STAT 2593

Lecture 022 - The Distribution of the Sample Mean

Dylan Spicker

The Distribution of the Sample Mean

Learning Objectives

1. Describe the mean and the variance of the sampling distribution for the sample mean.

2. Describe the limiting behaviour for the sampling distribution of the sample mean.


sodastream

The Sampling Distribution of the Sample Mean

- Suppose that we take $\overline{X}$ to represent the random variable of the sample mean.

The Sampling Distribution of the Sample Mean

- Suppose that we take $\overline{X}$ to represent the random variable of the sample mean.
- Regardless of the distribution of the underlying data we have

The Sampling Distribution of the Sample Mean

- Suppose that we take $\overline{X}$ to represent the random variable of the sample mean.
- Regardless of the distribution of the underlying data we have
  - $E[\overline{X}] = E[X] = \mu$.

The Sampling Distribution of the Sample Mean

- Suppose that we take $\overline{X}$ to represent the random variable of the sample mean.
- Regardless of the distribution of the underlying data we have
  - $E[\overline{X}] = E[X] = \mu$.
  - $\text{var}(\overline{X}) = \frac{1}{n}\text{var}(X) = \frac{\sigma^2}{n}$.

The Sampling Distribution of the Sample Mean

- Suppose that we take $\overline{X}$ to represent the random variable of the sample mean.
- Regardless of the distribution of the underlying data we have
  - $E[\overline{X}] = E[X] = \mu$.
  - $\text{var}(\overline{X}) = \frac{1}{n}\text{var}(X) = \frac{\sigma^2}{n}$.
  - Often, we will consider the standard deviation which is $\frac{\sigma}{\sqrt{n}}$. We call this the **standard error** in this context.

The Sampling Distribution of the Sample Mean

- Suppose that we take $\overline{X}$ to represent the random variable of the sample mean.
- Regardless of the distribution of the underlying data we have
  - $E[\overline{X}] = E[X] = \mu$.
  - $\text{var}(\overline{X}) = \frac{1}{n}\text{var}(X) = \frac{\sigma^2}{n}$.
  - Often, we will consider the standard deviation which is $\frac{\sigma}{\sqrt{n}}$. We call this the **standard error** in this context.
- If we consider totals instead of means we get $E[T] = n\mu$ and $\text{var}(T) = n\sigma^2$.

Limiting Distributions

- Note that the sample mean will always have the same mean as the underlying distribution.

Limiting Distributions

- Note that the sample mean will always have the same mean as the underlying distribution.
- As the sample size grows, $n \to \infty$, the variance of the sample mean goes to zero.

Limiting Distributions

- Note that the sample mean will always have the same mean as the underlying distribution.
- As the sample size grows, $n \to \infty$, the variance of the sample mean goes to zero.
  - As a result, the sample mean ends up centered exactly on the underlying mean, in large enough samples.

Limiting Distributions

- Note that the sample mean will always have the same mean as the underlying distribution.
- As the sample size grows, $n \to \infty$, the variance of the sample mean goes to zero.
  - As a result, the sample mean ends up centered exactly on the underlying mean, in large enough samples.
- If the underlying data follows a $N(\mu, \sigma^2)$ then the sample mean will also be **exactly** normally distributed.

Limiting Distributions

- Note that the sample mean will always have the same mean as the underlying distribution.
- As the sample size grows, $n \to \infty$, the variance of the sample mean goes to zero.
  - As a result, the sample mean ends up centered exactly on the underlying mean, in large enough samples.
- If the underlying data follows a $N(\mu, \sigma^2)$ then the sample mean will also be **exactly** normally distributed.
- If the underlying data are non-normal, the sample mean is **not exactly** normally distributed.

Limiting Distributions

- Note that the sample mean will always have the same mean as the underlying distribution.
- As the sample size grows, $n \to \infty$, the variance of the sample mean goes to zero.
  - As a result, the sample mean ends up centered exactly on the underlying mean, in large enough samples.
- If the underlying data follows a $N(\mu, \sigma^2)$ then the sample mean will also be **exactly** normally distributed.
- If the underlying data are non-normal, the sample mean is **not exactly** normally distributed.
  - However, in large enough samples, it will be *approximately* normal.

Limiting Distributions

- Note that the sample mean will always have the same mean as the underlying distribution.
- As the sample size grows, $n \to \infty$, the variance of the sample mean goes to zero.
  - As a result, the sample mean ends up centered exactly on the underlying mean, in large enough samples.
- If the underlying data follows a $N(\mu, \sigma^2)$ then the sample mean will also be **exactly** normally distributed.
- If the underlying data are non-normal, the sample mean is **not exactly** normally distributed.
  - However, in large enough samples, it will be *approximately* normal.
  - This is known as the **central limit theorem**.

Summary

- The sample mean has a distribution with expectation equal to the population expectation and variance which goes to zero in n.
- Normal samples will have an exactly normal sampling distribution.
- Non-normal samples will be approximately normally distributed, if the sample size is large enough.