Sampling Distributions

This is the distribution of individuals (taken one at a time). It has long tails because individuals vary a lot.

This distribution has $\mu = 100$ and $\sigma = 60$

---

This is a sampling distribution for $\overline{x}$.

It shows the distribution of sample means when samples are taken 9 at a time. (In other words, $n=9$.)

The center is still 100, but it’s less spread out. The standard deviation is 20 instead of 60.

This distribution has $\mu_{\overline{x}} = 100$ and $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{60}{\sqrt{9}} = 20$

---

This is another sampling distribution for $\overline{x}$.

It shows the distribution of sample means when samples are taken 36 at a time. (In other words, $n=36$.)

The center is still 100, but the standard deviation is 10 (so it’s even less spread out).

This happens because the samples “average out” to the middle.

For example, imagine a distribution of people’s heights.

You might sample one tall person fairly often, but you’re not likely to sample 36 tall people at once.

This distribution has $\mu_{\overline{x}} = 100$ and $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{60}{\sqrt{36}} = 10$

**The main idea:**
As the samples get larger, the standard deviation gets smaller.

This is important because all of the statistics that we do with hypothesis tests and confidence intervals are based on sampling distributions like this.

Created by David Grimes, HCC, 2019
The Central Limit Theorem (for the sample mean $\bar{x}$):

When the population is normal, the sampling distribution of $\bar{x}$ is normal too. However, if the population is “shaped funny,” the sampling distribution will be “shaped funny” too... unless we take large samples (30 or more). The large samples compensate for the weird shape of the population, allowing our sampling distribution to be “normal enough.”

Normal Approximation of Binomial (for the sample proportion $\hat{p}$):

When you take large samples, the binomial distribution starts to look like the normal distribution.

In order for the normal approximation to be good enough, two things must be true:

1. The sample size ($n$) can’t be more than 5% of the population size ($N$): $n \leq 0.05N$
2. When you plug in the $n$ and $p$ numbers, this expression must be 10 or bigger: $np(1-p) \geq 10$

The main idea: The normal distribution will give approximately the same answers as the binomial distribution. Any time we do a hypothesis test or a confidence interval for the proportion, we are using the normal approximation.

Example:
Suppose we have a binomial problem with $n=60$ and $p=0.4$ (60 trials and a 0.4=40% chance of success each time). What is the probability of getting more than 30 out of 60? The binomial calculation is shown on the left.

For the normal approximation, we use $\mu_{\hat{p}} = p = 0.4$ as the mean of sample proportions and $\sigma_{\hat{p}} = \sqrt{\frac{0.4(1-0.4)}{60}} = 0.06324$ as the standard deviation of sample proportions.

“30 out of 60” is represented by the sample proportion $\hat{p} = \frac{30}{60} = 0.5$. A normal probability histogram is shown on the right.