**Sampling Distributions Notes**

We have already learned some really valuable ideas about sampling distributions:

First, we have defined **sampling distributions** as **the distribution of a statistic**.

This is fundamental - I cannot stress the importance of this idea. We simulated the creation of sampling distributions in the previous ipython notebook for samples of size 5 and size 20, which is something you will do more than once in the upcoming concepts and lessons.

Second, we found out some interesting ideas about sampling distributions that will be iterated later in this lesson as well. We found that for proportions (and also means, as proportions are just the mean of 1 and 0 values), the following characteristics hold.

1. The sampling distribution is centered on the original parameter value.
2. The sampling distribution decreases its variance depending on the sample size used. Specifically, the variance of the sampling distribution is equal to the variance of the original data divided by the sample size used. This is always true for the variance of a sample mean!

In notation, we say if we have a random variable, \bold{X}X, with variance of \bold{\sigma^2}σ2, then the distribution of \bold{\bar{X}}X¯ (the sampling distribution of the sample mean) has a variance of \bold{\frac{\sigma^2}{n}}nσ2​

**Looking Ahead**

The rest of this lesson will reinforce some of these ideas that you saw at work in this notebook, but you are already being introduced to some big ideas that will continue to show up again and again.

d be a 'hat' on the \sigma^2*σ*2 in the statistics side at 0:47 (i.e. \hat{\sigma}^2*σ*^2).

As you saw in this video, we commonly use Greek symbols as parameters and lowercase letters as the corresponding statistics. Sometimes in the literature, you might also see the same Greek symbols with a "hat" to represent that this is an estimate of the corresponding parameter.

Below is a table that provides some of the most common parameters and corresponding statistics, as shown in the video.

Remember that all **parameters** pertain to a population, while all **statistics** pertain to a sample.

| **Parameter** | **Statistic** | **Description** |
| --- | --- | --- |
| \mu*μ* | \bar{x}*x*¯ | "The mean of a dataset" |
| \pi*π* | p*p* | "The mean of a dataset with only 0 and 1 values - a proportion" |
| \mu\_1 - \mu\_2*μ*1​−*μ*2​ | \bar{x}\_1-\bar{x}\_2*x*¯1​−*x*¯2​ | "The difference in means" |
| \pi\_1 - \pi\_2*π*1​−*π*2​ | p\_1-p\_2*p*1​−*p*2​ | "The difference in proportions" |
| \beta*β* | b*b* | "A regression coefficient - frequently used with subscripts" |
| \sigma*σ* | s*s* | "The standard deviation" |
| \sigma^2*σ*2 | s^2*s*2 | "The variance" |
| \rho*ρ* | r*r* | "The correlation coefficient" |

NEXT

Two important mathematical theorems for working with sampling distributions include:

1. **Law of Large Numbers**
2. **Central Limit Theorem**

The **Law of Large Numbers** says that **as our sample size increases, the sample mean gets closer to the population mean**, but how did we determine that the sample mean would estimate a population mean in the first place? How would we identify another relationship between parameter and statistic like this in the future?

Three of the most common ways are with the following estimation techniques:

* [**Maximum Likelihood Estimation**](https://en.wikipedia.org/wiki/Maximum_likelihood_estimation)
* [Method of Moments Estimation][**[https://en.wikipedia.org/wiki/Method\_of\_moments]](https://en.wikipedia.org/wiki/Method_of_moments%5d" \t "_blank)**
* [**Bayesian Estimation**](https://en.wikipedia.org/wiki/Bayes_estimator)

Though these are beyond the scope of what is covered in this course, these are techniques that should be well understood for Data Scientist's that may need to understand how to estimate some value that isn't as common as a mean or variance. Using one of these methods to determine a "best estimate", would be a necessity.

NEXT

The **Central Limit Theorem** states that **with a large enough sample size the sampling distribution of the mean will be normally distributed**.

The **Central Limit Theorem** actually applies for these well known statistics:

1. Sample means (\bar{x}*x*¯)
2. Sample proportions (p*p*)
3. Difference in sample means (\bar{x}\_1 - \bar{x}\_2*x*¯1​−*x*¯2​)
4. Difference in sample proportions (p\_1 - p\_2*p*1​−*p*2​)

And it applies for additional statistics, **but it doesn't apply for all statistics!** . You will see more on this towards the end of this lesson.

In the previous example, you saw how the **Central Limit Theorem** applies to the sample mean of 100 draws from a right-skewed distribution. However, it did not apply to a sample size of 3 draws from this same distribution.

In the next concepts, you will see that the with large sample sizes the sampling distribution of certain statistics will never become normally distributed. So how do we know which statistics will follow normal distributions, and which will not?

So, you might be wondering already why is the **Central Limit Theorem** such a big deal? In our new age of computers, it probably isn't as big of a deal, but more on this coming up soon!

**Central Limit Theorem - Part III**

You saw how the **Central Limit Theorem** worked for the sample mean in the earlier concept. The **Central Limit Theorem** states that **with a large enough sample size the sampling distribution of the mean will be normally distributed**.

The **Central Limit Theorem** actually applies for these well known statistics:

1. Sample means (\bar{x}*x*¯)
2. Sample proportions (p*p*)
3. Difference in sample means (\bar{x}\_1 - \bar{x}\_2*x*¯1​−*x*¯2​)
4. Difference in sample proportions (p\_1 - p\_2*p*1​−*p*2​)

And it applies for additional statistics, **but it doesn't apply for all statistics!** . Here, you will simulate the sampling distribution for the sample variance. Try out the notebook and quizzes.

**Bootstrapping** is sampling with replacement. Using **random.choice** in python actually samples in this way. Where the probability of any number in our set stays the same regardless of how many times it has been chosen. Flipping a coin and rolling a die are kind of like bootstrap sampling as well, as rolling a 6 in one scenario doesn't mean that 6 is less likely later.

ou actually have been bootstrapping to create sampling distributions in earlier parts of this lesson, but this can be extended to a bigger idea.

It turns out, we can do a pretty good job of finding out where a parameter is by using a sampling distribution created from bootstrapping from only a sample. This will be covered in depth in the next lessons.

Three of the most common ways are with the following estimation techniques for finding "good statistics" are as shown previously:

* [**Maximum Likelihood Estimation**](https://en.wikipedia.org/wiki/Maximum_likelihood_estimation)
* [**Method of Moments Estimation**](https://onlinecourses.science.psu.edu/stat414/node/193)
* [**Bayesian Estimation**](https://en.wikipedia.org/wiki/Bayes_estimator)

Though these are beyond the scope of what is covered in this course, these are techniques that should be well understood for data scientists who may need to understand how to estimate some value that isn't as common as a mean or variance. Using one of these methods to determine a "best estimate" would be a necessity.

NEXT

wo helpful links:

* You can learn more about Bradley Efron [**here**](https://en.wikipedia.org/wiki/Bradley_Efron).
* Additional notes on why bootstrapping works as a technique for inference can be found [**here**](https://stats.stackexchange.com/questions/26088/explaining-to-laypeople-why-bootstrapping-works).