d2l.ai. (n.d.). 18.1. Introduction to Gaussian processes. <https://www.d2l.com/blog/d2l-and-artificial-intelligence-the-past-present-and-future/>

This text included more actual content than ChatGPT did, but some of that content was vaguely presented.

The second and third references exist and are cited correctly. The first reference has a 1/2 point mistake and the last two have mistakes worth 1 point each. It should also be noted that using Wikipedia as a source is often considered bad practise and should be avoided, but we have chosen to not reduce any additional points relating to the quality of source.

Regarding the first reference, the latest edition to the Wikipedia article (as of March 2023) was made in January 2024, not November 2023. From the "View History" Section on Wikipedia, one can view archived versions of a Wikipedia article, so it technically could be correct. However, the dates for previous edits do not include November 19, 2023. Thus the date cited by Gemini is incorrect. [39]

The fourth reference fails to give credit to the authors J. Görtler, R. Kehlbeck, and O. Deussen [40]. The last reference links to a page that the reference text itself does not refer to. Both these mistakes are recognized as one-point mistakes.

These make for a total of $2.5/5 = 0.5$ points, which is just slightly more than the

score of ChatGPT (0.43).

Comparing the two texts is a prime example of choosing between an "empty content" text and a text with many mistakes, but more actual content as well. The student preferred the one with incorrect information, which is Gemini, whilst the professor preferred ChatGPT.

4.2 Other observations

Mathematical text usually has formulae, but neither ChatGPT or Gemini provided them in many of the texts. Only four texts by ChatGPT and another four texts by Gemini had formulae. Two of these were on the same topics, one of them being eigenvalue and the other Cauchy distribution. The other four were normal distribution and standard deviation for ChatGPT and linear regression and mean absolute deviation for Gemini. Granted, some of the topics, such as random variables, one would probably not expect a formula for such a short text, as the topics are quite broad or can be easily explained without a formula. However, overall the lack of formulae was unusual for mathematical text, even with the short length of the texts.

The prompt asked for in-text citations, which Gemini provided in each generated text, but ChatGPT did not. There were only 5 texts by ChatGPT that included in-text citations. Although we did not end up checking the in-text citations or doing anything with them, it is interesting that ChatGPT more often than not ignored that part of the prompt. It is unclear why this is, or if it correlates with the quality of references for both ChatGPT and Gemini. Perhaps a confusion about what references and in-text citations are could have been a disadvantage for Gemini, and ChatGPT knew not to create non sensible in-text citations which may or may not have cited the generated references.

The style of text by ChatGPT was slightly more professional and suitable for scientific writing, whereas Gemini was more expressive with its writing. This led to some interesting snippets of text, such as "Random variables empower us to analyze various aspects of data". Some mathematical concept empowering people is a quite goofy way of expressing its purpose. The different styles of writing can be seen quite well in just the generated titles alone. The titles, particularly by Gemini, were often times quite silly and repetitive. For example, a good portion of the titles by Gemini had the word "demystify" in them. Some of them include, "Demystifying the Unseen: A Dive into Statistical Inference", "Demystifying Chance: An Introduction to Random Variables", and "Demystifying Data Dispersion: A Dive into Standard Deviation". Even more so, the titles by Gemini would include the word "unveil", such as "Skewness: Unveiling Asymmetry in Data", "Eigenvalues: Unveiling the Hidden Structure of Linear Transformations" and "Unveiling Relationships: A Look at the Correlation Coefficient".

In contrast to this, ChatGPT sometimes simply titled the texts "Abstract" or "Understanding [topic]". There was far less repetitiveness in the titles of ChatGPT, apart from the word "abstract", but the common use of the word "abstract" makes sense. The words and sentences "understanding", "unraveling", "exploring" and "a comprehensive overview", were repeated in the titles a few times, but far less than

Gemini repeated the words "demystifying" and "unveiling". The titles by ChatGPT felt more appropriate for the content of the text and the style of writing that was expected.

Sometimes both ChatGPT and Gemini provided a word count at the end of the text. This value was always the incorrect number of words in the text.

The writing styles also differed in the amount of lists generated. For 11 of the 25 topics, Gemini would have some form of list in its text. For example, the sign test, t-test and Gaussian processes that were analysed earlier all had lists. This is quite unusual for abstracts, but a fairly useful tool for briefly explaining the basics of a topic. ChatGPT included lists in two of the 25 texts.

For the references, Gemini sometimes provided a separate info message, which included a source or a few sources that were supposedly applied in the text. These references were not included in the analysis as they were not a part of the generated texts themselves. Interestingly, often these references in the info message would not include those references that had actually been generated.

When fact-checking the information in the generated responses, it was clearly noticeable that a lot of the text, especially the first few sentences seemed to have been either directly plagiarized from Wikipedia or the words or their order had been changed slightly. For Google Gemini, this could be explained by the fact that Gemini uses all of Google as a training set, and it is likely that there are several other (non-academic) sources that have also plagiarized from Wikipedia, either directly or indirectly.

5 Discussion

One interesting result of this study was that the better LLM for generating mathematical text (Gemini) was not the same LLM that was better at generating references (ChatGPT).

The mistakes in the content of the texts were sometimes quite well hidden, meaning they could be hard to notice unless one knows the topic well or is in the field. The text often times looks very professional and thus is very convincing, however the underlying content can be wrong. Frequently the mistakes weren't far off from the correct information, making the mistakes look suspiciously accurate.

The professor also noted that choosing the better option of the LLMs was more about choosing the less bad option rather than about choosing the better option. This highlights how bad the texts were from a professional standpoint.

ChatGPT and Gemini simply don't have the skills as of how to produce accurate mathematical text. They can generate text just fine, but the text may not have correct information, as LLMs do not think the same way that humans do. They are made simply to generate text using a massive training set of string data, but especially for creating a new way of writing a specific topic, there is no guarantee that the content of that text is correct.

However, the popularity of LLMs are growing rapidly and they are being constantly developed, so in the future there could be major changes in how the LLMs understand mathematics. LLMs have a lot of potential, but can also effect the critical thinking skills of people. People shouldn't rely on texts produced by LLMs for their information.

To conclude, ChatGPT and Gemini are great text generators, but not information generators and should not be used as such.

References

- [1] E. Kasneci, K. Sessler, S. Küchemann, *et al.*, “Chatgpt for good? on opportunities and challenges of large language models for education”, *Learning and Individual Differences*, vol. 103, p. 102 274, 2023, ISSN: 1041-6080. DOI: <https://doi.org/10.1016/j.lindif.2023.102274>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1041608023000195>.
- [2] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, “Natural language processing: an introduction”, *Journal of the American Medical Informatics Association*, vol. 18, no. 5, pp. 544–551, Sep. 2011, ISSN: 1067-5027. DOI: [10.1136/amiajnl-2011-000464](https://doi.org/10.1136/amiajnl-2011-000464). eprint: <https://academic.oup.com/jamia/article-pdf/18/5/544/5962687/18-5-544.pdf>. [Online]. Available: <https://doi.org/10.1136/amiajnl-2011-000464>.
- [3] OpenAI, *Introducing chatgpt*, 2022. [Online]. Available: <https://openai.com/blog/chatgpt>.
- [4] K. Hu, “Chatgpt sets record for fastest-growing user base—analyst note”, *Reuters Technology*, 2023. [Online]. Available: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>.
- [5] OpenAI, *Introducing chatgpt plus*, 2023. [Online]. Available: <https://openai.com/blog/chatgpt-plus>.
- [6] S. Pichai, *A message from our ceo: An important next step on our ai journey*, 2023. [Online]. Available: <https://blog.google/technology/ai/bard-google-ai-search-updates/>.
- [7] Gemini, *Updates*. [Online]. Available: <https://gemini.google.com/updates>.
- [8] S. An, Z. Ma, Z. Lin, N. Zheng, J.-G. Lou, and W. Chen, *Learning from mistakes makes llm better reasoner*, 2023. arXiv: [2310.20689](https://arxiv.org/abs/2310.20689) [cs.CL].
- [9] K. Cobbe, V. Kosaraju, M. Bavarian, *et al.*, *Training verifiers to solve math word problems*, 2021. arXiv: [2110.14168](https://arxiv.org/abs/2110.14168) [cs.LG].
- [10] V. Cheng and Y. Zhang, “Analyzing chatgpt’s mathematical deficiencies: Insights and contributions”, in *Proceedings of the 35th Conference on Computational Linguistics and Speech Processing (ROCLING 2023)*, 2023, pp. 188–193.
- [11] Z. Yuan, H. Yuan, C. Tan, W. Wang, and S. Huang, *How well do large language models perform in arithmetic tasks?*, Mar. 2023.
- [12] S. Frieder, L. Pinchetti, R.-R. Griffiths, *et al.*, “Mathematical capabilities of chatgpt”, *Advances in Neural Information Processing Systems*, vol. 36, 2024.

- [13] P. Giannos and O. Delardas, “Performance of chatgpt on uk standardized admission tests: Insights from the bmat, tmua, lnat, and tsa examinations”, *JMIR Med Educ*, vol. 9, e47737, Apr. 2023, ISSN: 2369-3762. DOI: [10.2196/47737](https://doi.org/10.2196/47737). [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/37099373>.
- [14] J. de Winter, *Can chatgpt pass high school exams on english language comprehension?*, Sep. 2023. DOI: [10.1007/s40593-023-00372-z](https://doi.org/10.1007/s40593-023-00372-z). [Online]. Available: <https://doi.org/10.1007/s40593-023-00372-z>.
- [15] M. Rosoł, J. S. Gąsior, J. Łaba, K. Korzeniewski, and M. Młyńczak, “Evaluation of the performance of gpt-3.5 and gpt-4 on the polish medical final examination”, *Scientific Reports*, vol. 13, 1 2023. DOI: [10.1038/s41598-023-46995-z](https://doi.org/10.1038/s41598-023-46995-z). [Online]. Available: <https://doi.org/10.1038/s41598-023-46995-z>.
- [16] D. Stribling, Y. Xia, M. K. Amer, K. S. Graim, C. J. Mulligan, and R. Renne, “The model student: Gpt-4 performance on graduate biomedical science exams”, *Scientific Reports*, vol. 14, 1 2024. DOI: [10.1038/s41598-024-55568-7](https://doi.org/10.1038/s41598-024-55568-7). [Online]. Available: <https://doi.org/10.1038/s41598-024-55568-7>.
- [17] J. Kocoń, I. Cichecki, O. Kaszyca, *et al.*, “Chatgpt: Jack of all trades, master of none”, *Information Fusion*, vol. 99, p. 101 861, 2023, ISSN: 1566-2535. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S156625352300177X>.
- [18] R. Hogg, J. McKean, and A. Craig, *Introduction to Mathematical Statistics* (Pearson education international). Pearson Education, 2005, pp. 518–531, ISBN: 9780130085078. [Online]. Available: <https://books.google.fi/books?id=vIEZAQAIAAJ>.
- [19] T. K. Kim, “T test as a parametric statistic”, *Korean journal of anesthesiology*, vol. 68, no. 6, pp. 540–546, 2015. [Online]. Available: <https://doi.org/10.4097/kjae.2015.68.6.540>.
- [20] OpenAI, *Memory and new controls for chatgpt*, 2024. [Online]. Available: <https://openai.com/index/memory-and-new-controls-for-chatgpt>.
- [21] *Bard privacy notice*. [Online]. Available: <https://support.google.com/bard/answer/13594961?hl=en>.
- [22] I. Peterson, 2014. [Online]. Available: <https://mathcomm.org/a-student-guide-to-writing-an-abstract/>.
- [23] D. S. S. Ashkan Ertefaie and P. R. Rosenbaum, “Quantitative evaluation of the trade-off of strengthened instruments and sample size in observational studies”, *Journal of the American Statistical Association*, vol. 113, no. 523, pp. 1122–1134, 2018. DOI: [10.1080/01621459.2017.1305275](https://doi.org/10.1080/01621459.2017.1305275). [Online]. Available: <https://doi.org/10.1080/01621459.2017.1305275>.

- [24] H.-Y. Kim, “Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis”, *Restorative dentistry endodontics*, vol. 38, no. 1, pp. 52–54, 2013. [Online]. Available: <https://doi.org/10.5395/rde.2013.38.1.52>.
- [25] H.-Y. Kim, “Statistical notes for clinical researchers: Effect size”, *Restorative dentistry endodontics*, vol. 40, no. 4, pp. 328–331, 2015. [Online]. Available: <https://doi.org/10.5395/rde.2015.40.4.328>.
- [26] B. L. Welch, “The generalization of ‘Student’s’ problem when several different population variances are involved”, *Biometrika*, vol. 34, no. 1-2, pp. 28–35, Jan. 1947, ISSN: 0006-3444. DOI: [10.1093/biomet/34.1-2.28](https://doi.org/10.1093/biomet/34.1-2.28). eprint: <https://academic.oup.com/biomet/article-pdf/34/1-2/28/553093/34-1-2-28.pdf>. [Online]. Available: <https://doi.org/10.1093/biomet/34.1-2.28>.
- [27] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd ed. Routledge, 2002. DOI: [10.4324/9780203774441](https://doi.org/10.4324/9780203774441).
- [28] Google scholar - gurdip bakshi. [Online]. Available: <https://scholar.google.com/citations?user=Oigg-tkAAAAJ>.
- [29] *The Review of Financial Studies*, vol. 9, 2 1996. [Online]. Available: <https://www.jstor.org/stable/i352839>.
- [30] The Editors of Encyclopaedia, “Derivatives”, *Encyclopedia Britannica*, [Online]. Available: <https://www.britannica.com/money/derivatives>.
- [31] R. Pierce, *Definition of derivative*, 2023. [Online]. Available: <https://www.mathsisfun.com/definitions/derivative.html>.
- [32] R. Pierce, *Introduction to derivatives*, 2024. [Online]. Available: <https://www.mathsisfun.com/calculus/derivatives-introduction.html>.
- [33] D. Scherfgen, *Derivative calculator*, 2024. [Online]. Available: <https://www.derivative-calculator.net/>.
- [34] E. Bonilla, K. Chai, and C. Williams, “Multi-task gaussian process prediction”, *Proc. Adv. Neural Inf. Process. Syst*, vol. 20, pp. 153–160, Jan. 2008.
- [35] A. M. Alaa and M. van der Schaar, “Bayesian inference of individualized treatment effects using multi-task gaussian processes”, in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, et al., Eds., vol. 30, Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/6a508a60aa3bf9510ea6acb021c94b48-Paper.pdf.

- [36] A. Dezfouli and E. V. Bonilla, “Scalable inference for gaussian process models with black-box likelihoods”, in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., vol. 28, Curran Associates, Inc., 2015. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2015/file/3b3dbaf68507998acd6a5a5254ab2d76-Paper.pdf.
- [37] B. Carpenter, A. Gelman, M. D. Hoffman, *et al.*, “Stan: A probabilistic programming language”, *Journal of Statistical Software*, vol. 76, no. 1, pp. 1–32, 2017. DOI: [10.18637/jss.v076.i01](https://doi.org/10.18637/jss.v076.i01). [Online]. Available: <https://www.jstatsoft.org/index.php/jss/article/view/v076i01>.
- [38] A. Wilson, C. Dann, and H. Nickisch, “Thoughts on massively scalable gaussian processes”, Nov. 2015.
- [39] “Gaussian process: Revision history”, *Wikipedia*, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Gaussian_process&action=history.
- [40] J. Görtler, R. Kehlbeck, and O. Deussen, “A visual exploration of gaussian processes”, *Distill*, 2019, <https://distill.pub/2019/visual-exploration-gaussian-processes>. DOI: [10.23915/distill.00017](https://doi.org/10.23915/distill.00017).

A Appendix

All of the texts generated by ChatGPT and Gemini used in this thesis are below.

ChatGPT: Sign test

Abstract

The sign test, a non-parametric statistical method, is a valuable tool for analyzing data when the assumptions of parametric tests are not met or when dealing with small sample sizes. This test is particularly useful when dealing with ordinal or skewed data, or when the underlying distribution is unknown or non-normal. The sign test evaluates whether the median of a dataset differs significantly from a hypothesized value, making it robust against outliers and resistant to the effects of non-normality.

This paper provides an overview of the sign test, including its assumptions, calculation, and interpretation. Additionally, it discusses its application in various fields such as medicine, environmental science, and social sciences where parametric assumptions may be violated or where small sample sizes are common. Through a review of relevant literature and case studies, this paper demonstrates the versatility and utility of the sign test in practical research settings.

The sign test has been extensively studied and applied in research literature. In a study by Wilcox (2016), the sign test was utilized to analyze the efficacy of a new treatment for chronic pain, demonstrating its applicability in medical research. Furthermore, in a study by Ertefaie et al. (2018), the sign test was employed to assess the impact of air pollution on respiratory health, highlighting its use in environmental science.

In conclusion, the sign test is a valuable statistical method for analyzing data in situations where parametric assumptions are not met or when dealing with small sample sizes. Its robustness, simplicity, and versatility make it a valuable tool for researchers across various disciplines.

References

1. Ertefaie, A., Small, D. S., & Flory, J. H. (2018). sign tests for the matched-pair design: assessing the effect of air pollution on health outcomes. *Journal of the American Statistical Association*, 113(523), 1415-1427.
2. Wilcox, R. R. (2016). *Introduction to robust estimation and hypothesis testing*. Academic Press.

Gemini: Sign test

The sign test: A Powerful Non-Parametric Tool for Statistical Analysis

The sign test, a versatile non-parametric statistical method, stands out for its simplicity and broad applicability. Unlike parametric tests requiring specific assumptions about data distribution, the sign test relies solely on the directional difference between paired observations (Field, 2020). This characteristic makes it particularly valuable when:

- Data distribution is unknown or non-normal: Many real-world datasets deviate from normality, rendering parametric tests unsuitable (Siegel, 1956). The sign test bypasses this issue by focusing on the sign of change (+/-) rather than the magnitude of differences.
- Sample sizes are small: With limited data, the robustness of the sign test becomes advantageous (Cramèr, 1946). While parametric tests lose power with small samples, the sign test maintains its reliability.
- Matched pairs are present: The sign test thrives when observations exist as matched pairs, allowing comparisons within individuals before and after treatment, for example (McDonald, 2014).

Applications and Variations:

The sign test boasts diverse applications across various fields, including psychology, medical research, and economics. It can be used to:

- Compare two populations: Assess if a new medication truly impacts pain levels compared to a placebo (Spren, 1999).
- Test for a median difference: Examine if customer satisfaction scores differ significantly after website revamp (Siegel & Castellan, 1988).
- Identify outliers: Detect observations deviating drastically from the paired trends within groups (Conover, 1999).

While the basic sign test analyzes paired data, variations cater to specific scenarios:

- McNemar's test: Compares two dependent nominal categories, suitable for binary outcomes like pass/fail (McNemar, 1947).
- Wilcoxon signed-rank test: Assigns ranks to differences within pairs, offering more information than just sign (+/-) (Wilcoxon, 1945).

Strengths and Limitations:

The sign test's main strengths lie in its simplicity, flexibility, and robustness. However, limitations exist:

continued on next page

continued from next page

- Loss of information: Focusing only on sign discards valuable information about the magnitude of differences.
- Lower power than parametric tests: When data meets parametric assumptions, parametric tests might be more powerful.

Conclusion:

The sign test, with its ease of use and minimal assumptions, stands as a valuable tool for researchers and analysts. Understanding its applications, variations, and limitations empowers informed decision-making in various statistical analyses.

Sources:

1. Conover, W. J. (1999). Practical nonparametric statistics (3rd ed.). Wiley.
2. Cramèr, H. (1946). Mathematical methods of statistics. Princeton University Press.
3. Field, A. (2020). Discovering statistics using IBM SPSS Statistics (5th ed.). Sage.
4. McDonald, J. H. (2014). Handbook of parametric and nonparametric statistical procedures. Routledge.
5. McNemar, Q. (1947). Psychological statistics. John Wiley & Sons.
6. Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. McGraw-Hill.
7. Siegel, S., & Castellan, N. J. (1988). Nonparametric statistics for the behavioral sciences (2nd ed.). McGraw-Hill.
8. Sprent, P. (1999). Applied nonparametric statistical methods (2nd ed.). Chapman and Hall.
9. Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics Bulletin, 1(6), 80-83.

ChatGPT: p-value

Understanding the Significance of p-values in Statistical Inference

Abstract:

continued on next page

continued from next page

The p-value is a crucial statistical measure widely used in hypothesis testing, particularly in scientific research. It quantifies the strength of evidence against the null hypothesis and aids researchers in making informed decisions regarding the acceptance or rejection of hypotheses. This abstract explores the significance of p-values, their interpretation, and the controversies surrounding their usage.

Initially introduced by Sir Ronald Fisher in the early 20th century, the p-value has become a fundamental tool in statistical analysis. It represents the probability of obtaining results as extreme as or more extreme than the observed data under the assumption that the null hypothesis is true. A low p-value indicates strong evidence against the null hypothesis, suggesting that the observed data are unlikely to have occurred by chance alone.

However, the interpretation of p-values is often misunderstood. A common misconception is that a low p-value indicates the practical significance or importance of a result, which is not necessarily true. Furthermore, the arbitrary threshold of 0.05 as a cutoff for statistical significance has been criticized for promoting dichotomous thinking and disregarding the context of the research.

Recent discussions in the scientific community have emphasized the limitations and misuse of p-values. The American Statistical Association (ASA) released a statement in 2016 cautioning against the misuse of p-values and advocating for a more nuanced interpretation of statistical results. They emphasized the importance of considering effect sizes, confidence intervals, and practical significance in addition to p-values.

In conclusion, while p-values are valuable tools in statistical inference, their interpretation requires caution and consideration of various factors beyond a simplistic binary decision based solely on the numerical value. Researchers should aim for a more comprehensive understanding of statistical results, incorporating effect sizes and confidence intervals to ensure robust and meaningful conclusions in scientific research.

Sources:

1. Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd.
2. Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, 70(2), 129–133. <https://doi.org/10.1080/00031305.2016.1154108>

continued on next page

continued from next page

3. Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., ... Johnson, V. E. (2018). Redefine Statistical Significance. *Nature Human Behaviour*, 2(1), 6–10. <https://doi.org/10.1038/s41562-017-0189-z>

Gemini: p-value

Demystifying the P-Value: A Statistical Guidepost with Cautions

The p-value, a cornerstone of statistical hypothesis testing, holds immense power in shaping scientific conclusions. Yet, its interpretation is often fraught with pitfalls and misconceptions. This abstract aims to clarify the essence of p-values while highlighting their limitations and responsible usage.

In essence, the p-value represents the probability of observing results as extreme as, or more extreme than, those actually obtained, assuming the null hypothesis (no effect) is true (Wasserstein & Lazar, 2016). Lower p-values indicate less likelihood of chance occurrences, suggesting statistically significant effects (Greenwald et al., 1996). However, a crucial distinction exists between statistical significance and real-world importance (Cohen, 1994). Small p-values can arise from large studies detecting minute, inconsequential effects, while large p-values in smaller studies might mask clinically relevant findings.

Furthermore, arbitrary p-value thresholds (often 0.05) can create artificial dichotomies, leading to "p-hacking" manipulations to achieve desired significance (Head et al., 2015). Additionally, neglecting alternative explanations and multiple comparisons can inflate false positive rates (Rothman et al., 2014).

Therefore, responsible researchers require a nuanced understanding of p-values. Reporting effect sizes and confidence intervals alongside p-values provides additional context (Cumming, 2004). Prioritizing pre-registration of hypotheses and analysis plans fosters transparency and reduces bias (Nosek et al., 2018). Shifting focus from mere statistical significance to effect size, replicability, and real-world impact ensures research findings contribute meaningfully to scientific advancement.

Sources:

1. Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*,

continued on next page

continued from next page

- 49(12), 997-1003. <https://doi.org/10.1037/0003-066X.49.12.997>
2. Cumming, G. (2004). The new statistics: Why and how. *Psychological Science*, 15(5), 397-407. <invalid URL removed>
 3. Greenwald, A. G., Nosek, B. A., & Perlman, R. A. (1996). Meta-analysis of redundancy detection: Does size matter? *Psychonomic Bulletin & Review*, 3(4), 330-333. <invalid URL removed>
 4. Head, M. L., Holman, L., Kahn, R., Bahney, B. J., Daniel, J. K., & Prechter, M. (2015). The evidence for replicable effect sizes in psychological science. *Trends in Cognitive Sciences*, 19(5), 387-393. <https://doi.org/10.1016/j.tics.2015.02.007>
 5. Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breuning, C., .. & Wolbert, S. H. (2018). Preregistration is a research design feature, not a sign of fraud. *Proceedings of the National Academy of Sciences*, 115(11), 2590-2599. <invalid URL removed>
 6. Rothman, K. J., Greenland, S., & Lash TL. (2014). *Modern epidemiology* (3rd ed.). Lippincott Williams & Wilkins.
 7. Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2), 129-133. <invalid URL removed>

Note: This abstract is approximately 275 words, adhering to a one-page format.

ChatGPT: Linear regression

Abstract

Linear regression is a fundamental statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data. This method serves as a cornerstone in various fields such as economics, finance, sociology, and epidemiology, among others, providing insights into the nature of relationships between variables and facilitating predictive modeling.

This paper provides an overview of linear regression, discussing its basic principles, assumptions, and applications. The methodology involves estimating the parameters of the linear equation, typically using the method of least squares, which minimizes the sum of the squared differences between observed and predicted values. The resulting model can be used for prediction, inference,

continued on next page

continued from next page

and hypothesis testing.

Assumptions underlying linear regression include linearity, independence, homoscedasticity, and normality of errors. Violations of these assumptions can affect the validity and reliability of the regression model, emphasizing the importance of diagnostic checks and potential remedial actions.

Applications of linear regression are widespread. In economics, it is used to analyze relationships between variables such as supply and demand, inflation and unemployment, or income and consumption. In finance, linear regression is applied to predict stock prices, analyze risk factors, and estimate asset returns. In epidemiology, it helps understand the relationship between risk factors and disease outcomes.

This paper also explores extensions and variations of linear regression, including multiple linear regression, polynomial regression, and logistic regression, each tailored to specific research questions and data characteristics. Furthermore, it discusses the role of diagnostic tools such as residual analysis, leverage plots, and influential point detection in evaluating model assumptions and identifying outliers.

In conclusion, linear regression is a versatile and widely used statistical method with broad applications across various disciplines. Understanding its principles, assumptions, and limitations is essential for researchers and practitioners to effectively analyze data and draw meaningful conclusions.

Sources

1. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. John Wiley & Sons.
2. Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied Linear Regression Models*. McGraw-Hill Irwin.
3. Fox, J. (2015). *Applied Regression Analysis and Generalized Linear Models*. SAGE Publications.
4. Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach*. Cengage Learning.
5. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

continued on next page

continued from next page

- Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gemini: Linear regression

Unveiling Linear Relationships: A Glimpse into Linear Regression

Linear regression, a cornerstone of both statistics and machine learning, offers a powerful tool for understanding and predicting relationships between variables. At its core, it seeks to model a dependent variable (predicted) as a linear function of one or more independent variables (predictors) (Montgomery et al., 2021). Its simplicity and interpretability make it a widely used technique across diverse fields, from economics and healthcare to physics and social sciences (Hair et al., 2022).

The fundamental equation of linear regression expresses the predicted value of the dependent variable (y) as a linear combination of the independent variables (x_i) with associated weights (β_i) and an intercept term (α):

$$y = \alpha + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \beta_n \cdot x_n + \epsilon$$

where ϵ represents the error term capturing noise and unexplained variance (Hair et al., 2022).

The goal of linear regression is to estimate these unknown parameters (α , β_i) based on a set of observations (data points) relating the variables. Common estimation methods include ordinary least squares (OLS), which minimizes the sum of squared errors between predicted and actual values (James et al., 2013).

Interpreting the results of a linear regression is crucial. The estimated weights (β_i) quantify the individual impact of each independent variable on the dependent variable, indicating its strength and direction (positive or negative) of association (Hair et al., 2022). Statistical tests assess the significance of these relationships, helping determine which variables truly contribute to the model's predictive power (Montgomery et al., 2021).

Beyond prediction, linear regression can reveal underlying relationships between variables. Examining the model's goodness-of-fit measures, such as the R-squared, helps evaluate how well the linear model captures the data's variability (James et al., 2013). This information is crucial for understanding the limitations of the model and the potential

continued on next page

continued from next page

presence of non-linear trends or other factors influencing the dependent variable.

Despite its versatility, linear regression has limitations. It assumes a linear relationship between the variables, and deviations from this linearity can lead to inaccurate predictions (Hair et al., 2022). Additionally, it is sensitive to outliers and requires assumptions about the error term that may not always hold true (Montgomery et al., 2021).

In conclusion, linear regression remains a fundamental tool for data analysis and modeling. Its ease of use, interpretability, and broad applicability make it a valuable asset across various disciplines. However, understanding its assumptions and limitations is crucial for drawing reliable conclusions and exploring alternative methods for more complex relationships.

Sources:

1. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2022). *Multivariate data analysis* (8th ed.). Pearson Education.
2. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R* (Vol. 112). Springer.
3. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis* (Vol. 5). John Wiley & Sons.

Note: This abstract is approximately 300 words long, staying within the one-page limit. Please let me know if you'd like me to adjust the content or focus on specific aspects of linear regression further.

ChatGPT: Normal distribution

Understanding the Normal Distribution

The normal distribution, also known as the Gaussian distribution, is a fundamental concept in statistics and probability theory. It describes the probability distribution of a continuous random variable that is symmetrically distributed around its mean, forming a bell-shaped curve. This abstract provides an overview of the normal distribution, its characteristics, applications, and significance in various fields.

The normal distribution is characterized by two parameters: the mean (μ) and the standard deviation (σ). The mean determines the central tendency of the distribution, while the standard deviation measures the spread or dispersion of

continued on next page

continued from next page

the data points around the mean. The probability density function (PDF) of the normal distribution is expressed mathematically as:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where x represents the random variable.

The normal distribution exhibits several important properties:

- **Symmetry:** The distribution is symmetric around its mean, with the highest point (mode) at the mean.
- **Bell-shaped curve:** The curve is bell-shaped, with tails that extend infinitely in both directions.
- **Empirical rule:** Approximately 68% of the data falls within one standard deviation of the mean, 95% falls within two standard deviations, and 99.7% falls within three standard deviations.
- **Central limit theorem:** In many cases, the sum or average of a large number of independent and identically distributed random variables tends to follow a normal distribution, regardless of the underlying distribution of the individual variables.

The normal distribution has widespread applications in various fields such as:

- **Statistics:** It serves as a basis for many statistical tests and methods, including hypothesis testing, confidence intervals, and regression analysis.
- **Natural and Social Sciences:** Many natural phenomena and social variables follow a normal distribution, such as heights, weights, IQ scores, and errors in measurements.
- **Finance:** Stock prices, returns, and other financial variables often exhibit a normal distribution, which is utilized in risk management and option pricing models.
- **Quality control:** In manufacturing and process control, the normal distribution is used to analyze variations in product quality and to set quality control limits.

Understanding the normal distribution is essential for interpreting data, making predictions, and making informed decisions in various domains. Its ubiquity and mathematical properties make it a cornerstone of statistical analysis and a valuable tool for researchers, analysts, and practitioners across disciplines.

Sources:

continued on next page

continued from next page

1. Devore, J. L., & Peck, R. (2015). *Statistics: The Exploration & Analysis of Data* (7th ed.). Cengage Learning.
2. Rice, J. A. (2006). *Mathematical Statistics and Data Analysis* (3rd ed.). Cengage Learning.
3. Wackerly, D. D., Mendenhall III, W., & Scheaffer, R. L. (2014). *Mathematical Statistics with Applications* (7th ed.). Cengage Learning.

Gemini: Normal distribution

Unveiling the Bell Curve: An Abstract on the Normal Distribution

The normal distribution, also known as the Gaussian distribution, stands as a cornerstone of statistics and probability. Its characteristic bell-shaped curve embodies the ubiquitousness of randomness and underpins diverse applications across various fields (Larson, 2014). This abstract delves into the essence of the normal distribution, exploring its key features, applications, and enduring significance.

Central to the normal distribution's nature is its symmetry around the mean, indicating that values closer to the average occur more frequently (Moodie, Mackenrodt, & Stephens, 2014). This property is quantified by the mean and standard deviation, which govern the location and spread of the "bell curve," respectively (Larson, 2014). Interestingly, the total area under the curve always equals one, representing the probability of all possible values (Larson, 2014).

One crucial feature of the normal distribution is its universality. Countless natural phenomena, from human heights to exam scores, exhibit an approximate normal distribution (Moodie et al., 2014). This ubiquity stems from the Central Limit Theorem, which states that the sum of many independent random variables often approaches a normal distribution, regardless of the original distributions (Larson, 2014).

The applicability of the normal distribution extends far beyond descriptive statistics. It forms the foundation for hypothesis testing, enabling researchers to draw inferences from samples about entire populations (Moodie et al., 2014). Additionally, it plays a vital role in areas like quality control, where deviations from the expected normal distribution signify potential anomalies (Montgomery, 2021).

continued on next page

continued from next page

Despite its limitations, such as not always perfectly capturing real-world phenomena, the normal distribution remains a powerful tool in the statistical arsenal. Its simplicity, interpretability, and widespread applicability solidify its position as a fundamental concept in numerous scientific disciplines.

Sources:

1. Larson, H. P. (2014). An introduction to probability and statistics. Cengage Learning.
2. Moodie, E. M., Mackenrodt, U., & Stephens, C. (2014). The normal distribution: Usefulness and limitations. *International Journal of Operations Research and Information Systems*, 7(4), 336-346.
3. Montgomery, D. C. (2021). Introduction to statistical quality control. John Wiley & Sons.

ChatGPT: Correlation coefficient

Abstract

Correlation coefficient is a statistical measure that quantifies the degree to which two variables are linearly related. It provides insight into the strength and direction of the relationship between variables, aiding researchers and analysts in understanding patterns and making predictions. This paper explores the concept of correlation coefficient, its calculation methods, interpretation, and practical applications across various fields.

The Pearson correlation coefficient, denoted by r , is the most commonly used measure of correlation. It ranges from -1 to 1, where -1 indicates a perfect negative linear relationship, 0 indicates no linear relationship, and 1 indicates a perfect positive linear relationship between the variables. It is calculated by dividing the covariance of the variables by the product of their standard deviations.

Alternatively, the Spearman rank correlation coefficient, denoted by r_s , is used when variables are not normally distributed or when the relationship is nonlinear. It measures the strength and direction of association between the ranks or ordinal values of the variables.

Correlation coefficient finds extensive application in various fields such as finance, economics, psychology, and medicine. In finance, it helps investors assess the relationship between different assets in their portfolio, aiding in risk management and diversification strategies. In psychology, it is used to study relationships between variables like intelligence and academic performance. In

continued on next page

continued from next page

medicine, correlation coefficient assists in analyzing the association between risk factors and disease outcomes, contributing to disease prevention and treatment strategies.

Understanding the limitations of correlation coefficient is crucial. While it indicates the strength and direction of the linear relationship, it does not imply causation. Additionally, outliers can disproportionately influence the correlation coefficient, leading to potentially misleading interpretations.

In conclusion, correlation coefficient serves as a valuable tool in statistical analysis, providing insights into the relationship between variables in various fields. Its calculation methods, interpretation, and applications make it an essential component of research and decision-making processes.

Sources:

1. Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics*. Pearson Education.
2. Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Routledge.
3. Gujarati, D. N. (2009). *Basic Econometrics*. Tata McGraw-Hill Education.
4. Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications.
5. Myers, J. L., & Well, A. D. (2003). *Research Design and Statistical Analysis*. Lawrence Erlbaum Associates.

Gemini: Correlation coefficient

Unveiling Relationships: A Look at the Correlation Coefficient

In the realm of statistics, discerning patterns and quantifying relationships between variables lies at the heart of understanding complex systems. Among the key tools employed for this purpose is the correlation coefficient, a versatile measure that gauges the strength and direction of linear association between two variables (Hair et al., 2022). This one-page abstract delves into the essence of the correlation coefficient, exploring its interpretation, applications, and limitations.

continued on next page

continued from next page

The most widely used Pearson product-moment correlation coefficient (r) calculates the covariance between two variables, normalized by their standard deviations (Field, 2018). R values range from -1 to 1, with:

- 1 signifying a perfect positive correlation: As one variable increases, the other consistently increases proportionally.
- 0 indicating no linear relationship: Changes in one variable have no predictable impact on the other.
- -1 representing a perfect negative correlation: An increase in one variable is met with a consistent decrease in the other.

Interpreting r necessitates caution. While a high absolute value (closer to 1) suggests a strong correlation, it doesn't imply causation (Fiske et al., 2014). Additional analyses are crucial to establish and understand causal links.

The versatility of the correlation coefficient extends across numerous disciplines. In finance, it helps assess investment risk by measuring the co-movement of asset prices (Campbell et al., 1997). In psychology, it aids in investigating relationships between personality traits and behaviors (Cohen et al., 2003). Even fields like ecology utilize it to explore connections between environmental factors and species abundance (Legendre & Legendre, 1998).

Despite its widespread use, the correlation coefficient has limitations. It is sensitive to outliers, susceptible to non-linear relationships, and only measures linear association (Schoemann, 2012). Therefore, careful data analysis and the choice of appropriate correlation measures are essential.

Understanding the correlation coefficient empowers researchers and analysts to unveil hidden patterns and make informed decisions in various fields. However, interpreting it judiciously and acknowledging its limitations are crucial for drawing accurate conclusions and avoiding pitfalls.

Sources:

1. Campbell, R., Lo, A., & MacKinlay, A. (1997). *The Econometrics of Financial Markets*. Princeton University Press.
2. Cohen, J., Cohen, P., West, S., & Aiken, L. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Routledge.
3. Field, A. P. (2018). *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications.

continued on next page

continued from next page

4. Fiske, J., Kenny, D. A., & West, S. G. (2014). Correlation versus causation: An alternative to null-hypothesis testing. *Social Science Research*, 53(1), 712-725.
5. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2022). *Multivariate data analysis* (9th ed.). Pearson Education.
6. Legendre, P., & Legendre, L. (1998). *Numerical ecology* (2nd ed.). Elsevier.
7. Schoemann, S. (2012). A primer on correlation coefficients. *International Journal of Research and Method in Education*, 35(8), 676-693.

ChatGPT: Random variables

Abstract

Random variables are fundamental concepts in probability theory and statistics, serving as mathematical representations of uncertain quantities. They are widely used in various fields such as finance, engineering, and biology to model and analyze stochastic phenomena. This abstract explores the definition, types, properties, and applications of random variables, drawing from diverse scholarly sources in the field of mathematics and statistics.

Random variables are defined as numerical outcomes of random experiments, mapping the sample space of an experiment to real numbers. They can be classified into two main types: discrete random variables, which take on a countable number of distinct values, and continuous random variables, which can take on any value within a given range. The probability distribution of a random variable describes the likelihood of its possible values, and it can be represented by a probability mass function (PMF) for discrete random variables or a probability density function (PDF) for continuous random variables.

Properties of random variables include moments such as mean, variance, and higher-order moments, which provide measures of central tendency and dispersion. These properties are essential for understanding the behavior and characteristics of random variables in various contexts. For instance, the mean of a random variable represents its average value, while the variance measures its variability or spread around the mean.

Random variables find extensive applications in diverse fields. In finance, they are used to model stock prices, interest rates, and other financial variables, enabling risk assessment and portfolio management. In engineering, random

continued on next page

continued from next page

variables are employed to analyze the reliability and performance of systems subject to uncertain factors such as environmental conditions and manufacturing variations. In biology, they are utilized to model genetic variations and evolutionary processes, aiding in the study of population dynamics and disease spread.

This abstract synthesizes information from authoritative sources such as "Probability and Statistics" by Morris H. DeGroot and Mark J. Schervish, "Introduction to Probability Models" by Sheldon M. Ross, and "Mathematical Statistics with Applications" by Dennis D. Wackerly, William Mendenhall III, and Richard L. Scheaffer. Additionally, scholarly articles from journals like the "Journal of the American Statistical Association" and "Annals of Probability" contribute to the comprehensive understanding of random variables presented in this abstract.

In conclusion, random variables play a crucial role in probabilistic modeling and statistical analysis across various disciplines. Understanding their properties and applications is essential for making informed decisions and drawing meaningful insights from uncertain data.

Keywords: Random Variables, Probability Theory, Statistics, Discrete Random Variables, Continuous Random Variables, Probability Distribution, Mean, Variance, Applications.

Gemini: Random variables

Demystifying Chance: An Introduction to Random Variables

Random variables, the cornerstone of probability and statistics, bridge the gap between unpredictable events and numerical expressions. They transform uncertain outcomes into quantifiable measures, empowering us to analyze real-world phenomena with statistical rigor. This one-page abstract delves into the essence of random variables, exploring their types, applications, and fundamental properties.

At its core, a random variable assigns a numerical value to each outcome of a random experiment (Walpole et al., 2016). Imagine rolling a die: the outcome (a specific number) is uncertain, but we can associate a random variable "X" to it, where X takes values 1 through 6 depending on the rolled number. Random variables can be discrete if they assume a finite or countably infinite number of distinct values, like the die example (Mendenhall & Beaver, 2011). Alternatively, they can be continuous, taking on any value within a specific range, like the measured height of a plant (Larson, 2018).

continued on next page

continued from next page

Understanding the distribution of possible values for a random variable is crucial. This information is captured by the probability distribution, which specifies the likelihood of each value occurring (Ross, 2014). For the die roll, the probability distribution assigns equal probability ($1/6$) to each number. Conversely, the height of a plant might follow a normal distribution, where most plants fall within a certain height range, with fewer deviating significantly.

Random variables empower us to analyze various aspects of data. We can calculate their expected value, akin to the average outcome, to gain a central tendency understanding (Mood et al., 2014). Additionally, the variance measures how spread out the values are, providing insights into data variability (Larson, 2018). These fundamental properties, along with others like standard deviation and moment-generating functions, serve as essential tools for statistical analyses and inference (Walpole et al., 2016).

Beyond basic calculations, random variables play a pivotal role in hypothesis testing, parameter estimation, and modeling real-world phenomena (Larson, 2018). Whether analyzing financial markets, predicting election outcomes, or studying disease outbreaks, random variables serve as the foundation for statistical models that illuminate hidden patterns and inform decision-making.

In conclusion, random variables bridge the gap between uncertainty and quantification, enabling us to analyze and understand the world around us. Their applications span diverse fields, making them indispensable tools for anyone seeking to extract meaningful insights from data.

Sources:

1. Larson, H. P. (2018). Introduction to probability theory and statistical applications. Kendall Hunt.
2. Mendenhall, W., & Beaver, R. J. (2011). Introduction to probability and statistics. Cengage Learning.
3. Mood, A. M., Graybill, F. A., & Boes, D. C. (2014). Introduction to the theory of statistics. Cengage Learning.
4. Ross, S. M. (2014). Introduction to probability models. Academic Press.
5. Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, Z. (2016). Probability & statistics for engineers & scientists. Pearson Education.

continued on next page

continued from next page

Note: This abstract is approximately one page long, excluding the sources and in-text citations. Please adjust the content as needed to fit your specific requirements.

ChatGPT: t-test

Abstract:

The t-test is a fundamental statistical method used to determine if there is a significant difference between the means of two groups. Originating from the work of William Sealy Gosset in 1908, the t-test has since become a cornerstone in various fields including psychology, medicine, economics, and more. This abstract provides an overview of the t-test, its applications, assumptions, and variations.

The t-test is commonly employed when dealing with small sample sizes (typically less than 30) and assumes that the data is normally distributed. It is particularly useful in experimental research settings where researchers aim to compare the means of two groups, such as comparing the effectiveness of two different treatments or interventions. The t-test calculates a t-value, which represents the difference between the means of the two groups relative to the variance within the groups.

There are several variations of the t-test, each suited for different scenarios. The independent samples t-test is used when comparing the means of two independent groups, while the paired samples t-test is used when the same group is measured at two different points in time or under two different conditions. Additionally, the one-sample t-test is utilized when comparing the mean of a sample to a known value or population mean.

Assumptions of the t-test include the independence of observations, normality of data distribution, and homogeneity of variances between groups. Violations of these assumptions can affect the validity of the t-test results.

In conclusion, the t-test is a versatile statistical tool for comparing means between two groups in various research contexts. Its simplicity, applicability to small sample sizes, and ability to provide meaningful insights make it a widely used method in scientific inquiry.

Sources:

1. Gosset, W. S. (1908). The probable error of a mean. *Biometrika*, 6(1),

continued on next page

continued from next page

- 1-25.
2. Field, A. (2009). *Discovering statistics using SPSS*. Sage.
3. Howell, D. C. (2012). *Statistical methods for psychology*. Cengage Learning.
4. Kim, H. Y. (2015). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, 40(4), 328-331.
5. Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: the Bonferroni method. *BMJ: British Medical Journal*, 310(6973), 170.

Gemini: t-test

The T-Test: A Workhorse of Statistical Inference

The t-test, also known as Student's t-test, stands as one of the most fundamental and widely used statistical methods in diverse fields, from psychology and medicine to economics and social sciences (Field, 2009). Its versatility stems from its ability to assess differences between group means under different scenarios, making it crucial for drawing comparisons and testing hypotheses. This abstract delves into the core principles of the t-test, exploring its variations, assumptions, and applications.

Core Concepts and Variations:

At its heart, the t-test assumes normally distributed data and utilizes the t-statistic, a measure of how many standard deviations the observed mean difference between groups falls from a hypothesized value (Cohen et al., 2014). Depending on the research question, two main variants exist:

- **One-sample t-test:** Compares the mean of a single group to a hypothesized value (e.g., testing if exam scores differ significantly from the average of 70).
- **Two-sample t-test:** Compares the means of two independent groups (e.g., evaluating if new teaching methods impact learning outcomes compared to traditional ones).

Further variations include paired t-tests, where data from the same subjects are compared across conditions, and Welch's t-test, suitable for unequal group variances (Field, 2009).

continued on next page

continued from next page

Assumptions and Considerations:

While powerful, the t-test relies on crucial assumptions: normality of data distribution, homogeneity of variances (except for Welch's test), and independence of observations (Cohen et al., 2014). Violation of these assumptions can lead to inaccurate results. Therefore, researchers need to assess and, if necessary, transform data or employ alternative non-parametric tests.

Applications and Significance:

The t-test finds application in numerous areas. In clinical trials, it compares treatment efficacy across groups. In psychology, it assesses differences in cognitive performance or attitudes based on interventions. In economics, it evaluates the impact of policy changes on income levels. Its flexibility caters to a wide range of research questions, making it a cornerstone of statistical analysis.

Sources:

1. Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2014). Applied multiple regression/correlation analysis for the behavioral sciences. Routledge.
2. Field, A. (2009). Discovering statistics using SPSS (3rd ed.). SAGE Publications Ltd.

Limitations:

While providing valuable insights, the t-test has limitations. It is sensitive to outliers and violations of assumptions, and its interpretations require careful consideration of effect sizes and confidence intervals. For more complex comparisons or non-normal data, alternative statistical methods are necessary.

In conclusion, the t-test represents a powerful and versatile tool for researchers across disciplines. Understanding its core concepts, variations, assumptions, and limitations is essential for conducting and interpreting meaningful statistical analyses.

Note: This abstract is approximately one page long and includes in-text citations for the mentioned sources. It provides a basic overview of the t-test; further details and specific applications can be explored based on your specific research area.

ChatGPT: Standard deviation

Abstract: Standard Deviation in Statistics

continued on next page

continued from next page

Standard deviation is a fundamental concept in statistics that measures the dispersion or spread of a dataset around its mean. It provides valuable information about the variability or volatility within a dataset, making it a crucial tool in data analysis and decision-making across various fields such as finance, science, and social sciences.

This abstract aims to provide a concise overview of standard deviation, including its calculation, interpretation, and significance in statistical analysis. Furthermore, it explores its practical applications and limitations in real-world scenarios.

Calculation and Interpretation:

Standard deviation (σ) is calculated as the square root of the variance (σ^2) of a dataset, which is the average of the squared differences between each data point and the mean. Mathematically, it is represented as:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Where N is the number of data points, x_i represents individual data points, and \bar{x} denotes the mean of the dataset.

A higher standard deviation indicates greater variability among data points, whereas a lower standard deviation suggests that data points are closer to the mean, indicating less variability.

Significance in Statistical Analysis:

Standard deviation serves as a measure of risk and uncertainty in various statistical analyses. For instance, in finance, it is commonly used to measure the volatility of asset returns, helping investors assess investment risk. In experimental sciences, standard deviation helps quantify the reliability and consistency of experimental results. Additionally, in quality control processes, standard deviation aids in assessing the consistency and stability of manufacturing processes.

Practical Applications:

Standard deviation finds applications in a wide range of fields, including:

- Finance: Assessing investment risk and volatility.
- Science: Analyzing experimental data and assessing reliability.
- Social Sciences: Measuring variability in survey responses and opinion polls.

continued on next page

continued from next page

- **Quality Control:** Monitoring and improving manufacturing processes by minimizing variability.

Limitations:

While standard deviation is a powerful tool for measuring variability, it has certain limitations. It assumes a normal distribution of data, which may not always be the case in real-world datasets. Additionally, it can be influenced by outliers, skewing its interpretation in skewed distributions.

Conclusion:

In conclusion, standard deviation is a crucial statistical measure that quantifies variability within a dataset, providing valuable insights into the spread of data points around the mean. Its calculation, interpretation, and practical applications make it an indispensable tool in various fields of study and decision-making processes.

Sources and In-text Citations:

1. Introduction to Probability and Statistics by William Mendenhall, Robert J. Beaver, Barbara M. Beaver (2012).
2. Statistics for Business and Economics by Paul Newbold, William L. Carlson, Betty Thorne (2012).
3. An Introduction to Statistical Learning: with Applications in R by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani (2013).
4. Practical Statistics for Data Scientists: 50 Essential Concepts by Andrew Bruce, Peter Bruce (2017).

Gemini: Standard deviation

Demystifying Data Dispersion: A Dive into Standard Deviation

Standard deviation, often denoted by the Greek letter σ , reigns supreme as a fundamental statistic, quantifying the "spread" or "variability" within a dataset (Investopedia, 2024). In simpler terms, it tells us how much individual data points tend to deviate from the average (mean) value (Byju's, 2023). A low standard deviation signifies data clustered tightly around the mean, while a high value indicates wider dispersion (National Library of Medicine, 2023).

Understanding standard deviation unlocks its diverse applications across various fields. In finance, it measures the volatility or risk associated with an

continued on next page

continued from next page

investment, aiding informed decision-making (Investopedia, 2024). In quality control, it helps identify manufacturing inconsistencies, ensuring product consistency (Calculator.net, 2023). Even in education, it assesses test score variability, guiding instructional strategies (Byju's, 2023).

Calculating standard deviation involves a multi-step process. First, calculate the mean of the dataset. Then, for each data point, determine the squared difference from the mean. Finally, average these squared differences and take the square root (Wikipedia, 2023). While seemingly complex, statistical software and online calculators readily perform this calculation.

However, interpreting standard deviation requires context. Consider the data units. A standard deviation of 5 degrees Celsius in temperature data represents less variability than 5 years in age data. Additionally, outliers can significantly skew the standard deviation, necessitating further analysis (National Library of Medicine, 2023).

In conclusion, standard deviation plays a pivotal role in statistical analysis, offering valuable insights into data dispersion. Its applications span diverse fields, from finance and quality control to education and research. While its calculation involves specific steps, readily available tools simplify the process. However, mindful interpretation considering data units and outlier presence is crucial for accurate conclusions.

Sources:

1. Byju's. (2023, January 11). Definition, How to calculate the variance and standard deviation. <https://byjus.com/maths/standard-deviation/>
2. Calculator.net. (2023, November 9). Standard Deviation Calculator. <https://www.calculator.net/standard-deviation-calculator.html>
3. Investopedia. (2024, January 26). Standard Deviation Formula and Uses vs. Variance. <https://www.investopedia.com/terms/s/standarddeviation.asp>
4. National Library of Medicine. (2023, September 26). Standard Deviation. <invalid URL removed>
5. Wikipedia. (2023, October 26). Standard deviation. https://en.wikipedia.org/wiki/Standard_deviation

Note: This abstract is approximately 275 words, falling within the one-page limit. Sources are included both within the text (in-text citations) and in a list at the end.

ChatGPT: Causation

Abstract: Understanding Causation

Causation, a fundamental concept in philosophy, science, and everyday reasoning, encompasses the relationships between events where one event, the cause, brings about another, the effect. This abstract delves into various perspectives on causation, exploring philosophical debates, scientific approaches, and practical implications.

Philosophical Perspectives:

Philosophical inquiries into causation date back to ancient Greek philosophy, with Aristotle's distinction between efficient and final causes laying foundational groundwork. Modern philosophers have expanded on this framework, with David Hume's empiricist critique challenging the notion of necessary connections between cause and effect. Immanuel Kant's transcendental idealism proposed a middle ground, arguing for a synthetic a priori basis for causal connections.

Scientific Approaches:

In the realm of science, causation is central to establishing relationships between variables. The scientific method relies on experimental design to identify causal mechanisms while controlling for confounding factors. However, philosophical debates continue to influence scientific thinking, particularly regarding probabilistic causation and the nature of causation in complex systems.

Practical Implications:

Understanding causation has significant implications across various domains. In medicine, causal inference informs treatment protocols and public health interventions. In law, establishing causation is essential in determining liability and assigning responsibility. Additionally, in fields like economics and sociology, causal models help analyze social phenomena and predict outcomes.

Conclusion:

Causation remains a complex and multifaceted concept, with ongoing debates surrounding its philosophical underpinnings and practical applications. Philosophical inquiries continue to shape scientific methodologies, while practical implications underscore the importance of accurate causal inference in various disciplines.

Sources:

1. Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.

continued on next page

continued from next page

2. Lewis, D. (1986). Causal Explanation. In *Philosophical Papers Volume II*. Oxford University Press.
3. Woodward, J. (2005). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.
4. Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.
5. Hitchcock, C. (2012). Probabilistic Causation. *Stanford Encyclopedia of Philosophy*.
6. Cartwright, N. (2007). *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge University Press.

This abstract provides a concise overview of the multifaceted nature of causation, drawing on both philosophical discussions and practical applications, while referencing key sources in the field.

Gemini: Causation

Causation: Unveiling the Tapestry of Influence

Causation, the intricate relationship between events, fuels our understanding of the world around us. From everyday occurrences to scientific breakthroughs, identifying causes and effects lies at the heart of our inquisitive nature. However, unraveling this tapestry of influence proves remarkably complex, sparking philosophical and scientific debates for centuries. This abstract delves into the multifaceted nature of causation, exploring its key concepts, challenges, and contemporary interpretations.

One fundamental challenge lies in distinguishing true causation from mere correlation. Hume's classic problem of induction highlights the difficulty in inferring causation from observations alone, as events merely occurring together don't necessarily prove one causing the other (Hume, 1748). This issue underscores the need for rigorous methodologies in scientific endeavors, employing controlled experiments and statistical analysis to establish causal relationships (Pearl, 2009).

Beyond the challenge of identification, the very nature of causation itself invites debate. Traditional counterfactual theories posit that an event A causes event B if B wouldn't have occurred without A (Lewis, 1973). However, this simplistic view struggles to explain scenarios with multiple contributing

continued on next page

continued from next page

factors or probabilistic outcomes. Woodward (2003) proposes a probabilistic counterfactual approach, acknowledging that causes often increase the probability of effects without guaranteeing them, allowing for a more nuanced understanding of complex causal relationships.

Further complicating matters, the concept of causation takes on diverse forms across different disciplines. In physics, the focus lies on deterministic laws governing cause-and-effect relationships (Bunge, 1963). In social sciences, however, stochastic processes and individual agency introduce complexities, necessitating probabilistic and multi-level causal models (Hedström & Swedberg, 1998).

Recent advancements in artificial intelligence and machine learning are shedding new light on causation. Causal inference algorithms are being developed to extract causal relationships from vast datasets, aiding in scientific discovery and prediction across diverse fields (Peters et al., 2017). However, concerns regarding interpretability and potential biases in these algorithms highlight the need for careful considerations when relying on machine-driven causal insights.

As we continue to unravel the intricate tapestry of causation, a multifaceted approach embracing diverse perspectives and methodologies proves crucial. Understanding the limitations of traditional models, incorporating advancements in data analysis, and acknowledging the complexities inherent in different domains are essential in navigating this challenging yet fundamental concept. Through ongoing exploration and collaboration, we strive to weave a clearer picture of causation, one that illuminates the interconnectedness of events and empowers our understanding of the world around us.

Sources:

1. Bunge, M. (1963). *Causality: The philosophical aspects of science*. Harvard University Press.
2. Hedström, P., & Swedberg, R. (1998). *Social mechanisms: An analytical framework*. Cambridge University Press.
3. Hume, D. (1748). *An enquiry concerning human understanding*. Oxford University Press.
4. Lewis, D. (1973). Counterfactuals and indicative conditionals. *Philosophical Review*, 81(4), 355-376.
5. Pearl, J. (2009). *Causality: Models, reasoning, and inference*. Springer.

continued on next page

continued from next page

6. Peters, J., Bühlmann, P., & Meinshausen, K. (2017). Causal inference using directed acyclic graphs. *International Encyclopedia of Statistical Science*.

Note: This abstract is approximately one page long, excluding the list of sources.

ChatGPT: Time series analysis

Abstract: Time Series Analysis

Time series analysis is a statistical method used to analyze sequential data points measured over time intervals. It has diverse applications across various fields including finance, economics, meteorology, and engineering. This abstract provides an overview of time series analysis techniques, their applications, and recent advancements.

The primary objective of time series analysis is to understand the underlying patterns, trends, and behaviors within the data, enabling forecast and prediction of future values. It involves examining autocorrelation, seasonality, and trends within the data, along with identifying and modeling any underlying stochastic processes.

One of the fundamental techniques in time series analysis is autoregression (AR), where a time series is modeled as a linear combination of its past values. This method is widely used in economic forecasting and financial markets analysis (Hamilton, 1994). Another important concept is moving average (MA), which involves taking the average of a series of data points within a defined window. Combining AR and MA models gives rise to the autoregressive integrated moving average (ARIMA) model, which is extensively used in various fields for forecasting purposes (Box et al., 1994).

Recent advancements in time series analysis include the development of machine learning algorithms such as Long Short-Term Memory (LSTM) networks and Recurrent Neural Networks (RNNs). These deep learning techniques have shown promising results in handling complex temporal dependencies and have found applications in areas like natural language processing and speech recognition (Hochreiter & Schmidhuber, 1997).

Moreover, the emergence of big data and computational power has facilitated the analysis of massive time series datasets, leading to advancements in techniques like wavelet analysis and spectral analysis for time-frequency domain representation (Percival & Walden, 2000).

continued on next page

continued from next page

In conclusion, time series analysis is a vital tool for understanding and forecasting sequential data. With the integration of traditional statistical methods and modern machine learning techniques, researchers and practitioners can gain deeper insights and make more accurate predictions in various domains.

References:

1. Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
2. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (1994). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.
3. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780.
4. Percival, D. B., & Walden, A. T. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge University Press.

Gemini: Time series analysis

Unraveling the Flow of Time: An Introduction to Time Series Analysis

Time series analysis delves into the captivating realm of data collected over time, unveiling patterns and extracting valuable insights. From charting daily stock prices to predicting weather patterns, this versatile field holds immense potential across diverse domains (Chatfield, 2003). At its core, time series analysis involves a sequence of data points indexed chronologically, often collected at consistent intervals (Shumway & Stoffer, 2011).

The primary goals of this analysis are threefold: understanding the internal structure of the data, uncovering trends and seasonal variations, and ultimately, forecasting future values (Hamilton, 1994). To achieve these objectives, a plethora of techniques are employed, each catering to specific data characteristics and objectives.

Delving into the Depths: Common Techniques

One prominent approach is exponential smoothing, which assigns higher weights to recent data points, capturing trends and short-term fluctuations effectively (Hyndman & Athanasopoulos, 2018). For data exhibiting seasonal patterns, ARIMA models come into play, incorporating autoregressive and moving average components to model and forecast future values (Box et al., 2015). In the realm of complex data with non-linear relationships, machine

continued on next page

continued from next page

learning algorithms like recurrent neural networks offer promising avenues for uncovering hidden patterns and making accurate predictions (Bergmeir & Schmid, 2019).

Applications Spanning Industries

The impact of time series analysis spans far and wide, influencing critical decisions across various industries. In finance, it forms the bedrock for stock price prediction and risk management (Campbell et al., 1997). In healthcare, it aids in disease outbreak forecasting and patient monitoring (Konstatinov & Thissen, 2011). The field also underpins weather forecasting, enabling preparedness for extreme weather events (Smith, 2000).

Looking Ahead: Future Directions

As data collection becomes increasingly ubiquitous and sophisticated, the future of time series analysis is brimming with possibilities. The integration of deep learning algorithms with domain-specific knowledge holds immense promise for uncovering intricate patterns and achieving superior forecasting accuracy (Kang et al., 2017). Furthermore, the development of interpretable models that shed light on the "why" behind predictions is crucial for building trust and ensuring responsible application of these powerful tools (Rudin, 2020).

In conclusion, time series analysis serves as a potent lens for examining the dynamic tapestry of time-based data. By unraveling underlying patterns and trends, it empowers informed decision-making across diverse fields, shaping a future where the flow of time holds less secrets and reveals more opportunities for progress.

Sources:

1. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). Time series analysis: Forecasting and control. John Wiley & Sons.
2. Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). The Econometrics of Financial Markets. Princeton University Press.
3. Chatfield, C. (2003). The analysis of time series: An introduction. CRC press.
4. Hamilton, J. D. (1994). Time series analysis. Princeton University Press.
5. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.

continued on next page

continued from next page

6. Kang, Y., Hong, X., & Sohn, K. A. (2017). Deep learning for anomaly detection in time series data. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 1116-1125). IEEE.
7. Konstantinov, A., & Thissen, U. (2011). Time series forecasting in healthcare: a survey. *Health Inform J*, 17(2), 119-132.
8. Rudin, C. (2020). Stop explaining black box machine learning models for high stakes decisions. *Nature Machine Intelligence*, 2(11), 749-758.
9. Shumway, R. H., & Stoffer, D. S. (2011). *Time series analysis and its applications: with R examples*. Springer.
10. Smith, L. A. (2000). *Forecast verification: a practitioner's guide in atmospheric science*. John Wiley & Sons.

ChatGPT: Variance

Abstract: Variance in Statistical Analysis

Variance is a fundamental concept in statistics, widely used to measure the dispersion or spread of a set of data points around their mean or expected value. It provides essential insights into the variability within a dataset, making it a crucial tool in various fields, including economics, psychology, and natural sciences.

The concept of variance was first introduced by the British statistician Karl Pearson in the late 19th century. Pearson's work laid the groundwork for understanding the importance of variability in statistical analysis. Variance is calculated by averaging the squared differences between each data point and the mean of the dataset, thereby quantifying the extent to which individual observations deviate from the average.

In modern statistical analysis, variance serves as a key component in several important calculations, such as standard deviation, which is the square root of the variance. Standard deviation is widely used in inferential statistics and plays a crucial role in hypothesis testing and confidence interval estimation.

Furthermore, variance plays a significant role in the analysis of experimental designs, where it is used to assess the degree of variability between different treatment groups or conditions. For example, in the field of medical research, variance analysis is often employed to evaluate the effectiveness of new drugs or treatments by comparing the variability in outcomes among different patient groups.

continued on next page

continued from next page

Moreover, variance has practical applications in quality control and process improvement. In manufacturing and industrial settings, variance analysis is used to monitor and minimize variability in production processes, thereby ensuring consistency and reliability in product quality.

Despite its widespread utility, variance has its limitations and assumptions. For instance, it assumes that the data are drawn from a population with a normal distribution, which may not always be the case in real-world scenarios. Additionally, variance is sensitive to outliers, which can disproportionately influence its value and affect the interpretation of results.

In conclusion, variance is a fundamental concept in statistical analysis, providing valuable insights into the variability of data and serving as a cornerstone for various statistical calculations and applications. Its understanding is essential for researchers, analysts, and practitioners across diverse fields to make informed decisions based on data-driven evidence.

Sources:

1. Pearson, K. (1894). Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London. Series A*, 185, 343-414.
2. Kenney, J. F., & Keeping, E. S. (1951). *Mathematics of Statistics, Part One*. Van Nostrand.
3. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. John Wiley & Sons.
4. Box, G. E. P., Hunter, J. S., & Hunter, W. G. (2005). *Statistics for Experimenters: Design, Innovation, and Discovery (2nd ed.)*. John Wiley & Sons.
5. Montgomery, D. C. (2017). *Design and Analysis of Experiments (9th ed.)*. John Wiley & Sons.

Gemini: Variance B

Variance: Quantifying Data Spread

Variance is a fundamental concept in statistics, offering a quantitative measure of how spread out a dataset is in relation to its mean (average) value ([Everitt & Hothorn, 2011]). A low variance indicates that data points cluster tightly around

continued on next page

continued from next page

the mean, while a high variance signifies a wider dispersion. Mathematically, it represents the average squared deviation of each data point from the mean ([Chatterjee & Hadi, 2012]).

Understanding variance is crucial in various fields, including:

- Descriptive statistics: It provides a summary measure of data variability alongside the mean, offering a more complete picture of the data's distribution ([Gupta, 2015]).
- Hypothesis testing: Comparing variances across groups helps assess the likelihood of their means being different by chance ([Field, 2013]).
- Statistical modeling: Estimating and analyzing variance components allows for the exploration of factors contributing to data variability ([Gelman & Hill, 2007]).

Variance finds application in diverse contexts, such as:

- Financial markets: Analyzing the variance of stock prices helps assess investment risk ([Hull, 2010]).
- Psychology: Studying the variance in test scores can reveal individual differences and potential group effects ([Ferguson, 2015]).
- Quality control: Monitoring the variance of product measurements helps ensure consistency and identify potential defects ([Montgomery, 2012]).

In conclusion, variance serves as a powerful tool for quantifying data spread and plays a crucial role in various statistical analyses and applications across different fields.

Sources:

1. Everitt, B. S., & Hothorn, T. (2011). *A Handbook of Statistical Analyses using R*. Chapman and Hall/CRC.
2. Chatterjee, S., & Hadi, A. S. (2012). *Regression Analysis by Example*. Wiley.
3. Gupta, A. K. (2015). *Probability and Statistics*. Springer.
4. Field, A. P. (2013). *Discovering Statistics Using R*. Sage Publications.
5. Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

continued on next page

continued from next page

6. Hull, J. C. (2010). *Options, Futures, and Other Derivatives*. Pearson Education.
7. Ferguson, C. J. (2015). *Statistical Applications in Psychology and Education*. Routledge.
8. Montgomery, D. C. (2012). *Introduction to Statistical Quality Control*. John Wiley & Sons.

ChatGPT: Median absolute deviation

Title: Exploring the Median Absolute Deviation: A Robust Measure of Dispersion

Abstract:

The median absolute deviation (MAD) serves as a robust alternative to the standard deviation, particularly in scenarios where data exhibit non-normal distributions or contain outliers. This paper delves into the concept, computation, and applications of MAD as a measure of dispersion.

The MAD is calculated as the median of the absolute deviations of individual data points from the overall median of the dataset. Unlike the standard deviation, which is sensitive to extreme values, MAD focuses on the central tendency of the data, making it less influenced by outliers.

This study reviews various computational methods for MAD, including the traditional formula and the efficient median-of-medians algorithm. Additionally, it examines the relationship between MAD and other measures of dispersion, such as the interquartile range and the standard deviation.

Furthermore, the paper explores the practical applications of MAD across diverse fields. In finance, MAD is used to assess the volatility of financial assets, providing a more robust estimate compared to standard deviation in the presence of outliers. In environmental science, MAD is employed to analyze data with non-normal distributions, offering reliable estimates of variability in pollutant concentrations. Moreover, in healthcare, MAD aids in detecting outliers in patient data, contributing to more accurate assessments of treatment efficacy and patient outcomes.

This paper synthesizes findings from various scholarly sources, including statistical textbooks, research articles, and academic journals. Notable

continued on next page

continued from next page

contributions include Leys et al. (2013) who discuss the advantages of MAD over standard deviation in analyzing skewed data, and Rousseeuw and Croux (1993) who propose robust estimators of MAD using resampling techniques.

In conclusion, the median absolute deviation emerges as a robust and versatile measure of dispersion, particularly valuable in analyzing data with non-normal distributions or outliers. Its applications span across disciplines, offering researchers and practitioners a reliable tool for data analysis and interpretation.

Sources:

1. Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764-766.
2. Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88(424), 1273-1283.

Gemini: Median absolute deviation

Median Absolute Deviation: A Robust Measure of Spread

The median absolute deviation (MAD) is a robust measure of statistical dispersion, quantifying how spread out a set of data is [2]. Unlike the standard deviation, which is heavily influenced by extreme values (outliers), the MAD offers a more resilient alternative in the presence of skewed data or outliers [3].

Calculating the MAD:

Calculating the MAD involves finding the absolute deviations of each data point from the median, taking the median of these absolute deviations [2]. Mathematically, for a sample of data points $\{x_1, x_2, \dots, x_n\}$, the MAD is calculated as:

$$\text{MAD} = \text{median}(|x_1 - \text{median}(x)|, |x_2 - \text{median}(x)|, \dots, |x_n - \text{median}(x)|)$$

Advantages of MAD:

The primary advantage of MAD is its robustness to outliers. Due to its reliance on the median, the MAD is less affected by extreme values compared to the standard deviation, making it a valuable tool for analyzing data potentially contaminated by outliers [1, 3].

continued on next page

continued from next page

Applications of MAD:

The MAD finds applications in various fields, including:

- Descriptive statistics: Summarizing the spread of data alongside measures like the median [2].
- Boxplots: Visualizing the distribution of data, where the MAD is used to define the upper and lower quartile boundaries (the "whiskers") of the boxplot [1].
- Exploratory data analysis: Identifying potential outliers and assessing the normality of data [3].

Conclusion:

The MAD serves as a valuable tool for understanding the spread of data, particularly when dealing with skewed distributions or the presence of outliers. Its resistance to extreme values offers a robust alternative to the standard deviation, making it a preferred measure of dispersion in various statistical analyses.

Sources:

1. Wikipedia - Median absolute deviation: https://en.wikipedia.org/wiki/Median_absolute_deviation
2. Statistics How To - Median Absolute Deviation: [invalid URL removed]
3. YouTube - What is the Median Absolute Deviation (MAD)?: [invalid URL removed]

ChatGPT: Skewness

Abstract

Skewness, a fundamental concept in statistics, measures the asymmetry of the probability distribution of a real-valued random variable about its mean. In this abstract, we explore the definition, significance, calculation methods, and interpretation of skewness. We also discuss its applications in various fields such as finance, economics, and biology.

Skewness is crucial in understanding the shape of a distribution and provides valuable insights into the underlying data characteristics. It complements other measures of central tendency and dispersion such as mean, median, and standard deviation, offering a more comprehensive description of the dataset's

continued on next page

continued from next page

behavior. Positive skewness indicates a longer right tail, while negative skewness suggests a longer left tail relative to the central peak of the distribution.

The calculation of skewness can be performed using different formulas, including the Pearson's moment coefficient of skewness, Bowley's coefficient of skewness, and the Fisher-Pearson standardized moment coefficient of skewness. Each formula has its advantages and is suitable for specific types of data distributions.

Skewness finds extensive application in various disciplines. In finance, it helps in understanding the distribution of returns on investment portfolios, aiding in risk assessment and portfolio optimization strategies (Jegadeesh, 1990). In economics, skewness plays a vital role in analyzing income distributions and consumer behavior (Atkinson, 1970). Furthermore, in biology, skewness assists in studying asymmetric traits in populations, contributing to evolutionary biology research (Merilä and Björklund, 1995).

Understanding skewness is essential for researchers and practitioners across different domains as it provides valuable information about the nature of data distributions, facilitating better decision-making processes and deeper insights into the phenomena under investigation.

References

1. Jegadeesh, N. (1990). Evidence of Predictable Behavior of Security Returns. *The Journal of Finance*, 45(3), 881–898.
2. Atkinson, A. B. (1970). On the Measurement of Inequality. *Journal of Economic Theory*, 2(3), 244–263.
3. Merilä, J., & Björklund, M. (1995). Fluctuating Asymmetry and Measurement Error. *Journal of Evolutionary Biology*, 8(5), 531–539.

Gemini: Skewness

Skewness: Unveiling Asymmetry in Data

Skewness, a crucial concept in statistics and probability, quantifies the asymmetry present in a data distribution ([Scribbr, 2023]). It measures how the data deviates from perfect symmetry, often visualized by the bell-shaped normal distribution ([Investopedia, 2023]). In simpler terms, skewness tells us whether the data is tilted to one side, indicating a longer tail on either the left or

continued on next page

continued from next page

right end of the distribution.

Understanding skewness is essential because many statistical methods, like calculating the mean and standard deviation, assume symmetrical data. When data is skewed, these methods can be misleading, leading to inaccurate interpretations and erroneous conclusions ([Corporate Finance Institute, 2023]).

There are three main types of skewness:

- **Right skewness (positive skewness):** The distribution has a longer tail extending towards higher values and is pulled to the left. This is often seen in income data, where a few individuals have significantly higher incomes than the majority ([Scribbr, 2023]).
- **Left skewness (negative skewness):** The distribution has a longer tail extending towards lower values and is pulled to the right. An example is the distribution of test scores in a class where most students score well, but a few score significantly lower ([Investopedia, 2023]).
- **Zero skewness:** The distribution is perfectly symmetrical, resembling the normal distribution. This is uncommon in real-world data, but it can occur in theoretical scenarios or data sets specifically designed to be symmetrical.

Several methods exist to calculate skewness, with the most common being the Pearson's skewness coefficient. This coefficient provides a numerical value, typically interpreted as follows:

- **Positive value:** Indicates right skewness.
- **Negative value:** Indicates left skewness.
- **Value close to zero:** Suggests little to no skewness (near symmetry).

Understanding and addressing skewness is crucial for accurate data analysis and interpretation. By acknowledging and potentially adjusting for skewness, researchers and analysts can make more informed decisions and draw more reliable conclusions from their data.

Sources:

1. Corporate Finance Institute. (2023, January 25). Skewness - Overview, Types, How to Measure and Interpret <https://www.sciencedirect.com/science/article/abs/pii/S0378426698001198>

continued on next page

continued from next page

2. Investopedia. (2023, November 10). Positively and Negatively Skewed Defined With Formula. <https://www.investopedia.com/terms/s/skewness.asp>
3. Scribbr. (2023, May 10). Skewness | Definition, Examples & Formula. <https://www.scribbr.com/plagiarism-checker/guide/>

ChatGPT: Kurtosis

Abstract: Understanding Kurtosis in Statistics

Kurtosis, a statistical measure, provides valuable insights into the shape and distribution of data. This abstract explores the concept of kurtosis, its significance in statistical analysis, various types, and its applications across diverse fields. The purpose is to offer a comprehensive overview of kurtosis, its calculation methods, interpretation, and practical implications.

Kurtosis measures the degree of peakedness or flatness of a probability distribution compared to a normal distribution. It quantifies the tails of a distribution and determines whether the data is heavy-tailed or light-tailed. A distribution with high kurtosis has heavier tails and a more peaked shape, while low kurtosis indicates lighter tails and a flatter shape.

In statistical analysis, kurtosis plays a crucial role in assessing the deviation of a dataset from normality. It helps identify outliers, evaluate risk in financial models, and understand the behavior of variables in various scientific studies. For instance, in finance, understanding the kurtosis of stock returns aids in risk management and portfolio optimization.

There are different types of kurtosis measures, including Pearson kurtosis, Fisher's kurtosis, and excess kurtosis. Pearson kurtosis is based on moments and is suitable for symmetric distributions, while Fisher's kurtosis focuses on the tails of the distribution. Excess kurtosis, the most commonly used measure, subtracts 3 from Pearson or Fisher kurtosis to adjust for the normal distribution's kurtosis.

Calculating kurtosis involves higher moments of the distribution and is influenced by the sample size. In practice, software tools like R, Python's SciPy library, and statistical software packages facilitate kurtosis calculation for datasets of varying sizes and complexities.

continued on next page

continued from next page

Interpreting kurtosis involves comparing it to the kurtosis of a normal distribution. A kurtosis value of 0 indicates a normal distribution, while positive (leptokurtic) or negative (platykurtic) values represent deviations from normality. Extreme values of kurtosis indicate potential issues like data clustering, outliers, or fat-tailed distributions.

In conclusion, kurtosis is a valuable statistical measure with wide-ranging applications in finance, economics, biology, and other disciplines. Understanding kurtosis aids researchers and analysts in interpreting data distributions, identifying outliers, and making informed decisions based on the underlying data characteristics.

Sources:

1. Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). "Continuous univariate distributions." Vol. 1. Wiley Series in Probability and Statistics.
2. Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2014). "Mathematical statistics with applications." Cengage Learning.
3. Hill, T., & Lewicki, P. (2007). "Statistics: Methods and applications: A comprehensive reference for science, industry, and data mining." StatSoft.
4. Westfall, P. H., & Henning, K. S. (2013). "Understanding advanced statistical methods." CRC Press.
5. Kuhn, M., & Johnson, K. (2013). "Applied predictive modeling." Springer Science & Business Media.
6. Wilcox, R. R. (2017). "Introduction to robust estimation and hypothesis testing." Academic Press.
7. Ghasemi, A., & Zahediasl, S. (2012). "Normality tests for statistical analysis: a guide for non-statisticians." *International journal of endocrinology and metabolism*, 10(2), 486-489.

Gemini: Kurtosis

Kurtosis: A Measure of Tail Weight in Statistical Distributions

Kurtosis is a statistical measure that characterizes the heaviness of the tails of a distribution relative to a normal distribution. In simpler terms, it describes how much a distribution's tails are thicker or thinner compared to a normal bell-shaped curve [Everitt & Hancock, 2013].

continued on next page

continued from next page

A distribution with positive kurtosis is leptokurtic, meaning it has heavier tails than a normal distribution. This indicates that the distribution is more likely to have extreme values in the tails compared to a normal distribution. Conversely, a distribution with negative kurtosis is platykurtic, meaning it has thinner tails than a normal distribution. This suggests that the distribution is less likely to have extreme values in the tails compared to a normal distribution [Field, 2013].

Calculating kurtosis involves finding the fourth moment of a distribution, which is the average of the fourth power of the deviations from the mean. However, directly using the fourth moment can be sensitive to outliers, so excess kurtosis is often used instead. Excess kurtosis is the fourth moment divided by the square of the variance, effectively removing the influence of the scale of the data [Westfall, 2014].

Understanding kurtosis is crucial in various domains, including:

- Finance: Investors use kurtosis to assess the risk associated with investments. A leptokurtic return distribution indicates a higher probability of significant losses or gains compared to a normal distribution.
- Machine learning: Identifying outliers and assessing the normality of data are crucial steps in many machine learning algorithms. Kurtosis helps identify deviations from normality, which can be important for choosing appropriate algorithms and interpreting results.
- Science and engineering: Researchers often analyze data with non-normal distributions. Kurtosis provides valuable insights into the characteristics of such data, aiding in data interpretation and model building.

In conclusion, kurtosis is a valuable statistical measure that quantifies the tail weight of a distribution relative to a normal distribution. Understanding kurtosis is essential in various fields, from finance and machine learning to science and engineering, allowing for informed decision-making and accurate data analysis.

Sources:

1. Everitt, B. S., & Hancock, G. R. (2013). Handbook of statistical methods. John Wiley & Sons.
2. Field, A. (2013). Discovering statistics using R. SAGE Publications Limited.
3. Westfall, P. H. (2014). Kurtosis and its estimation. John Wiley & Sons.

ChatGPT: Cauchy distribution

Abstract: Understanding the Cauchy Distribution

The Cauchy distribution, named after the French mathematician Augustin-Louis Cauchy, is a continuous probability distribution that plays a significant role in various fields of science and engineering. This abstract aims to provide a concise overview of the Cauchy distribution, its properties, applications, and significance.

Properties:

The Cauchy distribution is characterized by its probability density function $f(x) = \frac{1}{\pi(1+x^2)}$, where x is a real number, and its cumulative distribution function $F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan(x)$. Notably, the Cauchy distribution lacks finite moments due to its heavy tails, making its mean and variance undefined.

Significance:

Despite its infinite variance, the Cauchy distribution finds applications in various fields such as physics, finance, and signal processing. In physics, it describes resonance phenomena in particle physics and spectral line broadening in spectroscopy. In finance, it models extreme events and is used in risk management. In signal processing, it models noise in communication channels and image processing.

Applications:

One notable application of the Cauchy distribution is in robust statistics, where it serves as a basis for robust estimation techniques. Robust estimators based on the Cauchy distribution, such as the Cauchy median estimator, are less sensitive to outliers compared to traditional estimators like the sample mean.

Sources and In-text Citations:

1. Wasserman, L. (2013). *All of Statistics: A Concise Course in Statistical Inference*. Springer Science & Business Media. (Source for properties and significance)
2. García, C. B. (2007). *Cauchy Distribution*. *Encyclopedia of Statistical Sciences*. (Source for applications)
3. Huber, P. J. (1981). *Robust Statistics*. *Wiley Series in Probability and Mathematical Statistics*. (Source for robust estimation techniques)

In conclusion, the Cauchy distribution, despite its unique properties, holds significant importance in various fields and serves as a fundamental concept

continued on next page

continued from next page

in statistical theory and applications, particularly in scenarios involving heavy-tailed data and robust estimation techniques.

Word Count: 248 words

Gemini: Cauchy distribution

The Cauchy Distribution: A Heavy-Tailed Enigma

The Cauchy distribution, also known as the Lorentz distribution, is a continuous probability distribution characterized by its distinctive shape and unique properties. While it shares similarities with the familiar normal distribution, the Cauchy distribution exhibits "heavy tails," meaning its probability density function (PDF) decays slower than the normal distribution at the extremes, leading to a higher likelihood of observing extreme values [1]. This characteristic makes the Cauchy distribution a valuable tool in various fields, but also presents certain challenges in its application.

Formally, the Cauchy distribution is defined by two parameters: the location parameter (t) and the scale parameter (s) [2]. The location parameter dictates the center of the distribution, while the scale parameter controls the width of the "peak" at its center [3]. The PDF of the Cauchy distribution is given by:

$$f(x) = 1/[\pi s(1 + ((x - t)/s)^2)]$$

One of the defining features of the Cauchy distribution is the absence of a well-defined mean and variance. Unlike the normal distribution where the mean and variance completely characterize the distribution, the Cauchy distribution's moments (mean, variance, etc.) are undefined [1, 4]. This "pathological" behavior stems from the heavy tails of the distribution, making it unsuitable for tasks like calculating the average value or the spread of data it represents.

Despite the lack of defined moments, the Cauchy distribution finds applications in various domains. It appears in signal processing, where it models the Lorentzian line shape in spectral analysis [5]. Additionally, the Cauchy distribution finds use in economics to model heavy-tailed phenomena like financial returns [6].

However, the lack of a well-defined mean and the presence of heavy tails necessitate careful consideration when employing the Cauchy distribution. Statistical techniques like median and median absolute deviation (MAD) are often preferred over mean and standard deviation for summarizing data

continued on next page

continued from next page

following a Cauchy distribution [1].

In conclusion, the Cauchy distribution stands out as a unique and valuable probability distribution with its distinctive shape and heavy tails. While its "pathological" nature presents challenges in certain applications, its characteristic features make it relevant in diverse fields like signal processing and economics. Understanding the properties and limitations of the Cauchy distribution is crucial for its effective application in various scientific and engineering disciplines.

Sources:

1. NIST/SEMATECH e-Handbook of Statistical Methods:
<https://www.itl.nist.gov/div898/handbook/eda/section3/eda3663.htm>
2. Wikipedia - Cauchy distribution:
https://en.wikipedia.org/wiki/Cauchy_distribution
3. MathWorld - Cauchy Distribution:
<https://mathworld.wolfram.com/CauchyDistribution.html>
4. Nagoya University - Cauchy Distribution: http://www.math.nagoya-u.ac.jp/richard/teaching/s2019/Cauchy_Distribution.pdf
5. Statistics How To - Cauchy Distribution:
<https://www.statisticshowto.com/cauchy-distribution/>
6. Chapman, M., & Payne, R. (2009). The econometrics of financial markets. Cambridge University Press.

ChatGPT: Nonlinear dependence

Nonlinear Dependence in Complex Systems: Unraveling Sources and Implications

The study of nonlinear dependence in complex systems is pivotal in various disciplines ranging from physics to economics and beyond. Nonlinear dependencies characterize relationships where changes in one variable do not proportionally affect another, often leading to emergent behaviors and patterns. This abstract explores the sources and implications of nonlinear dependence in complex systems through a multidisciplinary lens.

Sources of Nonlinear Dependence:

continued on next page

continued from next page

Nonlinear dependencies arise from various sources, including but not limited to feedback loops, threshold effects, and network structures. Feedback loops introduce nonlinearities by creating self-reinforcing or self-regulating mechanisms within a system, leading to non-proportional responses to stimuli. Threshold effects manifest when a system transitions between qualitatively different states based on the magnitude of inputs, resulting in discontinuous changes and hysteresis phenomena. Network structures, prevalent in biological, social, and technological systems, introduce nonlinear interactions among interconnected components, giving rise to emergent behaviors and complex dynamics.

Implications of Nonlinear Dependence:

Understanding nonlinear dependence has far-reaching implications across disciplines. In ecological systems, nonlinear dependencies drive ecosystem resilience and stability, with cascading effects following perturbations. In economics, nonlinear dependencies underpin phenomena such as market bubbles and crashes, where small fluctuations can lead to disproportionate outcomes. In neuroscience, nonlinear dependencies govern neural dynamics and information processing, shaping cognitive functions and behaviors. Additionally, nonlinear dependencies pose challenges for predictive modeling and decision-making, highlighting the need for sophisticated analytical approaches capable of capturing complex dynamics.

Conclusion:

Nonlinear dependence is a fundamental characteristic of complex systems, with diverse sources and profound implications across disciplines. Embracing nonlinear dynamics is crucial for advancing our understanding of complex phenomena and developing strategies to address real-world challenges. By elucidating the sources and implications of nonlinear dependence, this abstract provides a foundational framework for further research and interdisciplinary collaboration in the study of complex systems.

References:

1. Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems*. MIT Press.
2. Sornette, D. (2003). *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton University Press.
3. Strogatz, S. H. (2014). *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Westview Press.

continued on next page

continued from next page

4. May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, 261(5560), 459-467.
5. Barabási, A. L. (2016). *Network Science*. Cambridge University Press.

Gemini: Nonlinear dependence

Nonlinear Dependence: Unveiling Complex Relationships

Nonlinear dependence refers to a statistical phenomenon where changes in one variable do not lead to directly proportional changes in another. Unlike linear relationships, which can be represented by a straight line, nonlinear dependencies exhibit more intricate patterns. Understanding these patterns is crucial in various fields, including physics, economics, and biology.

One way to conceptualize nonlinear dependence is through the absence of a constant rate of change. For instance, imagine studying the relationship between temperature and the growth rate of plants. As temperature increases, the growth rate might initially accelerate, then plateau, and even decline at extremely high temperatures. This illustrates a nonmonotonic relationship, where the direction of change (positive or negative) in the dependent variable (growth rate) is not constant with respect to the independent variable (temperature).

Several statistical techniques can be employed to identify and analyze nonlinear dependencies. Nonparametric methods, like Spearman's rank correlation coefficient and Kendall's rank correlation coefficient, assess the strength and direction of association between variables without assuming a specific form for the relationship [3]. Conversely, parametric methods, such as generalized autoregressive conditional heteroscedasticity (GARCH) models, can capture specific types of nonlinearity, such as the changing volatility of financial markets [1].

Nonlinear dependence often unveils complex underlying mechanisms. In ecology, for example, studying the nonlinear relationship between predator and prey populations can provide insights into population dynamics and ecosystem stability [2]. Similarly, in economics, analyzing the nonlinear relationship between interest rates and economic growth can inform policymakers' decisions [4].

By delving into the complexities of nonlinear dependence, researchers can gain a deeper understanding of the intricate relationships between variables in various disciplines. This knowledge proves invaluable in making informed

continued on next page

continued from next page

predictions, developing effective interventions, and ultimately, fostering progress across diverse fields.

Sources:

1. ARCH and Bilinearity as Competing Models for Nonlinear Dependence [invalid URL removed]
2. Efficient test for nonlinear dependence of two continuous variables <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-015-0697-7>
3. Non-linear Dependence? Mutual Information to the Rescue! <https://oxfordre.com/publichealth/display/10.1093/acrefore/9780190632366.001.0001/acrefore-9780190632366-e-248>
4. Nonlinear system theory: Another look at dependence <https://www.pnas.org/doi/10.1073/pnas.0506715102>

ChatGPT: Derivative

Title: Exploring the World of Derivatives: A Comprehensive Overview

Abstract:

Derivatives have become integral tools in modern finance, playing a pivotal role in risk management, speculation, and hedging strategies. This paper provides a comprehensive overview of derivatives, encompassing their types, pricing models, applications, and associated risks.

The study begins by delineating the foundational concepts of derivatives, elucidating their definition as financial contracts whose value derives from an underlying asset. It further examines the various types of derivatives, including forwards, futures, options, and swaps, highlighting their distinctive features and applications in different market contexts.

Furthermore, this paper delves into the methodologies employed in pricing derivatives, discussing classical approaches such as the Black-Scholes model and more recent advancements like the Monte Carlo simulation method. Through empirical analysis and real-world examples, it evaluates the efficacy of these models in valuing derivatives accurately, considering factors such as volatility, interest rates, and time to maturity.

continued on next page

continued from next page

In addition, the research explores the multifaceted applications of derivatives in financial markets, ranging from risk mitigation to speculative trading strategies. It examines how derivatives facilitate risk management by providing mechanisms for investors to hedge against adverse price movements and diversify their portfolios effectively. Furthermore, it elucidates how derivatives enable market participants to capitalize on price fluctuations and leverage their investment positions through speculation.

Moreover, this paper scrutinizes the inherent risks associated with derivatives, emphasizing their potential for amplifying market volatility and systemic risks. It evaluates the regulatory frameworks governing derivatives markets globally and assesses their effectiveness in safeguarding market stability and investor interests.

Finally, drawing upon a synthesis of empirical evidence and theoretical frameworks, this paper offers insights into the future prospects of derivatives markets, anticipating trends such as increased innovation in derivative products, heightened regulatory scrutiny, and evolving risk management practices.

Through a comprehensive analysis of derivatives, this paper aims to provide practitioners, academics, and policymakers with a nuanced understanding of these financial instruments, elucidating their role in contemporary financial markets and their implications for market dynamics and stability.

Sources:

1. Hull, John C. *Options, Futures, and Other Derivatives*. Pearson, 2017.
2. McDonald, Robert L. *Derivatives Markets*. Pearson Education, 2013.
3. Wilmott, Paul. *Paul Wilmott Introduces Quantitative Finance*. John Wiley & Sons, 2007.
4. Black, Fischer, and Myron Scholes. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy*, vol. 81, no. 3, 1973, pp. 637–654.
5. Cox, John C., et al. "Option Pricing: A Simplified Approach." *Journal of Financial Economics*, vol. 7, no. 3, 1979, pp. 229–263.
6. Merton, Robert C. "Theory of Rational Option Pricing." *The Bell Journal of Economics and Management Science*, vol. 4, no. 1, 1973, pp. 141–183.
7. Bakshi, Gurdip S., et al. "Valuing European Options by Numerical Integration: The Gauss–Hermite Quadrature Method." *Review of Financial Studies*, vol. 9, no. 2, 1996, pp. 277–300.

continued on next page

continued from next page

8. Basel Committee on Banking Supervision. "The Basel Committee's Responses to Frequently Asked Questions." Basel Committee on Banking Supervision Publications, 2020, www.bis.org.
9. European Securities and Markets Authority (ESMA). "Derivatives." ESMA Publications, 2022, www.esma.europa.eu.
10. U.S. Commodity Futures Trading Commission (CFTC). "Regulations." CFTC Regulations, 2023, www.cftc.gov.

Gemini: Derivative

Derivatives: Understanding Change

The derivative, a cornerstone of calculus, unveils the intricate relationship between change and function behavior. In essence, it quantifies the instantaneous rate of change of a function's output (dependent variable) with respect to its input (independent variable) ([Math is Fun, n.d.]). Imagine a car speeding down a highway; the derivative, at any given moment, captures the car's instantaneous speed, not just its average speed over a certain distance.

Geometrically, the derivative is visualized as the slope of the tangent line to a function's graph at a specific point ([Britannica, 2024]). This tangent line provides a linear approximation of the function's behavior in the immediate vicinity of that point. For instance, the derivative of the function representing the position of a falling object over time would yield its instantaneous velocity at any given time instance.

Finding the derivative, a process known as differentiation, involves applying a set of well-defined rules. These rules, like the product rule, quotient rule, and chain rule, provide a systematic approach to calculating derivatives of various functions ([Derivative Calculator, n.d.]). Notably, not all functions are differentiable at all points, and understanding these limitations is crucial for accurate interpretation.

The significance of derivatives extends far beyond theoretical understanding. They underpin numerous applications across diverse fields, from physics and engineering (modeling motion, optimizing designs) to economics and finance (analyzing market trends, managing risk). By harnessing the power of derivatives, we gain a deeper understanding of change and its impact, empowering us to make informed decisions and solve complex problems in the real world.

continued on next page

continued from next page

Sources:

1. Britannica. (2024, January 24). Derivative. <https://www.britannica.com/money/derivatives>
2. Math is Fun. (n.d.). Introduction to Derivatives. <https://www.mathsisfun.com/definitions/derivative.html>
3. Derivative Calculator. (n.d.). Derivative Calculator • With Steps! <https://www.derivative-calculator.net/>

ChatGPT: Eigenvalue

Abstract: Understanding Eigenvalues in Linear Algebra

Eigenvalues are fundamental concepts in linear algebra, playing a crucial role in various mathematical and scientific disciplines. They are essential for solving systems of linear equations, studying transformations, and analyzing dynamical systems. This abstract provides an overview of eigenvalues, their significance, and their applications, drawing from reputable academic sources.

Eigenvalues represent the scaling factor by which a vector is stretched or compressed when acted upon by a linear transformation. In a square matrix A , an eigenvalue λ and its corresponding eigenvector v satisfy the equation $Av = \lambda v$. This equation encapsulates the idea that the transformation represented by A merely stretches or shrinks the vector v without changing its direction, making eigenvectors invaluable for understanding transformations.

Eigenvalues are extensively utilized in various fields. In quantum mechanics, they are central to the analysis of operators representing physical observables, such as position and momentum. Additionally, eigenvalues play a pivotal role in data analysis and machine learning algorithms. Principal Component Analysis (PCA), for instance, relies on eigenvalues and eigenvectors for dimensionality reduction and feature extraction.

Moreover, eigenvalues are crucial in studying dynamical systems and stability analysis. In the context of differential equations, eigenvalues help determine the behavior of the system over time, aiding in predicting long-term trends and stability conditions. This application is particularly pertinent in fields like physics, engineering, and economics.

continued on next page

continued from next page

Understanding the properties of eigenvalues is essential for solving practical problems efficiently. For instance, the spectral theorem provides a powerful tool for diagonalizing matrices, simplifying complex calculations and facilitating the analysis of linear transformations. Additionally, knowledge of eigenvalues enables the classification of matrices based on their behavior, aiding in the identification of critical points in optimization problems.

In conclusion, eigenvalues are foundational concepts in linear algebra with broad applications across various scientific and engineering disciplines. Their significance lies in their ability to provide insights into linear transformations, dynamical systems, and data analysis. A comprehensive understanding of eigenvalues empowers researchers and practitioners to tackle complex problems effectively.

Sources:

1. Strang, Gilbert. "Linear Algebra and Its Applications." Cengage Learning, 2019.
2. Boyd, Stephen, and Lieven Vandenberghe. "Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares." Cambridge University Press, 2018.
3. Trefethen, Lloyd N., and David Bau III. "Numerical Linear Algebra." SIAM, 1997.
4. Golub, Gene H., and Charles F. Van Loan. "Matrix Computations." Johns Hopkins University Press, 2013.

Gemini: Eigenvalue

Eigenvalues: Unveiling the Hidden Structure of Linear Transformations

Eigenvalues and eigenvectors are fundamental concepts in linear algebra, holding immense significance across various scientific disciplines [3]. An eigenvalue (λ), in essence, is a special scalar value associated with a square matrix (A) that, when multiplied by a non-zero vector (v), yields a scaled version of the original vector ($Av = \lambda v$) [2]. This intriguing property implies that the eigenvector retains its direction under the transformation represented by the matrix, while only its magnitude gets scaled by the eigenvalue [7].

continued on next page

continued from next page

Geometrically, eigenvalues can be visualized as the stretching or shrinking factors imposed by a linear transformation on eigenvectors [4]. For instance, an eigenvalue of 2 indicates a doubling of the vector's length, whereas an eigenvalue of -1 signifies a reflection across the origin along with a reversal of direction [1]. Notably, not all vectors in a space are eigenvectors, and a matrix might possess multiple eigenvalues, each paired with its corresponding eigenvector.

The determination of eigenvalues is achieved by solving the characteristic equation, which is derived from the equation $Av = \lambda v$. This equation essentially asks, "for what values of λ does the system have non-zero solutions for v ?" [8]. The solutions to this equation are the eigenvalues, and they provide valuable insights into the behavior of the linear transformation represented by the matrix.

The applications of eigenvalues and eigenvectors are far-reaching. In physics, they play a crucial role in analyzing vibrations, characterizing wave motion, and solving problems in quantum mechanics [6]. In computer graphics, they are employed for image compression and dimensionality reduction techniques [5]. Additionally, they find applications in engineering for analyzing stability of structures and signal processing.

In conclusion, eigenvalues offer a powerful tool for understanding the nature of linear transformations. By unveiling the hidden structure within matrices, they empower us to analyze a wide range of phenomena across diverse scientific fields.

Sources:

- 1 Strang, G. (2005). Introduction to linear algebra (4th ed.). Wellesley-Cambridge Press.
- 2 Wikipedia. (2023, February 15). Eigenvectors and eigenvalues. https://simple.wikipedia.org/wiki/Eigenvalues_and_eigenvectors
- 3 Wolfram: MathWorld. (n.d.). Eigenvalue. <https://www.wolframalpha.com/calculators/eigenvalue-calculator>
- 4 3Blue1Brown. (2016, December 12). Eigenvectors and eigenvalues | Essence of linear algebra [Chapter 14]. YouTube. YouTube: <https://www.youtube.com/watch?v=PFDu9oVAE-g>
- 5 GeeksforGeeks. (2023, January 10). Eigenvalues & Eigenvectors: Definition, Formula, Examples. <https://www.geeksforgeeks.org/how-to-determine-eigenvalues-of-a-matrix/>

continued on next page

continued from next page

- 6 Professor Dave Explains. (2018, April 12). Finding Eigenvalues and Eigenvectors. YouTube. YouTube: <https://m.youtube.com/watch?v=TQvxWaQnrqI>
- 7 Mitchell, J. (n.d.). Eigenvectors and eigenvalues. https://catalog.usu.edu/preview_course_nopop.php?catoid=38&coid=313040
- 8 Khan Academy. (n.d.). Introduction to linear algebra: Eigensystems & eigenvalues. <https://www.khanacademy.org/math/linear-algebra/alternate-bases/eigen-everything/v/linear-algebra-introduction-to-eigenvalues-and-eigenvectors>

ChatGPT: Statistical Inference

Statistical Inference: Unraveling Patterns from Data

Abstract:

Statistical inference serves as the cornerstone of modern data analysis, enabling researchers and practitioners to draw meaningful conclusions from observed data. This abstract presents a concise overview of statistical inference, its significance in various fields, and key methodologies employed in this domain.

Statistical inference encompasses two primary objectives: estimation and hypothesis testing. Estimation involves inferring unknown parameters of interest, such as population means or proportions, based on sample data. Hypothesis testing, on the other hand, entails assessing the validity of conjectures or hypotheses about population parameters. Through these inferential techniques, researchers can extrapolate findings from a sample to a larger population, facilitating evidence-based decision-making.

Central to statistical inference is the concept of probability distributions, which describe the likelihood of different outcomes occurring. Parametric inference relies on assuming a specific distributional form for the data, such as the normal or binomial distribution, while non-parametric methods make fewer distributional assumptions, leveraging techniques like bootstrapping or permutation tests.

Classical statistical inference, grounded in frequentist principles, emphasizes the long-run behavior of estimators and test statistics under repeated sampling. Confidence intervals provide a range of plausible values for an unknown

continued on next page

continued from next page

parameter, whereas p-values quantify the strength of evidence against a null hypothesis. However, this framework has faced criticism for its reliance on arbitrary significance thresholds and dichotomous decision-making.

Bayesian inference offers an alternative paradigm, treating unknown parameters as random variables with associated prior distributions. By updating these priors in light of observed data using Bayes' theorem, Bayesian methods yield posterior distributions that reflect updated beliefs about the parameters. This approach enables practitioners to incorporate prior knowledge and quantify uncertainty more intuitively.

Statistical inference finds applications across diverse domains, including healthcare, economics, social sciences, and engineering. In medical research, for instance, clinicians use inferential methods to evaluate the efficacy of treatments based on clinical trials. Economists employ inference to analyze trends in financial markets and forecast future outcomes, informing policy decisions.

In conclusion, statistical inference plays a pivotal role in extracting meaningful insights from data, guiding decision-making across various disciplines. Its methodologies continue to evolve, driven by advancements in computational techniques and a growing emphasis on robustness and interpretability.

Sources:

1. Casella, G., & Berger, R. L. (2002). *Statistical Inference* (2nd ed.). Duxbury Press.
2. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). CRC Press.
3. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
4. Lehmann, E. L., & Romano, J. P. (2006). *Testing Statistical Hypotheses* (3rd ed.). Springer.
5. Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer.

Gemini: Statistical Inference

Demystifying the Unseen: A Dive into Statistical Inference

Statistical inference, the cornerstone of data-driven decision-making, empowers us to peek beyond the limitations of a single data point and make informed judgments about entire populations (Wainwright & Jordan, 2008). Unlike descriptive statistics, which simply summarize observed data, inferential methods bridge the gap between the sample we can directly measure and the population we truly want to understand (Wasserman, 2004). This abstract navigates the key concepts and applications of this crucial statistical tool.

At the heart of statistical inference lies the notion of sampling. Due to practical or financial constraints, directly measuring every member of a population is often impossible. Instead, we carefully select a representative sample and use its characteristics to draw inferences about the larger group (Shafer, 2006). This process, however, introduces inherent uncertainty, as any single sample might not perfectly reflect the population. Quantifying and accounting for this uncertainty is paramount to drawing reliable conclusions.

One major branch of statistical inference involves estimation. By analyzing sample data, we aim to estimate population parameters such as the mean, median, or proportion (Casella & Berger, 2002). For instance, a pharmaceutical company might analyze a sample of patients to estimate the average drug efficacy within the entire population. Estimation methods, like confidence intervals, provide a range of plausible values for the population parameter, acknowledging the inherent sampling variability (Good, 2005).

Another crucial application is hypothesis testing. Here, we formulate a specific claim about a population parameter (the null hypothesis) and assess its validity based on sample evidence. By calculating a p-value, which represents the probability of observing such data if the null hypothesis were true, we can gauge the evidence against it. If the p-value falls below a pre-defined threshold (e.g., 0.05), we reject the null hypothesis, suggesting the claim has merit (Moore & McCabe, 2019). However, it's crucial to remember that p-values only reflect compatibility with the data, not absolute truth, highlighting the importance of careful interpretation (Wasserstein & Lazar, 2020).

Statistical inference finds application in diverse fields, from scientific research and business analytics to public policy and social science (Gelman & Hill, 2007). Whether evaluating the effectiveness of a new marketing campaign or understanding disease prevalence within a community, inferential methods empower us to make informed decisions based on limited data, ultimately

continued on next page

continued from next page

unlocking knowledge that would otherwise remain hidden.

Sources:

1. Casella, G., & Berger, R. L. (2002). *Statistical inference* (Vol. 2). Duxbury Press.
2. Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
3. Good, I. R. (2005). *Likelihood, a theory of frequency and weight*. Clarendon Press.
4. Moore, D. S., & McCabe, G. P. (2019). *Introduction to the practice of statistics* (9th ed.). W.H. Freeman and Company.
5. Shafer, G. (2006). *Am I being fooled? How to detect and deal with fallacies*. W.H. Freeman and Company.
6. Wainwright, M. J., & Jordan, M. I. (2008). *Graphical models, exponential families, and variational inference*. Now Publishers Inc.
7. Wasserman, L. (2004). *All of statistics: A concise course in statistical inference*. Springer Science & Business Media.
8. Wasserstein, R. L., & Lazar, N. A. (2020). The ASA statement on p-values: Context, process, and purpose. *The American Statistician*, 74(2), 129-136.

This abstract provides a concise overview of statistical inference, highlighting key concepts, applications, and sources for further exploration. While limited to one page, it serves as a springboard for delving deeper into this essential tool for extracting meaning from data and navigating the world of uncertainty.

ChatGPT: Functional data analysis

Abstract: Functional Data Analysis

Functional Data Analysis (FDA) is a statistical methodology tailored for analyzing data represented by curves, functions, or other smooth, continuous data structures. It offers powerful tools for modeling and understanding complex relationships within data, particularly in fields where traditional methods may fall short, such as in longitudinal studies, time series analysis, and functional neuroimaging.

continued on next page

continued from next page

FDA encompasses various techniques for handling functional data, including functional principal component analysis (FPCA), functional regression, and functional clustering. These methods allow researchers to extract meaningful information from continuous data streams, capturing underlying patterns and dynamics that may be missed by conventional approaches.

One of the key concepts in FDA is the notion of functional data objects, where each observation is treated as a curve or function rather than a single point in space. This shift in perspective enables researchers to analyze entire data trajectories, leading to more comprehensive insights into underlying processes.

In recent years, FDA has gained traction across diverse fields such as biology, finance, and engineering. For instance, in genomics, FDA has been used to analyze gene expression profiles over time, revealing temporal patterns and identifying biomarkers associated with disease progression (Ramsay et al., 2018).

Moreover, FDA has found applications in financial data analysis, where it is employed to model and forecast time-varying stock prices and volatility patterns (Ramsay et al., 2009). By treating financial data as functional objects, analysts can better capture the underlying stochastic processes driving market dynamics.

In the realm of neuroimaging, FDA techniques have been instrumental in studying brain function and connectivity over time. Functional magnetic resonance imaging (fMRI) data, for example, can be analyzed using FDA methods to uncover temporal dynamics in brain activity and identify regions of interest associated with specific cognitive tasks (Aston et al., 2010).

In conclusion, Functional Data Analysis offers a versatile framework for analyzing complex, continuous data structures. By treating data as functional objects, researchers can gain deeper insights into underlying processes and dynamics, leading to enhanced understanding and predictive capabilities across various domains.

References:

1. Ramsay, J. O., Hooker, G., & Graves, S. (2009). *Functional data analysis with R and MATLAB*. Springer Science & Business Media.
2. Ramsay, J. O., Hooker, G., Campbell, D., & Cao, J. (2018). *Functional data analysis with applications in biology*. Springer.
3. Aston, J. A., & Kirch, C. (2010). Recent developments in functional

continued on next page

continued from next page

data analysis and some directions for the future. *Statistica Sinica*, 20(1), 149-170.

Gemini: Functional Data Analysis

Functional Data Analysis: Unveiling Insights from Curves and Surfaces

Functional data analysis (FDA) is a burgeoning statistical framework dedicated to extracting knowledge from data that inherently varies over a continuum, encompassing curves, surfaces, and other objects residing in infinite-dimensional spaces ([Everitt & Hothorn, 2011]). In stark contrast to traditional statistical methods tailored for scalar or low-dimensional data, FDA treats each observation as a complete function, unlocking a wealth of information beyond summary statistics ([Ramsay & Silverman, 2005]).

One of the core strengths of FDA lies in its ability to leverage the inherent smoothness of the underlying functions generating the data. This smoothness assumption allows for the application of powerful techniques like smoothing and registration, enabling researchers to effectively address noise and extract meaningful patterns from the data ([James et al., 2021]).

FDA encompasses a diverse toolbox of methodologies catering to various analytical goals. Common tasks like descriptive analysis, classification, and regression can all be addressed within the FDA framework, although the high dimensionality of functional data necessitates specialized approaches to overcome computational and theoretical challenges ([Yao et al., 2023]).

The applications of FDA span a multitude of domains, particularly flourishing in fields where data naturally manifests as curves or surfaces. Examples include analyzing growth trajectories in biology, EEG signals in neuroscience, and spectral profiles in chemistry ([Ferraty & Vieu, 2006]). As the field continues to evolve, FDA is poised to become an even more valuable tool for researchers seeking to gain deeper insights from the ever-growing landscape of complex, high-dimensional data.

Sources:

1. Everitt, B. S., & Hothorn, T. (2011). *A handbook of statistical analyses using R*. Chapman and Hall/CRC.
2. Ferraty, P., & Vieu, P. (2006). *Nonparametric functional data analysis*. Springer Science & Business Media.

continued on next page

continued from next page

3. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning. Springer.
4. Ramsay, J. O., & Silverman, B. W. (2005). Functional data analysis. Springer Series in Statistics (Vol. 27). Springer.
5. Yao, F., Zhang, J., Zou, H., & Wang, Y. (2023). Functional data analysis: An introduction and recent developments. arXiv preprint arXiv:2312.05523.

ChatGPT: Trend series

Title: Analyzing Trend Series: A Comprehensive Overview

Abstract:

Trend series analysis plays a pivotal role in understanding and predicting various phenomena across multiple domains, from economics to epidemiology. This abstract provides a concise overview of trend series analysis, highlighting its significance, methodologies, and applications.

1. Introduction:

Trend series analysis involves the examination of data over time to identify patterns and tendencies. It is a fundamental tool for understanding the direction and magnitude of changes in various variables, making it invaluable in decision-making processes across disciplines.

2. Methodologies:

Several methodologies are employed in trend series analysis, including but not limited to:

- Time Series Analysis: This method involves studying the data collected at successive points in time to discern patterns, trends, and seasonality.
- Moving Averages: Utilizing moving averages helps smooth out fluctuations in data over time, making underlying trends more apparent.
- Exponential Smoothing: This technique assigns exponentially decreasing weights to past observations, with recent data points carrying more significance.
- Regression Analysis: By fitting a regression model to time-stamped data, analysts can estimate the relationship between variables and forecast future values.

continued on next page

continued from next page

3. Applications:

Trend series analysis finds applications across various fields:

- **Economics:** It aids in forecasting economic indicators such as GDP, inflation rates, and employment figures, guiding policy decisions and investment strategies.
- **Public Health:** Trend series analysis is crucial in tracking disease outbreaks, identifying epidemiological patterns, and allocating resources for effective healthcare interventions.
- **Marketing and Sales:** Businesses utilize trend series analysis to forecast consumer demand, optimize inventory management, and develop targeted marketing campaigns.
- **Environmental Science:** Understanding trends in environmental data, such as temperature patterns or carbon emissions, is essential for policymaking and sustainable resource management.

4. Sources and In-Text Citations:

1. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.
2. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.
3. Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and Applications*. John Wiley & Sons.

Trend series analysis is indispensable for understanding and predicting complex temporal dynamics, offering insights that drive informed decision-making across diverse sectors. By employing various methodologies and drawing from a range of applications, analysts can harness the power of trend series analysis to navigate uncertainties and anticipate future trends.

Word Count: 293 words

Gemini: Trend series

Trend Series: Capturing Dynamism in Data Analysis

continued on next page

continued from next page

Understanding and predicting change is fundamental across various fields, from economics and finance to social sciences and marketing. Trend series analysis plays a crucial role in this pursuit, offering tools to identify, quantify, and forecast patterns in data collected over time (Montgomery et al., 2019). This abstract delves into the core concepts of trend series, highlighting their applications and exploring recent advancements in the field.

Core Concepts and Applications:

A trend series consists of data points ordered chronologically, revealing shifts and variations over time. These changes can be linear, non-linear, seasonal, or cyclical, demanding different analytical approaches. Classical methods like linear regression and exponential smoothing remain widely used for their interpretability and efficiency (Chatfield, 2004). However, complex trends often necessitate more sophisticated techniques such as ARIMA models and state-space modeling, which capture underlying stochastic processes and dynamics (Hyndman & Athanasopoulos, 2018).

Trend series analysis finds diverse applications across numerous domains. In finance, it's crucial for forecasting stock prices, market trends, and risk assessments (Campbell et al., 1997). In social sciences, it helps uncover patterns in population dynamics, disease outbreaks, and public opinion (Chatterjee & Hadi, 2012). In marketing, it allows companies to track sales trends, customer behavior, and campaign effectiveness (Hanssens et al., 2019).

Recent Advancements and Future Directions: The field of trend series analysis is constantly evolving, fueled by advancements in computing power and statistical methods. Recent years have witnessed the rise of:

- Machine learning approaches: Deep learning algorithms, such as recurrent neural networks, are showing promise in capturing complex non-linear trends and making accurate forecasts, especially with large datasets (Diebold et al., 2019).
- Bayesian methods: By incorporating prior knowledge and uncertainty into analysis, Bayesian approaches offer a more flexible and realistic framework for trend estimation and forecasting (Frühwirth-Schnatter, 2019).
- Big data integration: The ability to seamlessly analyze massive datasets from diverse sources, including social media and sensor networks, opens new avenues for understanding complex trends and real-time dynamics (Li et al., 2017).

continued on next page

continued from next page

As we move forward, the continued development of these methods and their integration with domain-specific knowledge will be crucial for extracting even deeper insights from trend series data. This will empower researchers and practitioners to make more informed decisions in various fields, ultimately leading to better predictions and improved outcomes.

Sources:

1. Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton University Press.
2. Chatfield, C. (2004). *The analysis of time series: An introduction*. Chapman and Hall/CRC.
3. Chatterjee, S., & Hadi, A. S. (2012). *Regression analysis by example*. John Wiley & Sons.
4. Diebold, F. X., Giannone, D., & Hlousěk, V. (2019). Forecasting macroeconomic aggregates. In *A Companion to Econometric Forecasting* (pp. 424-462). John Wiley & Sons.
5. Frühwirth-Schnatter, S. (2019). *Specifying bayesian models: Examples in Stan*. CRC Press.
6. Hanssens, D. M., Pauwels, K., Srinivasan, V., & Dekimpe, M. G. (2019). *Marketing metrics: The definitive guide to measuring marketing performance*. Routledge.
7. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
8. Li, Y., Zhang, C., & Zimmermann, K. (2017). Big data in retailing: Opportunities and challenges. *International Journal of Retail & Distribution Management*, 45(1), 3-19.
9. Montgomery, D. C., Jennings, G. C., & Borror, M. (2019). *Introduction to time series analysis and forecasting*. John Wiley & Sons.

Note: This abstract is approximately one page long and includes in-text citations for each source mentioned.

ChatGPT: Extreme value index

Abstract

continued on next page

continued from next page

Extreme value theory (EVT) is a statistical branch focusing on the modeling and analysis of extreme events, which are characterized by their rarity and high impact. The extreme value index (EVI) is a fundamental parameter within EVT, representing the tail behavior of a distribution. This abstract provides an overview of the concept of EVI, its significance in risk assessment, and methods for its estimation.

The EVI, denoted by ξ , is crucial for understanding the tail behavior of a distribution. It quantifies the rate at which extreme values occur, providing insights into the likelihood of rare events. In EVT, the Generalized Extreme Value (GEV) distribution is commonly employed to model extreme events. The EVI governs the shape of the GEV distribution's tail, determining whether it is heavy-tailed, light-tailed, or bounded.

Estimating the EVI involves fitting the GEV distribution to the observed extreme values. Various methods exist for EVI estimation, including the Hill estimator, the Pickands estimator, and the moment estimator. Each method has its strengths and weaknesses, and the choice often depends on the characteristics of the data and the desired precision of estimation.

The EVI finds applications across diverse fields such as finance, environmental science, and engineering. In finance, understanding the EVI is crucial for risk management, especially in modeling tail risk associated with extreme market movements. In environmental science, the EVI helps in assessing the probability of rare natural disasters such as floods and hurricanes. In engineering, knowledge of the EVI aids in designing structures resilient to extreme loads.

In conclusion, the extreme value index is a fundamental parameter in extreme value theory, playing a vital role in understanding and quantifying rare events with significant impacts. Accurate estimation of the EVI is essential for effective risk assessment and decision-making across various domains.

Sources:

1. Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events: for insurance and finance*. Springer Science & Business Media.
2. Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer Science & Business Media.
3. Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society*.

continued on next page

continued from next page

Series B (Methodological), 52(3), 393-442.

4. Leadbetter, M. R., Lindgren, G., & Rootzén, H. (1983). Extremes and related properties of random sequences and processes. Springer Science & Business Media.

Gemini: Extreme value index

The Extreme Value Index: Unveiling the Tail Behavior of Extremes

Extreme value theory (EVT) emerges as a crucial branch of statistics, delving into the realm of exceptional deviations from a distribution's central tendency [2]. This theory empowers us to assess the likelihood of encountering events exceeding any previously observed extremes within a given dataset [2]. A key parameter within EVT is the extreme value index (denoted by γ or θ), which offers valuable insights into the tail behavior of these extreme events [1, 4].

For a broad range of stationary time series, EVT provides limiting distributions for rare events [3]. It goes beyond simply describing the magnitude of extremes, but also sheds light on their frequency of occurrence. A critical question regarding extreme events is whether they appear in isolation or cluster together. This is precisely where the extremal index steps in [3].

Intuitively, the extremal index (θ) can be grasped as a measure of the extent to which extremes tend to cluster [4]. It represents the reciprocal of the average cluster size [4]. A positive extremal index ($\theta > 0$) signifies that extremes tend to cluster, implying a higher likelihood of observing multiple extreme events close together. Conversely, a negative extremal index (< 0) suggests that extremes are more independent, meaning they are less likely to cluster [1]. The special case where $\theta = 0$ indicates that the tail behavior follows a generalized Pareto distribution (GPD), with the extremal index governing the shape of the tail [5].

Estimating the extremal index is an essential step in EVT applications. Statistical packages like the `exdex` package in R offer functionalities to implement various estimation methods for [4]. These methods play a vital role in various fields, including finance, where they can be used to model extreme market fluctuations, or environmental science, where they can aid in understanding and predicting the occurrence of extreme weather events.

In conclusion, the extremal index serves as a powerful tool within EVT, providing valuable insights into the clustering behavior of extreme events. Understanding its implications allows for a more comprehensive analysis of

continued on next page

continued from next page

rare events and their potential consequences across various disciplines.

Sources:

1. Einmahl, J. H., et al. (2016). Estimation of the extreme-value index and generalized quantile plots. *Project Euclid*, 11(6), 1009-1025.
2. Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London: Chapman and Hall/CRC.
3. Rootzen, H. (2018). An overview of the extremal index. [HAL] hal-02334271
4. Beguería, S. (2020). Introducing exdex: Estimation of the Extremal Index. <https://cran.r-project.org/package=extRemes>
5. Pereira, A. C., et al. (2019). Trends in Extreme Value Indices. [Taylor & Francis Online] doi.org/10.1080/01621459.2019.1705305

ChatGPT: Gaussian processes

Abstract

Gaussian processes (GPs) are powerful probabilistic models used in various fields, from machine learning to Bayesian optimization, due to their flexibility and ability to model complex data distributions. In recent years, there has been growing interest in extending GPs to incorporate information from multiple heterogeneous data sources, leading to the development of Gaussian processes with sources (GPS).

This abstract presents an overview of Gaussian processes with sources, highlighting their key concepts, methodologies, and applications. First, we discuss the basic framework of Gaussian processes, emphasizing their probabilistic nature and their use as non-parametric models for regression and classification tasks. We then delve into the extension of GPs to Gaussian processes with sources, which enables the incorporation of diverse data modalities or information from multiple domains into a unified modeling framework.

We explore various approaches for integrating multiple data sources within the GPS framework, including concatenation-based methods, feature augmentation techniques, and multi-output Gaussian processes. These methods allow for the seamless fusion of information from different sources while preserving the

continued on next page

continued from next page

inherent uncertainty estimates provided by Gaussian processes.

Furthermore, we discuss the applications of Gaussian processes with sources across various domains, such as healthcare, finance, and environmental science. In healthcare, GPS models have been used for personalized medicine, integrating genomic, clinical, and imaging data to predict patient outcomes and recommend treatment strategies. In finance, GPS have been applied to portfolio optimization, combining information from diverse asset classes to improve investment decision-making. In environmental science, GPS have been used for climate modeling, assimilating data from satellites, weather stations, and numerical simulations to enhance predictive accuracy.

In conclusion, Gaussian processes with sources offer a flexible and principled framework for modeling heterogeneous data sources, enabling more accurate predictions and informed decision-making across a wide range of applications. Further research in this area holds great potential for advancing the capabilities of probabilistic modeling in handling complex, multi-source data environments.

In-text citations:

1. Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian processes for machine learning. MIT Press.
2. Bonilla, E. V., Chai, K. M. A., & Williams, C. K. I. (2008). Multi-task Gaussian process prediction. In *Advances in neural information processing systems* (pp. 153-160).
3. Vanhatalo, J., Riihimäki, J., Hartikainen, J., Jylänki, P., Tolvanen, V., & Vehtari, A. (2013). GPstuff: Bayesian modeling with Gaussian processes. *Journal of Machine Learning Research*, 14(1), 1175-1179.
4. Dezfouli, A., & Bonilla, E. V. (2015). Gaussian processes with Beta-Sherman-Woodbury kernels: Application to multi-source time-series data. In *Advances in Neural Information Processing Systems* (pp. 1191-1199).
5. Lázaro-Gredilla, M., Quiñonero-Candela, J., Rasmussen, C. E., & Figueiras-Vidal, A. R. (2010). Sparse spectrum Gaussian process regression. *Journal of Machine Learning Research*, 11(Feb), 1865-1881.
6. Wilson, A. G., Dann, C., & Nickisch, H. (2016). STAN: A probabilistic programming language. *Journal of Statistical Software*, 20(1), 22-43.
7. Alaa, A. M., & van der Schaar, M. (2017). Bayesian inference of individualized treatment effects using multi-task Gaussian processes. In *Advances in Neural Information Processing Systems* (pp. 4028-4038).

Gemini: Gaussian processes

Gaussian Processes: A Probabilistic Approach to Function Learning

Gaussian processes (GPs) offer a powerful and versatile framework for machine learning tasks involving function learning. They model data as a collection of random variables with a joint multivariate normal distribution, enabling the representation of uncertainty and the prediction of unseen data points [2]. This one-page abstract delves into the core concepts of GPs, their applications, and their unique advantages in various machine learning scenarios.

Conceptual Foundations:

At its core, a GP defines a distribution over functions, where any finite collection of function values follows a multivariate normal distribution [1]. This probabilistic approach allows GPs to capture the relationship between input and output variables, while also providing uncertainty estimates for predictions [5]. The key component in defining a GP is the kernel function, which determines the similarity between data points and governs the smoothness and behavior of the inferred function [4]. Different kernel functions can be chosen based on the specific problem and prior knowledge about the data.

Applications and Advantages:

GPs have found applications in various machine learning domains, including:

- **Regression:** Predicting continuous-valued outputs based on input features. GPs excel in tasks where the underlying function is complex or non-linear, as they can learn flexible relationships without overfitting the data [3].
- **Classification:** Classifying data points into discrete categories. Gaussian Process Classifiers (GPCs) provide probabilistic outputs, indicating the confidence in the predicted class labels [3].
- **Time series forecasting:** Predicting future values in a sequence of data points. GPs can capture temporal dependencies and uncertainties in time series data, making them suitable for various forecasting tasks.

Several advantages distinguish GPs from other machine learning methods:

- **Non-parametric nature:** GPs do not require explicit assumptions about the underlying functional form, making them adaptable to diverse data distributions.
- **Uncertainty quantification:** GPs inherently provide uncertainty estimates for predictions, allowing for a more nuanced understanding of the model's confidence in its outputs.

continued on next page

continued from next page

- Bayesian framework: GPs naturally integrate with Bayesian inference, enabling the incorporation of prior knowledge and the construction of robust models.

Conclusion:

Gaussian processes offer a powerful and flexible framework for function learning in machine learning. Their ability to model complex relationships, provide uncertainty estimates, and integrate with Bayesian inference makes them valuable tools for various applications. While their computational complexity can be a limitation in certain scenarios, ongoing research continues to explore efficient algorithms and improve the practical applicability of GPs.

Sources:

1. Wikipedia. (2023, November 19). Gaussian process. https://en.wikipedia.org/wiki/Gaussian_process
2. Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian processes for machine learning. The MIT Press.
3. scikit-learn [scikit-learn.org](https://scikit-learn.org/stable/modules/gaussian_process.html). (n.d.). 1.7. Gaussian processes. https://scikit-learn.org/stable/modules/gaussian_process.html
4. Distill. (2019, August 28). A visual exploration of Gaussian processes. [invalid URL removed]
5. d2l.ai. (n.d.). 18.1. Introduction to Gaussian processes. <https://www.d2l.com/blog/d2l-and-artificial-intelligence-the-past-present-and-future/>

ChatGPT: Statistical depth

Abstract: Statistical Depth Measures and Their Applications

Statistical depth measures are powerful tools used in data analysis and machine learning to assess the centrality and outlyingness of data points within a dataset. These measures provide a robust and comprehensive understanding of the distributional characteristics of the data, offering insights beyond traditional summary statistics like mean and variance. This abstract provides an overview of statistical depth measures, their properties, applications, and recent advancements.

continued on next page

continued from next page

Statistical depth measures quantify the centrality of a data point relative to the entire dataset by assessing its distance from the dataset's center. One of the most widely used depth measures is the Tukey depth, which computes the minimum probability of a point being outside a closed convex hull containing it. Other popular depth measures include the half-space depth, zonoid depth, and projection depth, each offering unique perspectives on data centrality.

These measures possess several desirable properties, including affine invariance, maximality, and outlier identification capabilities. Affine invariance ensures that depth measures remain unchanged under linear transformations of the data, facilitating their applicability across various domains. Additionally, maximal depth points, defined as points with the highest depth values, serve as robust estimates of centrality and are often used in robust statistical inference.

Statistical depth measures find applications in various fields, including multivariate analysis, clustering, classification, and outlier detection. In multivariate analysis, depth-based methods aid in visualizing high-dimensional data by representing it in lower-dimensional spaces while preserving its structural properties. In clustering and classification tasks, depth measures facilitate the identification of representative prototypes and the detection of outliers, enhancing the performance and interpretability of machine learning models.

Recent advancements in statistical depth measures focus on developing efficient algorithms for computing depths in large-scale datasets and extending their applicability to non-Euclidean spaces. Additionally, researchers are exploring the integration of depth measures with deep learning techniques to enhance the robustness of neural networks and improve their interpretability.

In conclusion, statistical depth measures offer a robust framework for analyzing data centrality and identifying outliers across various domains. Their properties and applications make them valuable tools in data analysis, machine learning, and statistical inference. Continued research in this field is essential for advancing the understanding and utilization of depth measures in real-world data analysis tasks.

Sources and In-text Citations:

1. Liu, Rong, and Dianne P. O'Leary. "Statistical depth measures and their applications." *Journal of Statistical Planning and Inference* 139.5 (2009): 1249-1260. [1]

continued on next page

continued from next page

2. Zuo, Yijun, and Hui Zou. "Projection-based depth functions and associated medians." *The Annals of Statistics* 37.6A (2009): 3696-3717. [2]
3. Liu, Rongmao, et al. "Halfspace depth and regression depth characterizations of regression ranks and regression medians." *Statistica Sinica* (2022). [3]
4. Liu, Rongmao, and Yijun Zuo. "An overview of projection depth and its applications." *Statistical Analysis and Data Mining: The ASA Data Science Journal* 6.6 (2013): 515-537. [4]
5. Mizera, Ivan, and Jiří Zvára. "Centrality and statistical depth: New perspectives and applications." (2018). [5]

Gemini: Statistical depth

Statistical Depth: Unveiling Centrality in Complex Data

Statistical depth, a concept developed in the 1980s, equips us with a powerful tool to assess the "centrality" of data points within a complex, often high-dimensional, data landscape. Unlike traditional approaches solely focused on location measures like mean or median, statistical depth delves deeper, gauging how "representative" a point is compared to the underlying data distribution (Liu, 1990). This paves the way for ranking and ordering data points, and subsequently, enables nonparametric statistical inference, particularly useful for datasets lacking readily available distributional assumptions [Zuo & Yao, 2006].

Formally, a statistical depth function, denoted as $D(x; F)$, assigns a numerical value to each data point "x" within a distribution "F." Points residing deeper within the "core" of the distribution receive higher depth values, indicating greater centrality. Conversely, points lying on the periphery are assigned lower depths, reflecting their relative "outlyingness" (Zuo & Yao, 2006). Notably, statistical depth functions adhere to specific properties, including:

- **Affine invariance:** The depth of a point remains unchanged under linear transformations of the data space.
- **Maximality at the center:** The central point(s) within the data distribution possess the highest depth values.
- **Monotonicity:** The depth of a point decreases as it moves away from the "deepest" point (the data point with the highest depth).

continued on next page

continued from next page

- Vanishing at infinity: As the distance from a point to the core of the data distribution increases, its depth value approaches zero [Liu, 1992].

These properties ensure the robustness and interpretability of statistical depth measures. Several depth functions have been proposed, each exhibiting unique characteristics and suitability for different types of data and applications, such as the simplicial depth and the halfspace depth function [Liu, 1990, Riani & Cerioli, 2005].

Beyond ranking individual data points, statistical depth empowers us to construct "depth-trimmed regions" by identifying the subset of data points exceeding a specific depth threshold. These regions offer valuable insights into the central tendency, spread, and shape of the underlying distribution, providing a nonparametric alternative to traditional location and scale estimators [López-Pintado & Romo, 2012].

The versatility of statistical depth extends beyond traditional data analysis. Recent research explores its application in diverse fields, including ranking and characterizing complex objects like transformer models in natural language processing [Zhang et al., 2023]. As statistical depth continues to evolve, it serves as an increasingly valuable tool for unveiling the hidden structure and relationships within complex data, offering valuable insights across various scientific disciplines.

Sources:

1. Liu, R. Y. (1990). On a notion of data depth based on random simplices. *Annals of Statistics*, 18(1), 405-414.
2. Liu, R. Y. (1992). Contributions to multivariate analysis. In *Multivariate Analysis and Its Applications* (pp. 207-223). Springer, New York, NY.
3. López-Pintado, S., & Romo, J. (2012). On the use of spatial depth for exploring multivariate data. *Journal of Applied Statistics*, 39(8), 1946-1962.
4. Riani, M., & Cerioli, A. (2005). Depth measures and classical multivariate location. *Statistical Methods and Applications*, 14(3), 229-244.
5. Zuo, Y., & Yao, Q. (2006). Depth-based rank tests for censored data. *Statistics & Probability Letters*, 76(13), 1393-1402.
6. Zhang, Y., Yuan, S., & Ma, X. (2023). Statistical depth for ranking and characterizing transformer. In *Proceedings of the 2023 Conference*

continued on next page

continued from next page

on Empirical Methods in Natural Language Processing (EMNLP) (pp. 9366-9376).

ChatGPT: Influence function

Title: Understanding Influence Functions in Machine Learning: A Comprehensive Overview

Abstract:

Influence functions play a crucial role in understanding the impact of individual data points on model predictions in machine learning. This abstract provides a comprehensive overview of influence functions, their applications, and recent advancements in this field.

Firstly, influence functions were introduced by Cook and Weisberg (1982) in the context of linear regression models to measure the effect of individual observations on parameter estimates. Since then, their utility has expanded to various machine learning models, including deep neural networks (Koh & Liang, 2017).

Influence functions quantify the sensitivity of model predictions to changes in the training data by assessing how perturbations in individual data points affect model parameters or predictions. This property makes them valuable for various tasks, such as outlier detection, robustness analysis, and model debugging.

Recent research has focused on developing efficient algorithms for computing influence functions in complex models, including deep learning architectures. For instance, Koh et al. (2019) proposed fast algorithms for estimating influence functions in neural networks, enabling scalable computation even for large datasets.

Moreover, influence functions have been utilized in diverse applications across different domains. For instance, in healthcare, they have been used to identify influential patient records affecting clinical decision-making (Shalit et al., 2017). In finance, influence functions have been employed for risk assessment and portfolio optimization (Kogelman et al., 2020).

Despite their utility, influence functions have certain limitations, such as sensitivity to model assumptions and computational challenges in high-dimensional settings. Addressing these limitations remains an active area of research, with

continued on next page

continued from next page

ongoing efforts to improve robustness and scalability.

In conclusion, influence functions are powerful tools for analyzing the impact of individual data points on model predictions in machine learning. They have diverse applications across various domains and continue to drive research advancements in understanding model behavior and improving model performance.

References:

1. Cook, R. D., & Weisberg, S. (1982). Residuals and influence in regression. Chapman and Hall.
2. Koh, P. W., & Liang, P. (2017). Understanding black-box predictions via influence functions. Proceedings of the 34th International Conference on Machine Learning.
3. Koh, P. W., et al. (2019). On the accuracy of influence functions for measuring group effects. Proceedings of the 36th International Conference on Machine Learning.
4. Shalit, U., et al. (2017). Estimating individual treatment effect: generalization bounds and algorithms. Proceedings of the 34th International Conference on Machine Learning.
5. Kogelman, L., et al. (2020). Efficient algorithms for adversarial portfolio optimization with transaction costs. Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics.

Gemini: Influence function

Unveiling the Inner Workings: Understanding Models Through Influence Functions

In statistics, the concept of influence functions plays a crucial role in assessing the sensitivity of an estimator to individual data points ([Kyle Butts, All about influence functions], [MyWeb, Statistical functionals and influence functions]). An estimator is a statistical method used to approximate an unknown population parameter based on a sample. The influence function quantifies how much a single data point can sway the estimated value.

Mathematically, the influence function of an estimator $\hat{\theta}$ at a data point z is defined as the rate of change of $\hat{\theta}$ with respect to an infinitesimal contamination of the data at z ([Wikipedia, Influence function]). In simpler terms, it tells us

continued on next page

continued from next page

how much the estimate would change if we added a slightly different data point instead of the original one.

Influence functions offer a powerful tool for analyzing various aspects of statistical models, particularly in the realm of robust statistics. Robust statistics deals with methods that are less susceptible to outliers or influential points in the data ([Kyle Butts, All about influence functions]). By examining the influence function of an estimator, we can:

- Identify influential points: Data points with a large influence function can significantly impact the estimate. This information helps us diagnose potential outliers or biases in the data ([OpenReview, UNDERSTANDING INFLUENCE FUNCTIONS AND DATA- MODELS VIA HARMONIC ANALYSIS]).
- Explain model predictions: In the context of machine learning models, influence functions can be used to understand how individual data points contribute to a specific prediction. This sheds light on the "black-box" nature of complex models, allowing us to interpret their decision-making process ([arXiv, Understanding Black-box Predictions via Influence Functions]).
- Assess model bias: Influence functions can reveal how sensitive an estimator is to certain data features. This can help us identify potential biases in the model's training data or its learning algorithm ([OpenReview, UNDERSTANDING INFLUENCE FUNCTIONS AND DATA- MODELS VIA HARMONIC ANALYSIS]).

Beyond these core applications, influence functions have broader implications for various statistical tasks. They are used to construct confidence intervals, estimate Shapley values (which explain feature contributions in machine learning models), and even detect data poisoning attacks where malicious actors attempt to manipulate the model's training data ([OpenReview, UNDERSTANDING INFLUENCE FUNCTIONS AND DATA- MODELS VIA HARMONIC ANALYSIS]).

In conclusion, influence functions provide a valuable lens for understanding the inner workings of statistical models. By analyzing how individual data points influence the estimated value, we can gain insights into model robustness, identify potential biases, and explain complex predictions. As the field of data science continues to evolve, influence functions are poised to play an increasingly important role in building robust, interpretable, and trustworthy models.