AP Statistics: Introduction to Sampling Distribution

Objective: To understand the concept of a Sampling Distribution, introduce proportion, mean, and standard deviation as they relate to a Sampling Distribution of Proportion

Random Coin Toss Activity

A very good introduction into Sampling Distribution:
https://www.youtube.com/watch?v=kPIPOXhSF_E
Porinchak, D. (2014, Jan. 21). AP Statistics: Sampling Distributions Part 2 Examples [video]

Terminology:

A Sampling Distribution is used to predict what might be true for a population.
A parameter is a value that is true for the population.
A statistic is a value the comes from sample.
Sample Variability (sampling error) is the value of any statistic ($\hat{p}$ or $\bar{X}$) varies in repeated random sampling.

Population

↓

Parameter
(fixed quantity)

Proportion p | **Mean** μ

Sample

↓

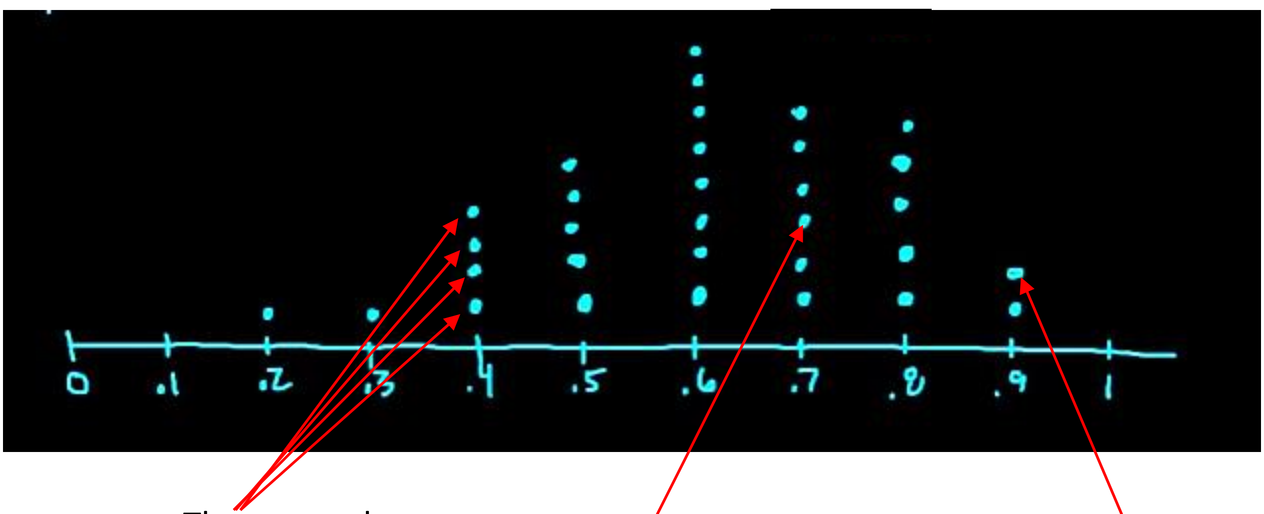
Statistic
(varies based on sample)

Proportion $\hat{p}$ (p-hat) | **Mean** $\bar{X}$ (x-bar)

What does a Sampling Distribution look like?

Consider this example. We have a bag of 200 marbles. Half the marbles are red and half are blue. We reach into the bag and pull out 10 marbles and record what proportion are blue. The Sampling Distribution is shown below:

Each dot on the graph represents an independent sample of 10 marbles from the bag.



These sample had 4 blue marbles

This sample had 7 blue marbles

This sample had 9 blue marbles

Looks a lot like the Normal Curve!

Video demonstration of the differences between a population Distribution, a sample distribution and a sampling distribution

At the 18 minute mark
https://www.youtube.com/watch?v=faM8upJa-Ns
Porinchak, D. (2015, Jan. 15). Sampling Distributions Intro (7.1) [video]

The M&M's web site claims that 24% of all M&M's are blue, 2o% are orange, 16% green, 14 % yellow, 13% are red and 13% are brown.

Draw the population distribution, the sample distribution and the sampling distribution in the space below:

Criteria for using the **Normal Model** for the Sampling Distribution or Proportion:

1. Randomness – the samples must be randomly selected from the population
2. 10% Condition – the samples size must be less than 10% of the population to ensure impendence
3. Is the sample big enough? Are there enough successes and failures in the sample of "n"? (np ≥10 and nq ≥10)

To describe the characteristics of the distribution curve, you need:

1. **Center** (mean) $\mu_{\hat{p}} = p$

   We expect the mean of our sample to be the same as the ~~mean~~ proportion of the population.

2. **Spread** (standard deviation) SD ($\hat{p}$) = $\sqrt{\frac{pq}{n}}$ where p is the probability of success in our sample and q is the probability of failure.

   Our spread vary since proportion will change depending on the sample. Sometimes we pull 3 blue marbles, sometimes we pull 7 blue marbles…

Note: Large samples produce low variability.

Example of Sample Distribution of Proportion

Suppose that about 13% of the population is left-handed. A 200-seat auditorium has been built with 15 "lefty seats," seats that have the built-in desk on the left rather than on the right arm of the chair. In a class of 90 students, what's the probability that there will not be enough seats for the left-handed students?

What is the problem asking?

What is the probability that in a group of 90 students, more than 15% will be left handed? Since 15/90 = 16.7%, I need the probability of finding more than 16.7% left handed students out of a sample of 90 if the proportion of lefties is 13%.

Can we use the Normal model for the distribution curve?

Test the 3 conditions:

- ☑ Randomness –The 90 students in the class can be thought of as a random sample of students.
- ☑ 10% Condition – 90 is definitely less than all students (those that attend that school and those who do not.)
- ☑ Big Enough? np = 90(0.13) = 11.7 ≥ 10 and nq = 90(0.87) = 78.3 ≥ 10

State the parameters and the sampling distribution model:

The population proportion is p=0.13. The conditions are satisfied for using the Normal model. Our sampling distribution of $\hat{p}$ with a Normal model with mean of 0.13 and a standard deviation of :

$$SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.13)(0.87)}{90}} \approx 0.035$$

We now have enough information to calculate the Z-score.

Proportion of the sample

Proportion of the population

$$Z = \frac{\hat{p} - p}{SD(\hat{p})} = \frac{0.167-0.13}{0.035} = 1.06$$

$$P(\hat{p} \geq 0.167) = P(z > 1.06) = 0.1446$$

There is about a 14.5% chance that there will not be enough seats for the left handed students in the class.