

12.4: The Central Limit Theorem

The Central Limit Theorem tells us that as sample sizes get larger, the sampling distribution of the mean will become normally distributed, *even if the data within each sample are not normally distributed*.

We can see this in real data. Let's work with the variable `AlcoholYear` in the NHANES distribution, which is highly skewed, as shown in the left panel of Figure ???. This distribution is, for lack of a better word, funky – and definitely not normally distributed. Now let's look at the sampling distribution of the mean for this variable. Figure 12.2 shows the sampling distribution for this variable, which is obtained by repeatedly drawing samples of size 50 from the NHANES dataset and taking the mean. Despite the clear non-normality of the original data, the sampling distribution is remarkably close to the normal.

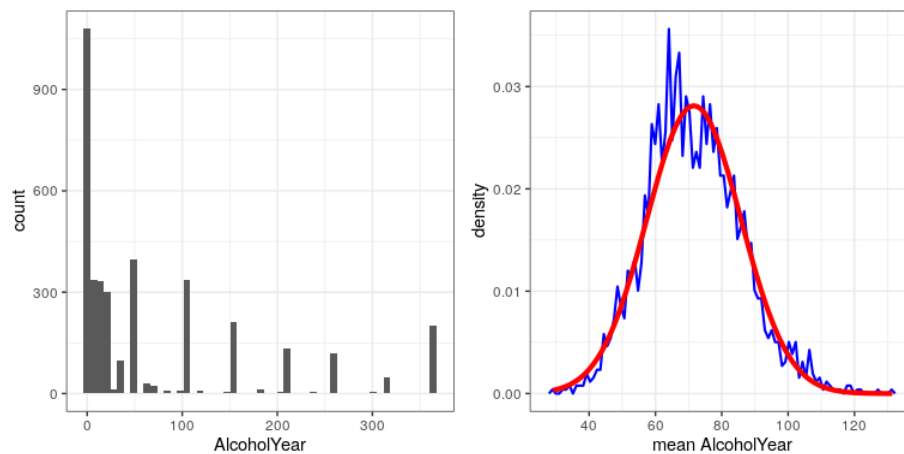


Figure 12.2: Left: Distribution of the variable `AlcoholYear` in the NHANES dataset, which reflects the number of days that the individual drank in a year. Right: The sampling distribution of the mean for `AlcoholYear` in the NHANES dataset, obtained by drawing repeated samples of size 50, in blue. The normal distribution with the same mean and standard deviation is shown in red.

The Central Limit Theorem is important for statistics because it allows us to safely assume that the sampling distribution of the mean will be normal in most cases. This means that we can take advantage of statistical techniques that assume a normal distribution, as we will see in the next section.

This page titled [12.4: The Central Limit Theorem](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Russell A. Poldrack](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.