"contexts": "The ANOVA can be a powerful tool to tame the variance in field experiments or more complex laboratory experiments, as it allows to account for variance in repeated experimental measures of experiments that are built around replicates. The ANOVA is thus the most robust method when it comes to the design of deductive experiments, yet with the availability of more and more data, also inductive data has increasingly been analysed by use of the ANOVA. This was certainly quite alien to the original idea of Fisher, who believed in clear robust designs and rigid testing of hypotheses.",

        "question": "What is the ANOVA a powerful for?",

        "ground\_truths": "Reducing variance in field experiments or to account for variance in repeated experimental measures of experiments that are built around replicates"

"contexts": "Centuries ago, Thomas Bayes proposed a dramatically different approach. Here, an imperfect or a small sample would serve as a basis for statistical inference. Very crudely defined, the two approaches start at exact opposite ends. While frequency statistics demand preconditions such as sample size and a normal distribution for specific statistical tests, Bayesian statistics build on the existing sample size; all calculations base on what is already there. Experts may excuse my dramatic simplification, but one could say that frequentist statistics are top-down thinking, while Bayesian statistics work bottom-up. The history of modern science is widely built on frequentist statistics, which includes such approaches as methodological design, sampling density and replicates, and diverse statistical tests. It is nothing short of a miracle that Bayes proposed the theoretical foundation for the theory named after him more than 250 years ago. Only with the rise of modern computers was this theory explored deeply, and builds the foundation of branches in data science and machine learning. The two approaches are also often coined as objectivists for frequentist probability fellows, and subjectivists for followers of Bayes theorem.",

        "question": "What is the difference between frequentist and Bayesian approaches to probability, and how have they influenced modern science?",

        "ground\_truths": "Thomas Bayes introduced a different approach to probability that relied on small or imperfect samples for statistical inference. Frequentist and Bayesian statistics represent opposite ends of the spectrum, with frequentist methods requiring specific conditions like sample size and a normal distribution, while Bayesian methods work with existing data."

"contexts": "== Cluster Analysis ==\n'''So you decided for a Cluster Analysis - or Classification in general.''' In this approach, you group your data points according to how similar they are, resulting in a tree structure. Check the entry on [[Clustering Methods]] to learn more.\n\nThere is a difference to be made here, dependent on whether you want to classify the data based on prior knowledge (supervised, Classification) or not (unsupervised, Clustering).",

        "question": "What is Cluster Analysis?",

        "ground\_truths": "Cluster Analysis is a approach of grouping data points based on similarity to create a structure. It can be supervised (Classification) or unsupervised (Clustering)."

"contexts": "Analysis of covariance (ANCOVA) is a statistical test that compares the means of more than two groups by taking under the control the \"noise\" caused by covariate variable that is not of experimental interest. This is done in order to see the true effect of the variable of interest on a dependent variable.",

        "question": "What is ANCOVA in statistical analysis?",

        "ground\_truths": "ANCOVA (analysis of covariance) is a statistical test that compares the means of two or more groups, while treating the covariance as a noise into account. "

"contexts": "Regression assumptions\n# The relationship between dependent and independent variables must be linear for each treatment group.\nANOVA assumptions\n# Variances between groups are homogeneous.\n# Residuals are randomly and normally distributed.\nSpecific ANCOVA assumptions\n# A further specific (but optional) assumption is homogeneity of slopes. It is optional because it is only required to simplify the model for estimation of adjusted means.",

        "question": "What are the assumptions associated with ANCOVA?",

        "ground\_truths": "ANCOVA assumptions include linearity, homogeneity of variances, normal distribution of residuals, and optionally, homogeneity of slopes."

"contexts": "Telemetry is another method that was further developed in recent years, although it has been used already for decades in wildlife ecology. Telemetry is \u201cthe system of determining information about an animal through the use of radio signals from or to a device carried by the animal\u201d (11). For birds, this method can be applied in areas ranging in size from restricted breeding territories of resident bird species to movement patterns of international migratory species. Also, the distribution patterns of infectious diseases of migratory species can be tracked (11). However, for some birds, negative effects on nesting behavior were observed (12).",

        "question": "What is telemetry?",

        "ground\_truths": "Telemetry is a method used in wildlife ecology that uses radio signals to gather information about an animal."

"contexts": "The most abundant reason for a deviance from the normal distribution is us. We changed the planet and ourselves, creating effects that may change everything, up to the normal distribution. Take weight. Today the human population shows a very complex pattern in terms of weight distribution across the globe, and there are many reasons why the weight distribution does not follow a normal distribution. There is no such thing as a normal weight, but studies from indigenous communities show a normal distribution in the weight found in their populations. Within our wider world, this is clearly different.",

        "question": "What is a common reason for deviation from the normal distribution?",

        "ground\_truths": "A common reason for deviation from the normal distribution is human actions, which have caused changes in patterns such as weight distribution."

"contexts": "System thinking gained prominence during the last century, allowing to take interactions and interdependencies into account when investigating a system. To this end, a \u2018system\u2019 stands for something that is in the focus of the investigation, such as a catchment area, a business enterprise, a social institution, or a population of wasps. Such systems show signs of complexity, which means that the interactions of smaller entities in the system may be unpredictable or chaotic, which makes systems more than the sum of their parts. Under this definition, solving problems within a system is often the greatest challenge: while the system dynamics may be understandable, the solutions for an emerging problem are not instantly available. Consequently, solutions for complex problems emerge, and this demands a new line of thinking about systems.",

        "question": "What is system thinking?",

        "ground\_truths": "System thinking is a method of investigation that considers interactions and interdependencies within a system, instead of breaking it into parts."

"contexts": "Generalized Linear Models are statistical analyses, yet the dependencies of these models often translate into specific sampling designs that make these statistical approaches already a part of an inherent and often deductive methodological approach. Such advanced designs are among the highest art of quantitative deductive research designs, yet Generalized Linear Models are used to initially check/inspect data within inductive analysis as well. Generalized Linear Models calculate dependent variables that can consist of count data, binary data, and are also able to calculate data that represents proportions. Mechanically speaking, Generalized Linear Models are able to calculate relations between continuous variables where the dependent variable deviates from the normal distribution. The calculation of GLMs is possible with many common statistical software solutions such as R and SPSS. Generalized Linear Models thus represent a powerful means of calculation that can be seen as a necessary part of the toolbox of an advanced statistician.",

        "question": "What types of data can Generalized Linear Models handle and calculate?",

        "ground\_truths": "Generalized Linear Models can handle and calculate dependent variables that can be count data, binary data, or proportions. It also can calculate continuous variables that deviates from the normal distribution."

"contexts": "A heatmap is a graphical representation of data where the individual numerical values are substituted with colored cells. In other words, it is a table with colors in place of numbers. As in the regular table, in the heatmap, each column is a feature and each row is an observation. Heatmaps are useful to get an overall understanding of the data. While it can be hard to look at the table of numbers it is much easier to perceive the colors. Thus it can be easily seen which value is larger or smaller in comparison to others and how they are generally distributed.",

        "question": "What is a heatmap and why is it useful?",

        "ground\_truths": "A heatmap is a graphical representation of data where numerical values are replaced with colors. It is useful for understanding data as it allows for easy comparison of values and their distribution."

 "contexts": "Many concrete steps brought us closer to the concrete application of scientific methods, among them - notably - the approach of controlled testing by the Arabian mathematician and astronomer Alhazen (a.k.a. Ibn al-Haytham). Emerging from the mathematics of antiquity and combining this with the generally rising investigation of physics, Alhazen was the first to manipulate experimental conditions in a systematic sense, thereby paving the way towards the scientific method, which would emerge centuries later. Alhazen is also relevant because he could be considered a polymath, highlighting the rise of more knowledge that actually enabled such characters, but still being too far away from the true formation of the diverse canon of scientific disciplines, which would probably have welcomed him as an expert on one matter or the other. Of course Alhazen stands here only as one of many that formed the rise of science on the Islamic world during medieval times, which can be seen as a cradle of Western science, and also as a continuity from the antics, when in Europe much of the immediate Greek and Roman heritage was lost.",

        "question": "How did Alhazen contribute to the development of scientific methods?",

        "ground\_truths": "Alhazen contributed to the development of scientific methods by being the first to systematically manipulate experimental conditions, paving the way for the scientific method."

"contexts": "Traditionally in the fields of computer science and mathematics, you have some input and some rule in form of a function. You feed the input into the rule and get some output. In this setting, you need to know the rule exactly in order to create a system that gives you the outputs that you need. This works quite well in many situations where the nature of inputs and/or the nature of the outputs is well understood. However, in situations where the inputs can be noisy or the outputs are expected to be different in each case, you cannot hand-craft the rules that account for every type of inputs or for every type of output imaginable. In such a scenario, another powerful approach can be applied: Machine Learning.",

        "question": "What is the advantage of using Machine Learning over traditional rules or functions in computer science and mathematics?",

        "ground\_truths": "Machine Learning can handle scenarios where inputs are noisy or outputs vary, which is not feasible with traditional rules or functions."

        "contexts": "We define ''Scientific Methods'' as follows: \* First and foremost, scientific methods produce knowledge. \* Focussing on the academic perspective, scientific methods can be either '''reproducible''' and '''learnable'''; can be '''documented''' and are '''learnable'''; or are '''reproducible''', can be '''documented''', and are '''learnable'''. \* From a systematic perspective, methods are approaches that help us '''gather''' data, '''analyse''' data, and/or '''interpret''' it. Most methods refer to either one or two of these steps, and few methods refer to all three steps. \* Many specific methods can be differentiated into different schools of thinking, and many methods have finer differentiations or specifications in an often hierarchical fashion. These two factors make a fine but systematic overview of all methods an almost Herculean task, yet on a broader scale it is quite possible to gain an overview of the methodological canon of science within a few years, if you put some efforts into it. This Wiki tries to develop the baseline material for such an overview, yet can of course not replace practical application of methods and the continuous exploring of empirical studies within the scientific literature.",

        "question": "What are the characteristics of scientific methods?",

        "ground\_truths": "Scientific methods are reproducible, learnable, and documentable. They help in gathering, analyzing, and interpreting data. They can be differentiated into different schools of thinking and have finer differentiations or specifications."

 "contexts": "Mixed Effect Models were a continuation of Fisher's introduction of random factors into the Analysis of Variance. Fisher saw the necessity not only to focus on what we want to know in a statistical design, but also what information we likely want to minimize in terms of their impact on the results. Fisher's experiments on agricultural fields focused on taming variance within experiments through the use of replicates, yet he was strongly aware that underlying factors such as different agricultural fields and their unique combinations of environmental factors would inadvertedly impact the results. He thus developed the random factor implementation, and Charles Roy Henderson took this to the next level by creating the necessary calculations to allow for linear unbiased estimates. These approaches allowed for the development of previously unmatched designs in terms of the complexity of hypotheses that could be tested, and also opened the door to the analysis of complex datasets that are beyond the sphere of purely deductively designed datasets. It is thus not surprising that Mixed Effect Models rose to prominence in such diverse disciplines as psychology, social science, physics, and ecology.",

        "question": "Who developed the calculations that allowed for linear unbiased estimates in Mixed Effect Models?",

        "ground\_truths": "Charles Roy Henderson developed the calculations that allowed for linear unbiased estimates in Mixed Effect Models."

"contexts": "Pomodoro is a method to self-organize, avoid losing yourself, stay productive and energized. Pomodoro is very simple. All you need is work to be done and a timer. There are six steps in the technique: Decide on the task to be done. Set the pomodoro timer (traditionally to 25 minutes = 1 Pomodoro). Work on the task. End work when the timer rings (Optionally: put a checkmark on a piece of paper). If you have fewer than four checkmarks (i.e. done less than 4 Pomodoros) , take a short break (5 minutes), then go back to step 2. After four pomodoros, take a longer break (15\u201330 minutes), reset your checkmark count to zero, then start again at step 1.",

        "question": "How does the Pomodoro technique work?",

        "ground\_truths": "The Pomodoro technique works by deciding on a task, setting a timer for 25 minutes, working on the task until the timer rings, taking a short break if fewer than four intervals have been completed, and taking a longer break after four intervals, then resetting the count and starting again."

 "contexts": "Thought Experiments are the philosophical method that asks the \"What if\" questions in a systematic sense. Thought Experiments are typically designed in a way that should question our assumptions about the world. They are thus typically deeply normative, and can be transformative. Thought Experiments can unleash transformation knowledge in people since such experiments question the status quo of our understanding of the world. The word \"experiment\" is insofar slightly misleading, as the outcome of Thought Experiments is typically open. In other words, there is no right or wrong answer, but instead, the experiments are a form of open discourse. While thus some Thought Experiments may be designed to imply a presumpted answer, many famous Thought Experiments are completely open, and potential answers reflect the underlying norms and moral constructs of people. Hence Thought Experiments are not only normative in their design, but especially in terms of the possible answers of results.",

        "question": "What is the purpose of Thought Experiments?",

        "ground\_truths": "The purpose of Thought Experiments is to systematically ask \"What if\" questions, challenging our assumptions about the world and potentially transforming our understanding of it."

"contexts": "Temporal autocorrelation. This principle defines that we humans value everything that occurs in the close past or future to be more relevant than occurrences in the distant future or past. This is even independent of the likelihood whether future events will actually happen. As an example, imagine that you want to have a new computer every few years, and you can pay 5\u20ac to have a 50% chance to get a new computer tomorrow. If you actually needed one, you would surely do that. However, even adjusted for inflation, many people would not pay the same 5\u20ac to have a 50 % chance to win the latest computer in 10 years. What difference does it make? It is the latest tech in any case. Temporal discounting is one of the main ways how people act unreasonably. This even extends well beyond the point where we are dead already, although this also plays a role. Our unreasonable inabilities to transcend temporal discounting extend likewise into the past. The strongest argument to this end is how many people insist that they are worse off today when compared to the past - in the 'good old days'. The mistake that is typically made is that some aspects in the past were better, while many other aspects were worse. People hence have a tendency to value the time period closer to them different from the time periods more distant. For the past, only the good things are remembered, and for the present, the negative sides are also acknowledged. It widely depends on other circumstances whether this unreasonable evaluation is better or worse than a more objective view on each time period. In any case, the sheer difference between how people evaluate distant time periods or events compared to closer ones is one indirect motivation for science to put this knowledge into perspective.",

        "question": "What is temporal autocorrelation?",

        "ground\_truths": "Temporal autocorrelation is a principle that states humans value events in the near past or future more than those in the distant past or future."