Lab | Reading About Statistic Concepts

Challenge 1: What is the difference between expected value and mean?

You know both concepts but, is there a difference? Are they synonyms?

Mean: If you have a collection of numbers for instance 1,2,...6 (rolling die) the average is a single (3,5) number that describes the whole collection.
The **Mean** of a random variable is the long-term average of its possible values over the **entire population of individuals** (or trials). It's found by taking the weighted average of the x-values multiplied by their probabilities.

Expected value: In probability theory, the expected value of a random variable is a key aspect of its probability distribution. If I want to find the average of a random variable it is also called expected value. The expected value is defined as the weighted average of the values in the range.
The **Expected value** is often referred to as the "long-term" average or mean. This means that over the long term of doing an experiment over and over, you would expect this average.

- ➔ In the end, both Expected value and Mean are very similar. The Mean in general is the average of an entire population of individuals. The Expected value can be the mean but it is only the estimated mean of a sample or experiment, the average of the outcomes. We compute the sample average on a given set of random variables (sample), that is a set of outcomes of a distribution.

Challenge 2: What is the "problem" in science with p-values?

The general problem is that null hypothesis significance testing has several shortcomings. We should not rely on all or nothing hypothesis tests. The authors of the article argue that: "We should never conclude there is 'no difference' or 'no association' just because a P value is larger than a threshold such as 0.05 or, equivalently, because a confidence interval includes zero. Neither should we conclude that two studies conflict because one had a

statistically significant result and the other did not". They explain, that this statistically approach to solve complex problems can lead to several conflicts. Different inferential methods may be most suitable for different types of research questions. Although, "The same problems are likely to arise under any proposed statistical alternative that involves dichotomization". Besides that, it is recommended that whenever researchers use null hypothesis, they should justify its use. It might also be helpful to publish pre-study power calculations and effect sizes, including negative findings. A pre-registration of hypothesis-testing studies is also recommended because some researchers might be encouraged to choose data and methods that "yield statistical significance for some desired (or simply publishable) result".
The researchers try to make clear that they don't want to abandon the p-values. Instead the values shouldn't be treated in categories.
Uncertainty should not be an issue. "One practical way to do so is to rename confidence intervals as 'compatibility intervals' and interpret them in a way that avoids overconfidence". Another idea is to consider more than one model for the research and to report all results.

Challenge 3: Applying testing to a specific case: A/B testing.

Basecamp example:

The company noticed fewer sign ups after changing the marketing site. If they would have done an AB test, they probably wouldn't have lost some profit.

In web analytics, the idea is to challenge an existing version of a website (A) with a new one (B), by randomly splitting traffic and comparing metrics on each of the splits.

The company Basecamp could have used this experiment technique to determine whether the new marketing site will bring more people to sign up or not.

The metrics that matter to the company are the average sign-ups during a certain amount of time.

H0: The average sign up from the old page is equal to the average amount of sign ups from the new page.

Alternative Hypothesis: The average amount of sign-ups is different from the new to the old page.

Since the author of the article mentioned that the company has regular seasonal effects, it would be good to consider them in the tests as well.

Then we could use a z-test to check if the average amount of sign ups is the same.