Lecture 3: Stats Week!

Sampling Distributions & 1 Sample Z Test for Proportions

Sample Distributions and Sampling Distributions

Sample vs. Sampling Distributions

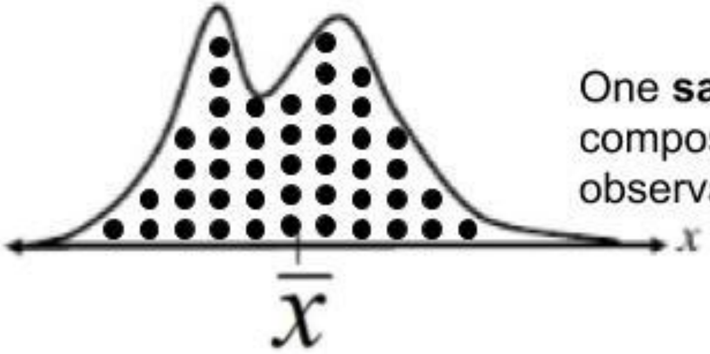
Confusing sample distribution with sampling distribution happens a lot, but it's a huge error. They are fundamentally different and that difference is *very* important.

- **Sample Distribution**: distributions built from the sample you've taken. These are what we've seen so far. You have a set of data and you plot it - that's a sample distribution. Each point on the dotplot stands for an individual observation.
- **Sampling Distribution**: distributions that allow us to see how statistics describing samples (like x-bar and s) vary from sample to sample. Think of a sampling distribution as a distribution of many *sample distributions*, or at least some statistic describing these sample distributions.

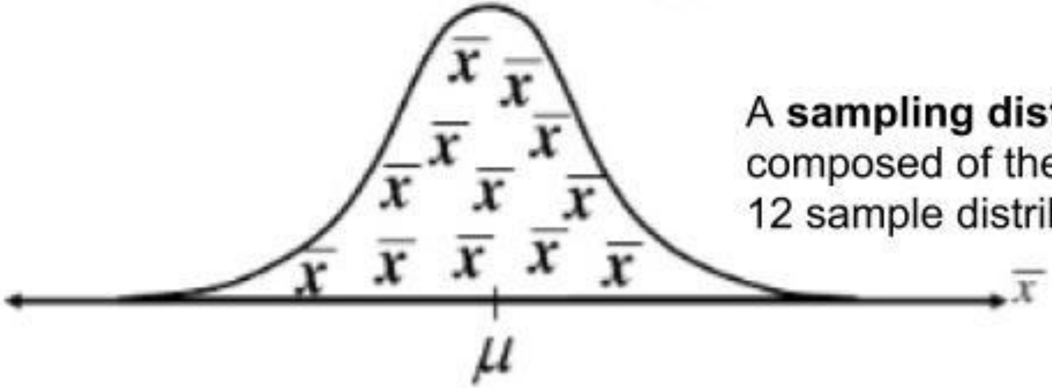
Individual Activity #1

1. Navigate to rolladie.net
2. Click “roll” to roll the die. Record the outcome. Repeat until you have 10 values.
3. Create a dotplot of your ten values.
4. Find the mean of your distribution.
5. Tell Kelly your mean.

Sample vs. Sampling Distribution



One **sample distribution** composed of 44 individual observations



A **sampling distribution** composed of the means of 12 sample distributions

Sampling Distributions for Proportions

What's a proportion?

- A proportion is categorical data. It's basically just a like a fraction, where X is the number of successes and n is the total number of trials.

$$\hat{p} = \frac{X}{n}$$

Note: $\hat{p}$ is the notation for a sample or observed proportion

Ex. You flip a coin ten times. Seven of those times you get heads. Your proportion, where heads is successes, is 7/10. What would the proportion be is tails were successes?

Mean and SD for a Proportion

- Just like means, proportions have symbols/formulae for both their means and standard deviations.

	Mean	Standard Deviation
Population Proportion	p	$\sqrt{\frac{pq}{n}}$

Note: p is the probability of success. q is the probability of failure. n is the number of trials.

Z-Scores for Proportions

- Z-Scores for Proportions are calculated very similarly to z-scores for means. The notation changes, but the process is the same.

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Mean and SD for Sampling Dist of Proportions

Mean	Standard Deviation
$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$

Checking Conditions for Sampling Dist of Proportions

As long as we meet the...

- **Randomization Condition** (random sample)
- **Success/Failure Condition**, which says that the expected number of successes (n*p) and the expected number of failures (n*q) must be greater than or equal to 10, and
- **10% Condition**, which says that the sample size, n, must be no larger than 10% of the population,

...then the sampling distribution for proportions can be described as:

$$\sim N\left(p, \sqrt{\frac{pq}{n}}\right)$$

Example #1

1. Assume I flip a fair coin (aka P(heads)=.5) 10 times. Find the mean and standard deviation of the sampling distribution of proportions when n=10.
    a. Should a Normal model be used here?
2. Now I flip the coin 20 times. Find the mean and sd of the sampling distribution of proportions when n=20.
    a. Should a Normal model be used here?
3. Using the mean and sd from #2, answer the following:
    a. What is the probability of having 45% or fewer heads?
    b. What is the probability of having between 40% and 65% heads?
    c. What is the probability of having at least 80% heads?

Group Work #1

In groups of 2-3, complete the two problems on the in-class activity page.

1 Sample Hypothesis Test for Proportions

1 Sample Z Test for Proportions (see handout)

- Essentially, what a test statistic is asking is: ASSUMING that your distribution did have a mean of p, how likely is it that you observed the sample mean, p-hat If the test statistic is really big (in absolute value), it is really unlikely that it could have been observed if the true mean was really p.
- If the p-value is smaller than some *alpha* or *significance level*, we will reject the null hypothesis and say the mean is *statistically significantly* different/greater than/less than from the null hypothesis.
- Typically, we use alpha = 0.001, 0.01, 0.05, 0.10

Notes and example 2 on handout.

Group Work #2

In groups of 2-3, consider the questions provided.

Sampling Distribution for Means

Sampling Distribution of Means

Mean	Standard Deviation
$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Note: The SD formula requires that we know the population standard deviation. If we are trying to make predictions about the population mean, it's incredibly unlikely we'll know the population standard deviation.

Sampling Distribution of Means

So… we have to estimate the standard deviation. When we estimate a parameter, it's called the **standard error** rather than the standard deviation.

Mean	**Standard Error**
$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$

In other words, the standard error is usually the standard deviation of the sampling distribution of means.

Checking Conditions for Sampling Dist of Means

As long as we meet the…

- **Randomization Condition** (random sample)
- **10% Condition**, which says that the sample size, n, must be no larger than 10% of the population,
- **Nearly Normal Condition** - satisfiable by (1) population is normal, (2) dataset appears normal, or (3) Large Enough by CLT.

…then the sampling distribution for means can be described as:

Known σ	Unknown σ
$\sim N(\mu, \frac{\sigma}{\sqrt{n}})$	$\sim N(\mu, \frac{s}{\sqrt{n}})$

Example #3

Consider the approximately normal population of heights of male college students with mean of 70 inches and standard deviation of 5 inches.

a. What is the probability that a single randomly selected male college student has a height greater than 73 inches? *Note that we know orig pop is normal.*
b. A random floor of the men's dormitory is selected and the 25 male students on the floor are sampled. What is the mean and standard deviation of the average heights for sample sizes of 25?
c. Find the probability that the sample mean of the 25 college males will be greater than 73 inches.

Group Work #3

In groups of 2-3, answer the two questions on the in-class activity sheet.