You might consider that with the increasing accessibility of user-friendly statistical software and data that performs the mathematical heavy-lifting for you, there's less require to be taught in statistical methods.

However, the easiness with which we can now access data and analyze data has caused an increase in the utilization of statistical figures and graphics as a way of providing allegedly objective evidence for claims. Nowadays, it's not only scientists who utilize statistics as evidence; however, political campaigns, advertisements, and the media also use statistics. As statistics are separated from their scientific foundation, their aim is transforming to convince instead of informing.

Also, the people making those kinds of statistical claims are not essentially qualified in statistical methods. A progressively various number of sources make and disseminate statistics with really small mistakes to make sure of their reliability. Also, when data is made by scientists in charge of the research, mistakes, and misrepresentations of statistical claims can happen at any given time in the cycle – from errors in the research to misrepresentations by the media as well as the public.

Therefore, in this present world, data literacy is now vital so as to correctly appraise the credibility of the various news stories, social media posts, and debates that make use of statistics as evidence. This book summary will offer you the entire tools you require to better assess the statistics you met regularly.

Chapter 1 - Statistics can assist in providing answers to the questions about the universe.

Have you ever thought of what statisticians really do?

To a lot of people, statistics is a mysterious part of mathematics, just a bit more interesting than the others since it uses pictures.

However, nowadays, the mathematical aspect of statistics is viewed as just one part of the discipline. Statistics handle the whole lifecycle of data, that entails five phases which can be

summarized by the acronym PPDAC: Problem, Plan, Data, Analysis, and Conclusion. The work of a statistician is to detect an issue, make a strategy to fix it, collect the significant data, analyze it, and interpret a proper conclusion.

Let's demonstrate how this process functions by taking a look at a real-life example that the author was one time involved in the case of Harold Shipman, the serial killer.

Harold Shipman was the United Kingdom's most prolific serial killer, with 215 sure victims and 45 likely ones. In 1998, before he got arrested, he made use of his status of authority as a doctor to kill a lot of his elderly patients. His method was to inject his patients with a poisonous amount of morphine and afterward modify their medical records in order to make their deaths look more natural.

The author was part of the task force created by a public inquiry to know if Shipman's murders could have been noticed earlier. This entails the first phase of the investigative cycle – the issue.

The next phase is the plan – was to gather evidence concerning the deaths of Shipman's patients and relate this with evidence concerning other patient deaths in the region to notice if there were any doubting incongruities in the information.

The third phase of the cycle is data –which entails the real process of collecting data. In this case, that entailed checking hundreds of physical death certificates from the year 1977 upwards.

In the fourth phase, that was data analysis whereby data were analyzed, and the data were entered into the software and compared with graphs. The analysis revealed two things: the first thing was, Shipman's practice recorded a very higher number of deaths than normal for his region. Secondly, while patient deaths for the other general practices were dispersed the entire day, Shipman's victims had a tendency to die between the time 01:00 p.m. and 05:00 p.m. –exactly when Shipman went for his home visits.

The last phase is the conclusion. The report of the author concluded that if a person had been observing the data, Shipman's deeds could have been detected as soon as 1984 – 15 years before– which could have saved close to175 lives.

Therefore, what is the work of statisticians? They observe trends in data to fix real-world issues.

Chapter 2 - The precision of data is usually skewed by systematic bias.

Data is not only cold, real facts – it's prone to biases and human judgment just like any other kind of knowledge.

As a matter of fact, human judgment is needed in the actual first phase. At times, we need to make fairly random choices about what we're measuring before we can gather data. If our issue is to count the number of trees that are present on the planet, then we have to define what precisely a "tree" is. For example, the majority of the studies of this kind only has trees that have reached a diameter of a minimum of 4 inches.

Subsequently, if the definition of what is being measured modifies halfway through measurement data can be skewed. For instance, the number of sexual crimes recorded by the police in the UK between the years 2014 and 2017 nearly increased two-folds from 64,000 cases to 121,000 cases. It might look as if crime increased in those years. But, the actual cause for the crime was that sexual offenses were taken very extreme after a 2014 report critiqued police recording operations.

Therefore, we should never conclude that data is a completely precise representation of truth. Think of the fact that a lot of data is gathered from surveys that question people about their experience, like how happy they feel. Definitely, those kinds of questions can't be assumed to get the full range of human experience on a spreadsheet. Also, the data can be skewed from biases on how people interpret and answer the question.

This is the reason why creating the right questions is one of the huge problems of statistics. The language used in asking questions can affect how the respondent feels about the question. When a UK survey questioned respondents how they felt about "providing 16 to17-year-olds the right to vote," 52% were for it and 41 percent opposed it. However, when the exact respondents

were asked the reasonably same question about how they felt about "dropping the voting age from 18 to 16," the people that supported it reduced to 37 percent with 56 percent opposed.

In other cases, it's not the question that leads to bias; however, the answers the survey allows. In 2017, Ryanair proudly declared that 92% of its passengers were contented with their flight experience. But, as a matter of fact, that their customer satisfaction survey just allowed the replies "excellent, very good, good, fair, and ok."

What this means is that before statisticians even get the data, they're usually already handling false data.

Chapter 3 - The presentation of data influences how it's interpreted.

The general issue of human interpretation doesn't just have an effect on the collection of data; however, it affects its presentation too.

Current years have witnessed an increase in the study of data visualization as a way of communicating statistical findings. Data visualizations are graphical methods utilized to make data discernible to the eye. Statisticians explain visual representations of data as inter-ocular, which means that patterns found in the data become visible from sight only– without needing to do any form of mental mathematics.

An illustration of this would be when a bar chart is made use of to compare the number of deaths as a result of heart surgery across various hospitals. Without needing to look at the numbers, any hospital that extremely drifts away from the usual will be noticeable to the naked eye.

However, graphics need a careful plan if they have to precise and effective. Things such as language, color, order, and font have an effect on how the data is interpreted. This is the reason why, presently, statisticians work together with psychologists to assess how alternative graphics are likely to be viewed.

To abide by the illustration of comparing death rates in hospitals, assuming a statistician showed her data in a table. She needs to choose in what order she's going to list the hospitals. It might look reasonable to list the hospitals by the death rate. The issue is, this way might give the feeling that the hospitals are ranked based on their quality, which would be really deceptive; the best hospitals usually get more death rates since the most extreme conditions need to be treated there.

A different well-documented case of how presentation impacts interpretation is the effect of framing. The language whereby a statistical claim is framed has an effect on its emotional impact.

In the past few years, an advertising campaign on the London underground asserted that 99% of young Londoners do not commit any extreme youth violence. The purpose, seemingly, was to reassure London's citizens of their protection. However, we could modify this claim's emotional impact if we change the statistic to state that "1% of young Londoners commit extreme youth violence." That's a bit really threatening. If we used a real number in place of a percentage the consequence is even more noticeable: "London has 10,000 violent young offenders!"

Statistics communicators usually make the most of framing, reliant on if they want to shock or reassure their audience. Researchers have to be cautious to obtain wrong gut reactions to data by making use of careful design and clear language.

Chapter 4 - In scientific literature, there is a positive bias brought about by selective reporting.

A lot of researchers used their whole lives searching data for significant findings that they hardly discover. At times, the pressure to publish important work makes researchers massage the data a bit.

In spite of their commitment to the truth, scientists as well have been known to partake in some doubtful research practices. Multiple testing is one form of such practice, and this is when researchers repeat tests until they get the outcomes they wish for. The more a researcher

repeats a test, the more the likelihood that they will get wrong positives –findings that look to verify a hypothesis; however, they are really as a result of the chance error.

In order to know the reason why this is an issue, let's consider a study that was conducted by a team of really respectable researchers in 2009. They used brain imaging to check which regions of a subject's brain would light up when they were presented with a range of pictures of people showing various emotions. The issue was that the "subject" was a four-pound, Atlantic salmon that was already dead. From the 8,064 areas assessed in the fish's brain, 16 displayed a reaction to the pictures. Instead of concluding that the fish possessed extraordinary powers, the team properly concluded that more than 8,000 tests are certain to cause a few false positives.

False positives aren't essentially an issue in themselves; however, usually, they're the only findings that become published. In scientific reporting as well, just positive or interesting results have a tendency to be reported. This has unavoidably caused a positive bias within the scientific literature, this means that the public only gets the studies that look to back up a hypothesis – not the ones that don't. Obviously, this affects how the findings are interpreted.

For instance, you might be surprised if a study discovered that consuming bacon sandwiches increases your risk of getting cancer. However, if you were aware that 20 former studies had discovered no association at all, your shock would probably be silenced.

The cause for that kind of selective reporting is difficult, owing both to extreme academic pressure and our taste for dramatic, innovative stories.

Positive bias is what made a professor of statistics at Stanford University named John Ioannidis, to state that "the majority of the published research findings are incorrect." Although Ioannidis was being deliberately provocative, his assertion works as a caution not to take findings from research for granted only because they're reported in a scientific journal.

Chapter 5 - The media has a tendency to highlight storytelling at the cost of accuracy.

As soon as research has been published, then the media reports it. But, the media has a tendency to use creative license.

The good news is that data journalism is thriving. Journalists are progressively being taught on how to interpret and communicate data. Statistics can enhance stories by bringing clarity and understanding to significant problems.

However, there's usually the danger that statistical assertions will be biased in the process of storytelling. Stories frequently need an emotional hit, which science journals hardly give. For institutions more interested in growing web traffic than reporting correct research, the temptation will usually be to avoid nuanced conclusions in favor of sensationalized ones.

The author was one time aware of that kind of sensationalizing after a careless comment he made during a conversation. He was replying to the outcomes of a national survey that observed the sexual behaviors of the people of Britain. The study found out that young adults in Britain were having sexual intercourse 20% less often than they had been a decade ago. The assumption by the author that a rise in content such as Netflix may be related to the decrease triggered a flurry of ridiculous headlines like "as stated by one lone scientist, sex will be outdated by 2030 as a result of Netflix."

Apart from such obvious lies, exaggerating of statistical claims of risk is one of the most popular means the media schemes an emotional punch.

During the time that a report from the World Health Organization discovered that frequent consumption of processed meat caused an 18% increased risk of having bowel cancer, this 18% number was extensively stated by the media. And, certainly, it actually seems scary. However, how concerned should we actually be?

Although the media did report the 18% figure correctly, they didn't differentiate between relative and absolute risk.

The increased risk of bowel cancer from the constant consumption of processed meat was relative to the 6% risk for people who consume processed meat frequently. Therefore, an 18% increase from 6 is 7.08 percent. In absolute terms, the increase in risk faced by people who

constantly consume processed meat is just roughly 1% higher than the risk faced by people who don't consume it regularly–really less scary.

Exaggerating risk is only one of the popular means that statistical claims are distorted. In the next chapters, we will examine some other usual interpretation errors.

Chapter 6 - Reported averages can be deceptive when the kind of average used isn't identified.

Wrong uses of averages have stirred some very bad jokes among statisticians.

Think of this one: "The majority of us possess more legs than average." This is correct, in some way, if you calculate the average number of legs making use of the mean average, which is approximately like 1.9999 – brought down from 2 by people who don't have legs again.

Okay, if that one wasn't for you, think of this: "Averagely, the whole public has a testicle." This as well is correct–however, just if you make use of a mean average that takes account of women in the calculation.

Both of these strange assertions are gotten by wrongly using the mean average when the median or mode would have brought about more reasonable approximations.

Let's do a fast summary of these three kinds of average. The mean average is gotten by adding the entire numbers in a data set and then dividing it by the numbers there are. The median average is the number that is found in the middle when the entire numbers in the data set are arranged in ascending order. Lastly, the mode average is the number that appears the most in the data set.

Different kinds of average are suitable for different conditions. For instance, the mean average is utilized best when the entire numbers in a data set group symmetrically around a middle value. However, in a lot of other circumstances, it can be really deceptive.

Let's take a lot at a different case. The UK National Survey of Sexual Attitudes and Lifestyle told respondents to mention the number of sexual partners they'd had. This is how the data looks like: 1 was the most frequently reported number of sexual partners, the majority reported between 0 and 20, and just a few reported numbers between 20 and 500.

Due to the small number of outliers mentioning numbers very higher than 20, the mean average of sexual partners is most likely to be very higher than the huge majority of people's experience and is, hence, a deceptive average to use. The median average is going to offer a number very close to the average person's experience, and the mode provides us an understanding of the most common experience.

The kind of average utilized is nearly never mentioned together with reported conclusions. The most common, is the mean average, which we've understood is usually wrong–therefore, a lot of statistical claims we hear in the media are deceptive and not important to our experiences.

Chapter 7 - Correlation does not mean causation is the statistician's slogan.

 The charge correlation does not mean causation has to turn into something of a motto among statisticians. Still, the error of deducing cause from correlation is still frequently made by the media and the public alike; therefore, it's a message that has to be repeated.

This fallacy is what brings about the comical headlines, like: "The reason why attending a university increases the risk of having a brain tumor."

The study that this headline is founded on discovered that a slightly higher number of people who got a brain tumor were from a higher socioeconomic status. However, this in itself doesn't entail that there is a causal association between the two.

As a matter of fact, even the authors of the study thought that the correlation was a result of some form of ascertainment bias. Meaning, people who come from a higher socio-economic

background were very more likely to get tested for and consequently diagnosed with brain cancer.

Therefore, when two data sets correlate, we shouldn't deduce that one leads to the other. The correlation could be clarified by any of the three other possibilities.

Firstly, it might be just a mere coincidence that two data sets correlate, as this silly case shows: between the years 2000 and 2009, there is a strong correlation between the per capita intake of mozzarella cheese in the US and the number of engineering doctorates given. In spite of the correlation, it's not likely that the increase in cheese intake has something to do with the number of people becoming engineers.

Secondly, correlated data could similarly be clarified by the reverse causal relationship to the one we anticipate. For instance, a lot of studies that compare alcohol intake with health consequences show that people who don't drink have a more death rate than people who drink alcohol moderately. It is studies such as these that create such hopeful headlines such as "drinking a glass of wine a night is really good for your health." But, this is supposed to be an illustration of reverse causation, because people who are sick already have a tendency to avoid alcohol.

Lastly, a correlation between two sets of data could be the cause of a hidden factor –this is a thing that isn't taken into consideration by a study; however, that affects both of the observed components which are. For instance, a correlation between the sales of ice-cream and drowning is possible as a result of the weather, which affects both.

Therefore, should in case you didn't get this the first time: correlation does not mean causation.

Chapter 8 - Probability is usually misinterpreted.

One time, when the author was asked on the reason why people have a tendency to find probability hard and counterintuitive, he answered by saying that probability basically really is hard and counterintuitive.

Even the people directing the country have issues understanding it as well. In the year 2012, 97 people from the parliament in the UK were told to answer this question: "If you flip a coin twice, what's the probability of having two heads?" The answer to the question is one quarter because having two heads is one possibility out of a possible four. Most of the parliament members – 60 people out of 97 people– could not answer the question correctly.

Think of this other probability question: think that approximately 1% of women have breast cancer and that mammography screening is 90% correct in identifying breast cancer. Say a woman is diagnosed with breast cancer, what's the probability that she truly has breast cancer?

Obviously, we might think that the woman has a 90% probability of getting breast cancer because that's how precise the scan is. However, the fact is, she just has an 8% probability. This cause for this counterintuitive result is that there are likely to be very more false positives due to the much bigger group of women who don't get breast cancer than true positives among the lesser group who have.

Gambler's fallacy is another frequent probability fallacy, this is when people modify their expectations of the chance of individual events according to what's come previously. For instance, after a long run of black on the roulette wheel, the temptation is to think that red is someway "due." Right to its name, this fallacy supports the success of casinos all around the globe.

But, although there's no device that forces individual random events to even out, it's an extraordinary fact of statistics that proportions of random events do stay approximately even eventually. If you continue flipping a fair coin ad infinitum, the percentage of heads and tails will both approach 50% each. It's viewed incredible that such uniformity should come out of seeming disorder.

In very much the same manner; at a macro level, unpredictable social events show an extraordinary uniformity. Similar to how random movement of molecules in a gas creates even physical features, the unpredictable happenings of millions of human lives join together to create even social features–like suicide statistics that barely transform from yearly.

Therefore, when used well, statistics are just like "social physics." It can be used to form consistent long-term guesses about completely unpredictable events.

The Art of Statistics: How to Learn from Data by David Spiegelhalter Book Review

Statisticians study trends in data to help provide solutions to the questions about the world. When data is reported well, statistical research can enhance storytelling and enlighten the public about significant problems. Unluckily, there are a lot of misleading filters that research has to go through before it gets to the public, as well as scientific journals and the media. As statistical data sneaks into our lives increasingly, there is an increasing must for all of us to enhance our data literacy in order for us to correctly evaluate the findings.

Do not take statistics at face value.

See statistical information as the manner you might see your friends: they're the basis of some amazing stories; however, they're not usually the most precise. Statistical information should be viewed with the same doubt you have to other types of claims, quotes, and facts. Also, anywhere possible, you have to check the origins of statistics behind the headlines in order for you to check how precisely how the information has been reported.

https://goodbooksummary.com/the-art-of-statistics-by-david-spiegelhalter-book-summary/