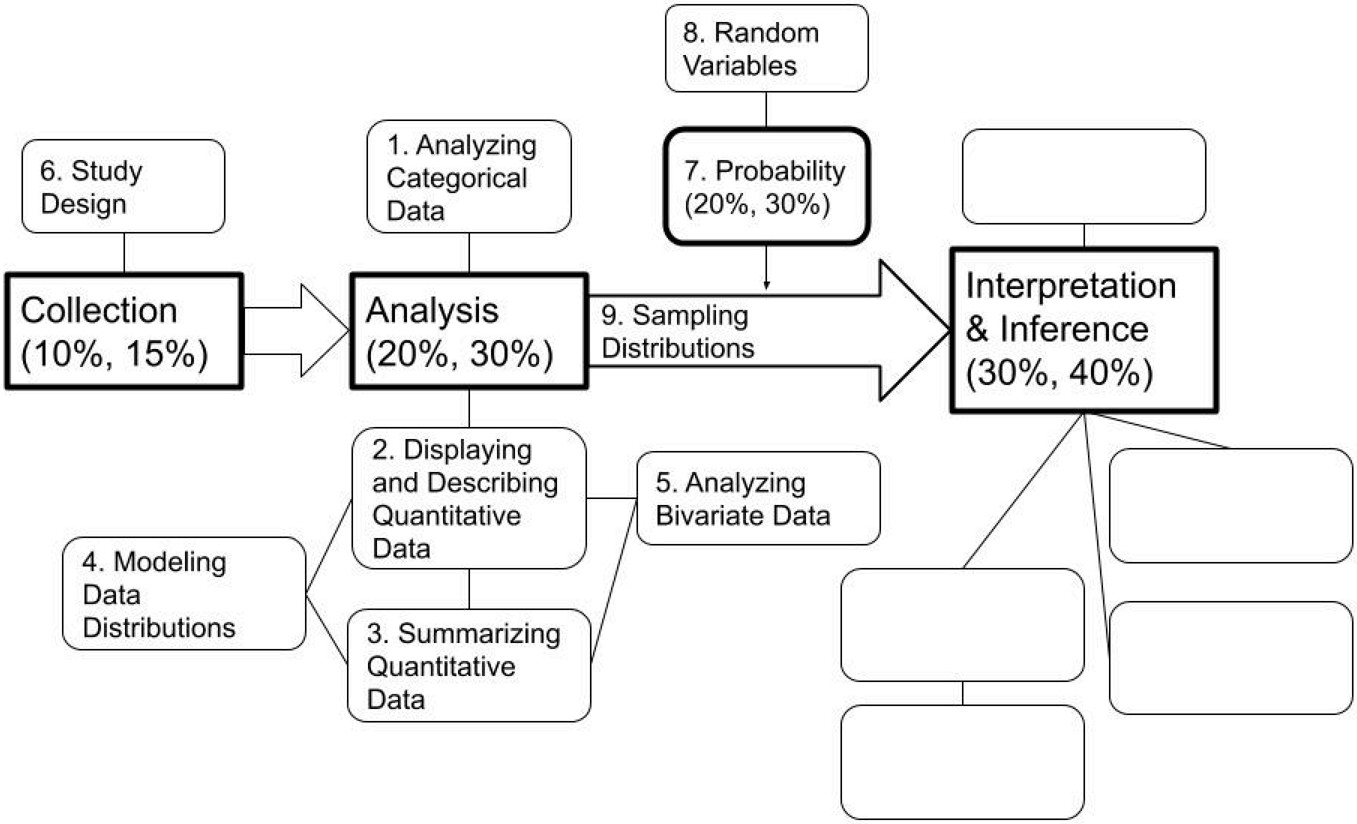
Unit 09: Sampling Distributions

Unit Objectives

- Connect previous unrealistic knowledge to real world situations by understanding sampling distributions

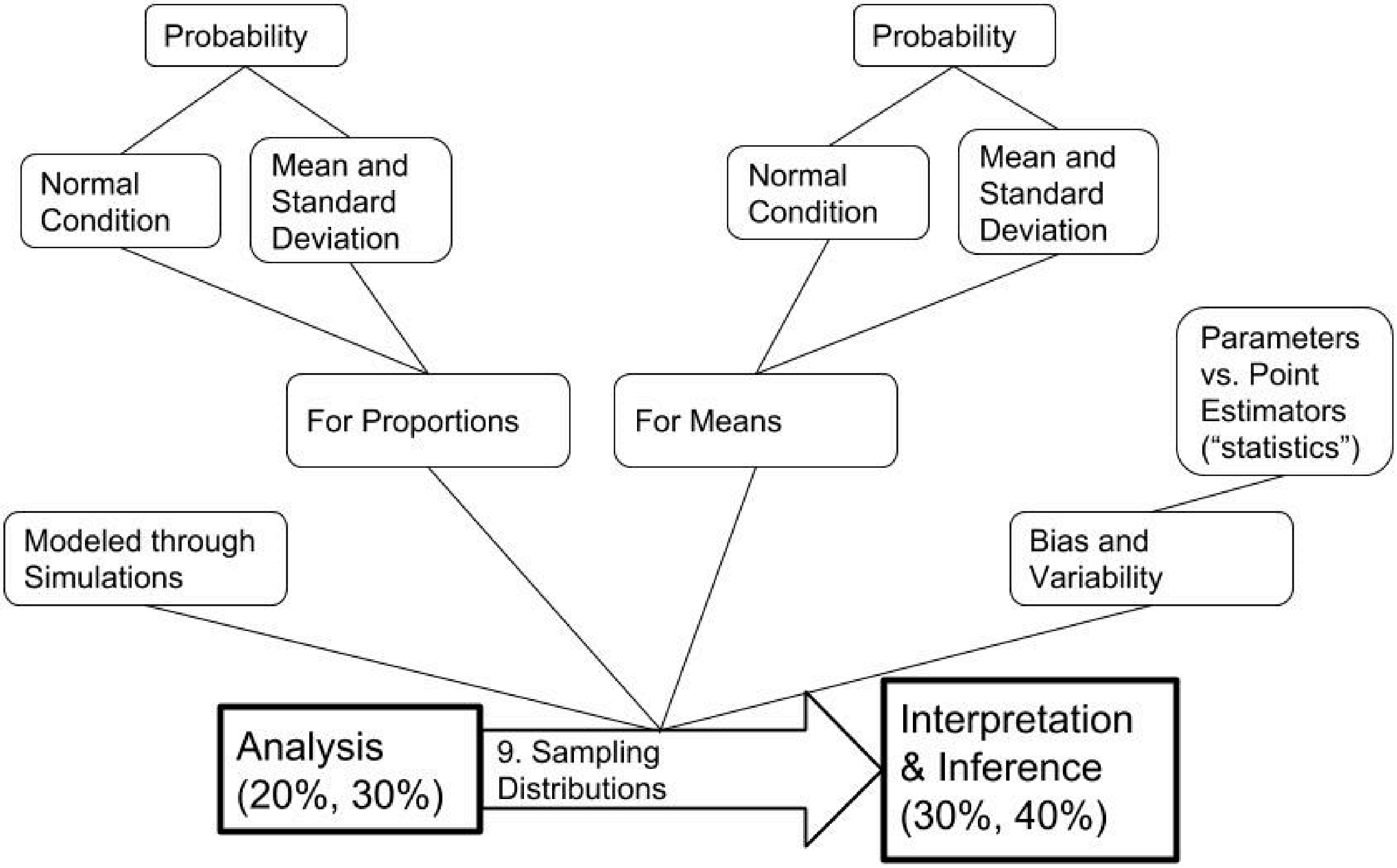

8. Random Variables
6. Study Design
1. Analyzing Categorical Data
7. Probability (20%, 30%)
Collection (10%, 15%)
Analysis (20%, 30%)
9. Sampling Distributions
Interpretation & Inference (30%, 40%)
2. Displaying and Describing Quantitative Data
5. Analyzing Bivariate Data
4. Modeling Data Distributions
3. Summarizing Quantitative Data

The horrible truth(s)

1. If we already knew the mean and standard deviation of a population, we would never waste time and money ______________________ a _________________________.

2. In order to truly know the population mean and population standard deviation, we would have to ______________________ the ___________________ ______________________.

3. The only reason anyone would ever care about a _____________________ distribution is because they want to know the ________________ distribution. **But there is never a guarantee that your sample distribution will be ________________________ of your population distribution. There will always be a chance** that you randomly select too many ____________ values or too many ___________ values from the population.


Probability
Normal Condition
Mean and Standard Deviation
Probability
Normal Condition
Mean and Standard Deviation
Parameters vs. Point Estimators ("statistics")
For Proportions
For Means
Modeled through Simulations
Bias and Variability
Analysis (20%, 30%)
9. Sampling Distributions
Interpretation & Inference (30%, 40%)

Unit 09 Lesson 01: Understand Sampling Distributions and Bias

Wait, how would we ever actually know a population mean or standard deviation? Have I been living a lie?

- Answer questions about sampling distributions given simulation output.
- Identify estimators as biased vs unbiased, and having high variability vs. low variability.

Population Distribution

- SOCS for a ______________________________.
- Values (mean, standard deviation, etc) are called __________________________.

Sample Distribution

- SOCS for a __________________________ of the ___________________________.
- Values (mean, standard deviation, etc) are called ____________________ ______________________ or

  ______________________.
- The purpose of a sample is to use a ____________________ ______________________ /

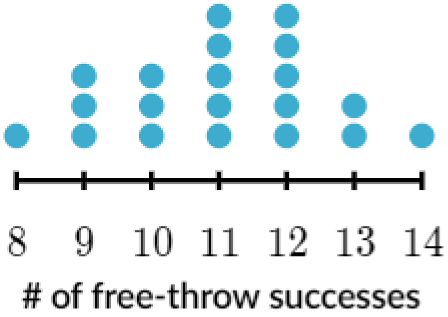
  ______________________ in order to estimate a ___________________________

Sampling Distribution

- SOCS for a ____________________ ______________________ / ______________________ of multiple samples.

Not all of the following questions involve a sampling distribution

1. [Source: Khan Academy] Noaya performed 20 samples of 15 free-throws. Naoya counted how many free-throws were successful in each sample. Here are his results:

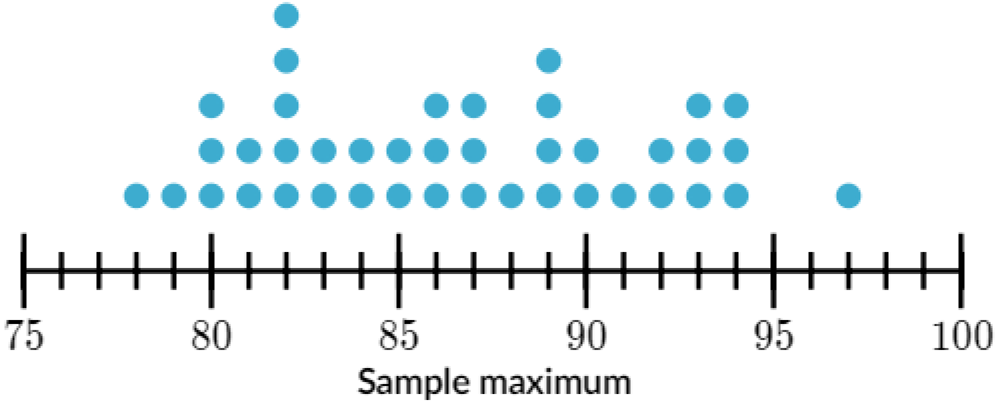


Use his results to estimate the probability that he succeeds at greater than 10 free-throws in a sample of 15 free-throws.

P(greater than 10 successes)≈

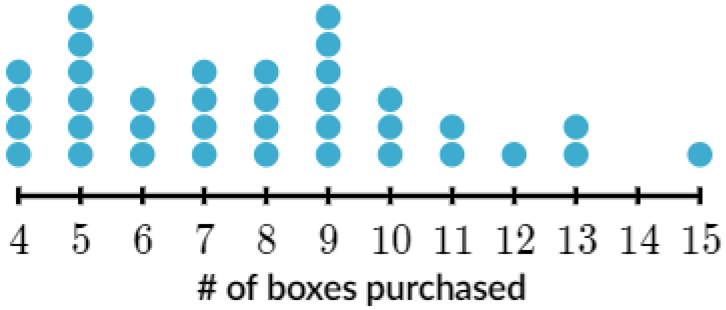
Sampling Bias: Rossman/Chance: Sampling Pennies

2. [Source: Khan Academy] Isabelle was curious if sample maximum was an unbiased estimator of population maximum. She started with a large normally distributed population of test scores whose maximum was 99 points. She then took a random sample of 6 tests and calculated the maximum of the sample. She replaced those tests and repeated this process for a total of 40 trials. Her results are summarized in the dotplot below, where each dot represents the maximum score from a sample of 6 tests.



Based on these results, does sample maximum appear to be a biased or unbiased estimator of population maximum? Why?

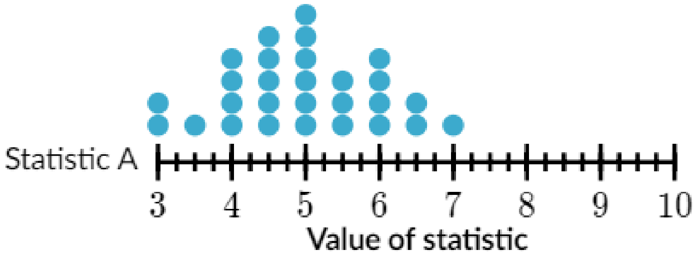
3. [Source: Khan Academy] A cereal company is putting 1 of 4 prizes in each box of cereal. The probability of getting any prize is 0.25. Vera wonders how many boxes she should expect to buy to get all 4 prizes. She and her 35 friends (36 people total) all recorded how many boxes it took before they got all 4 prizes. Each dot represents how many boxes it took to get all 4 prizes for that person.

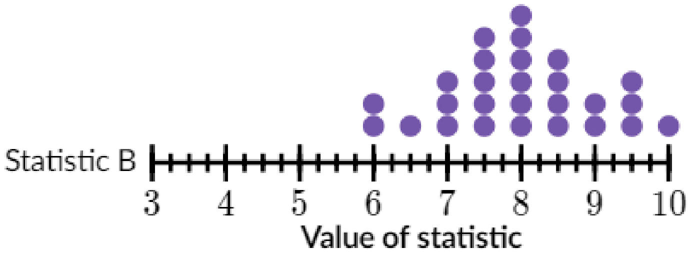


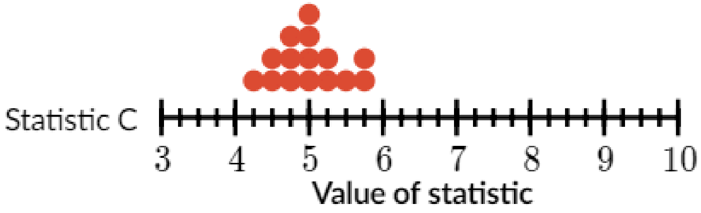
Use her results to estimate the probability that it takes 9 or fewer boxes to get all 4 prizes.

P(9 or fewer boxes)≈

4. [Source: Khan Academy] The dotplots below show an approximation to the sampling distribution for three different estimators of the same population parameter. If the actual value of the population parameter is 5, which dotplot displays the estimator with both low bias and high variability?





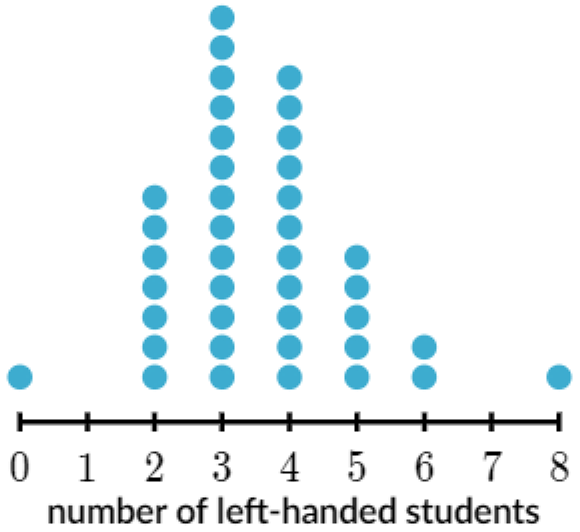


**~~~U09L01 Homework~~~**

1. Probability: Interpreting results of simulations
2. Sampling distributions: Biased and unbiased estimators

**~~~U09L01 Classwork~~~**

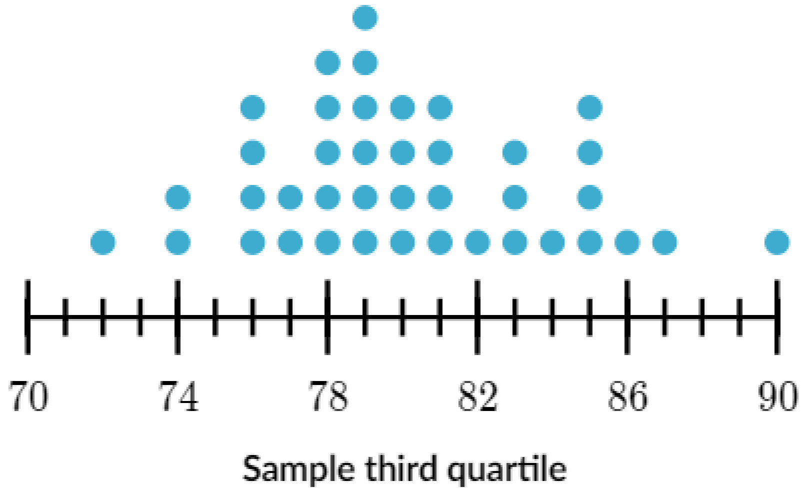
1. [Source: Khan Academy] Kaylee read a report that said the probability that a randomly selected American is left-handed is 14%. She was curious how many left-handed students to expect in a class of 25 students. She sampled 40 classes of 25 students. Kaylee counted how many left-handed students were in each class. Here are her results:



Use her results to estimate the probability that there are fewer than 3 left-handed students in a class of 25 students.

P(fewer than 3 left-handed)≈

2. [Source: Khan Academy] Serge was curious if a sample third quartile (or Q_3) was an unbiased estimator of a population third quartile. He started with a large normally distributed population of test scores whose third quartile was Q_3 = 80 points. He then took a random sample of 6 tests and calculated the third quartile of the sample. He replaced those tests and repeated this process for a total of 40 trials. His results are summarized in the dotplot below, where each dot represents the third quartile from a sample of 6 tests.

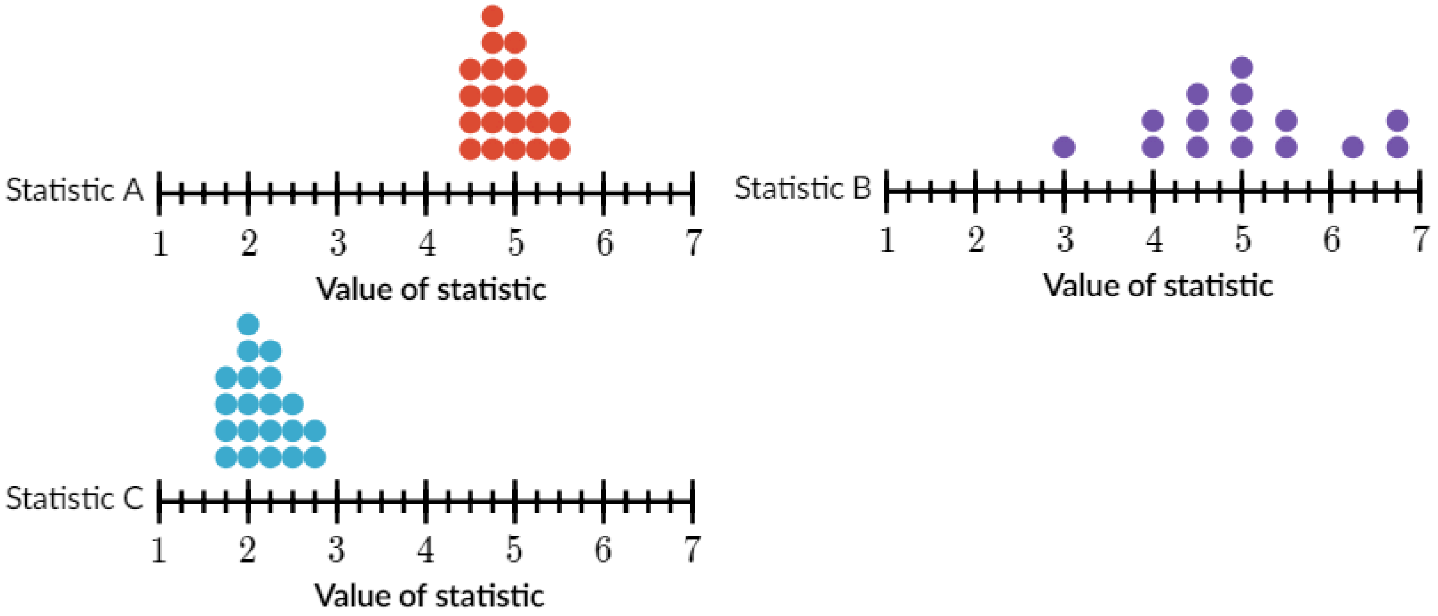


Based on these results, does sample third quartile appear to be a biased or unbiased estimator of population third quartile? Why?

3. **AP**: Casinos often put a board near their roulette wheels that list the winning colors from the last several games. The odds of red winning in American roulette is about 47%. The board indicates that red won 48 of the last 100 games. The most likely explanation for the difference between the observed results and the expected results in this case is

   (A) bias
   (B) variability due to sampling
   (C) nonsampling error
   (D) a sampling frame that is incomplete
   (E) confounding
   (F) divine intervention
   (G) for one spin that would normally have been black, the customer rubbed his lucky rabbit's foot and bet on red.
   (H) government conspiracy

4. [Source: Khan Academy] The dotplots below show an approximation to the sampling distribution for three different estimators of the same population parameter.



   If the actual value of the population parameter is 5, which dotplot displays the estimator with both low bias and low variability?

Unit 09 Lesson 02: Calculate the Mean and Standard Deviations of Sampling Distributions

How do we calculate these two critical estimators for sampling distributions?

- Given either population proportion or population mean/standard deviation, calculate the mean and standard deviation for a corresponding sampling distribution

Symbol roundup

	Proportion	Mean	Standard Deviation
Population Distribution			
Sample Distribution			
Sampling Distribution			

1. [Source: Khan Academy] Suppose that 93% of approximately 1000 total dogs sold by a shelter are up-to-date on their shots. Administration plans to take an SRS of 40 dogs sold by the shelter to see what proportion of the dogs sampled are up-to-date on their shots. What are the mean and standard deviation of the sampling distribution of the proportion of dogs who are up-to-date on their shots?

   $\mu_{\hat{p}} =$

   $\sigma_{\hat{p}} =$

2. [Source: Khan Academy] An organization is considering having a "Bring your pet to work" day. The administration takes an SRS of 35 of their employees and finds that 20% of those sampled are allergic to pets. Administration is considering taking more samples like this. Suppose that of all 500 employees of the organization, it's actually 10% that are allergic. Let $\hat{p}$ represent the proportion of a sample of 35 employees that are allergic to pets. What are the mean and standard deviation of the sampling distribution of $\hat{p}$?

3. [Source: Khan Academy] A car manufacturer produced 5,000 cars for a limited edition model. Dealers sold all of these cars at mean price of $36,000 with a standard deviation of $3,000. Suppose we were to take random samples of 9 of these cars and calculate the sample mean price $\bar{x}$ for each sample. Calculate the mean and standard deviation of the sampling distribution of $\bar{x}$.

$\mu_{\bar{x}} =$

$\sigma_{\bar{x}} =$

4. [Source: Khan Academy] Karina plays a video game that is scored by round. In about 500 rounds, her career average is 20 points per round with a standard deviation of 5 points per round. Suppose we take random samples of 3 past rounds and calculate the mean number of points $\bar{x}$ that she scored in each sample. Calculate the mean and standard deviation of the sampling distribution of $\bar{x}$.

$\mu_{\bar{x}} =$

$\sigma_{\bar{x}} =$

5. **AP**: Which of the following pairs of sample size *n* and population *p* would produce the greatest standard deviation for the sampling distribution of a sample proportion $\hat{p}$?
   (A) n = 500 and *p* close to 0
   (B) n = 500 and *p* close to 1
   (C) n = 500 and *p* close to $\frac{1}{2}$
   (D) n = 250 and *p* close to 0
   (E) n = 250 and *p* close to $\frac{1}{2}$

6. **AP**: A simulation was conducted using 8 fair six-sided dice, where the faces were numbered 1 through 6, respectively. All 8 dice were rolled, and the average of the 8 numbers appearing faceup was recorded. The process was repeated 16 times. Which of the following best describes the distribution being simulated?

    (A) A sampling distribution of a sample mean with n = 8, $\mu_{\bar{x}}$ = 3.5, and $\sigma_{\bar{x}} \approx$ 0.60

    (B) A sampling distribution of a sample mean with n = 8, $\mu_{\bar{x}}$ = 3.5, and $\sigma_{\bar{x}} \approx$ 1.71

    (C) A sampling distribution of a sample mean with n = 16, $\mu_{\bar{x}}$ = 3.5, and $\sigma_{\bar{x}} \approx$ 0.43

    (D) A sampling distribution of a sample proportion with n = 8, $\mu_{\hat{p}}$ = 1/6, and $\sigma_{\hat{p}} \approx$ 0.132

    (E) A sampling distribution of a sample proportion with n = 16, $\mu_{\hat{p}}$ = 1/6, and $\sigma_{\hat{p}} \approx$ 0.093

5. **AP**: Researchers working at the Environmental Protection Agency are investigating the average number of miles people drive each year. The researchers will use the mean miles driven in the last year of a random sample of 600 people. Which of the following best describes the effect on the bias and the variance of the estimator if the researchers increase the sample size to 900?

    (A) The bias will decrease and the variance will remain the same.
    (B) The bias will increase and the variance will remain the same.
    (C) The bias will remain the same and the variance will decrease.
    (D) The bias will remain the same and the variance will increase.
    (E) The bias will decrease and the variance will decrease.

5. **AP**: A small business is keeping track of how much money they make each week. The amount they make is proportional to the number of customers they get, which changes from week to week. They select a simple random sample of weeks and record how much money they make during each of those weeks. The business owners will report the sample mean as a point estimate for the weekly mean. Which of the following statements is correct for the sample mean as a point estimator?

    (A) A sample of size 20 will produce more variability of the estimator than a sample of size 45.
    (B) A sample of size 20 will produce less variability of the estimator than a sample of size 45.
    (C) A sample of size 20 will produce a biased estimator, but a sample of size 45 will produce an unbiased estimator.
    (D) A sample of size 20 will produce a more biased estimator than a sample of size 45.
    (E) A sample of size 20 will produce a less biased estimator than a sample of size 45.

**~~~U09L02 Homework~~~**

1. Sampling distributions: Mean and standard deviation of sample proportions
2. Sampling distributions: Mean and standard deviation of sample means

~~~U09L02 Classwork~~~

1. [Source: Khan Academy] The birth weights of newborn babies in the United States follow a normal distribution with a mean of 3.4 kg and a standard deviation of 0.6 kg. Researchers interested in studying how children gain weight decide to take random samples of 100 newborn babies and calculate the sample mean birth weight $\bar{x}$ for each sample. Calculate the mean and standard deviation of the sampling distribution of $\bar{x}$.

2. [Source: Khan Academy] Suppose that 42% of students of a high school play video games at least once a month. The computer programming club takes an SRS of 30 students from the population of 792 students at the school and finds that 40% of students sampled play video games at least once a month. The club plans to take more samples like this. Let $\hat{p}$ represent the proportion of a sample of 30 students who play video games at least once a month. What are the mean and standard deviation of the sampling distribution of $\hat{p}$?

3. **AP**: Toxoplasmosis is a single celled parasite that lives in the brain tissue of its hosts, which includes humans. The Center for Disease Control estimates the proportion of Americans with the disease is around 0.23. The true population proportion of Americans with the disease is 0.40. For samples of size 1000 that are selected at random from this population, what are the mean and standard deviation, respectively, for the sampling distribution of the sample proportion of Americans who have the disease?

   (A) 0.23, $\sqrt{1000(0.23)(0.77)}$

   (B) 0.23, $\sqrt{\dfrac{(0.40)(0.60)}{1000}}$

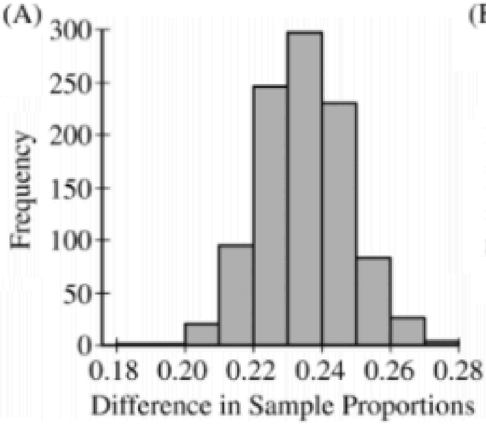
   (C) 0.23, $\sqrt{\dfrac{(0.23)(0.77)}{1000}}$

   (D) 0.40, $\sqrt{\dfrac{(0.40)(0.60)}{1000}}$

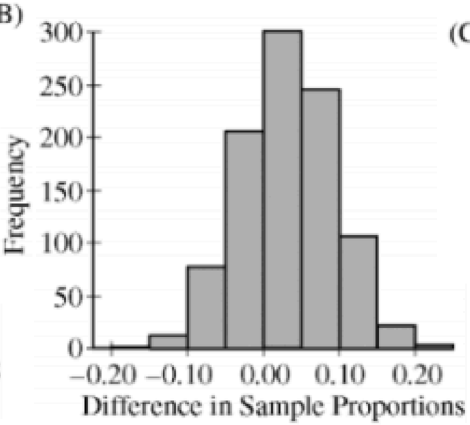
   (E) 0.40, $\sqrt{1000(0.40)(0.60)}$

4. **AP**: Suppose that 30 percent of adults and 27 percent of adolescents would answer yes to a particular question. In a simulation, a random sample of 100 adults and 100 adolescents were selected, and the difference in sample proportions of those who answered yes, $\hat{p}_{\text{adults}} - \hat{p}_{\text{adolescents}}$, was calculated. The process was repeated 1000 times. Which of the following is most likely to be a representation of the simulated sampling distribution of the difference between the two sample proportions?

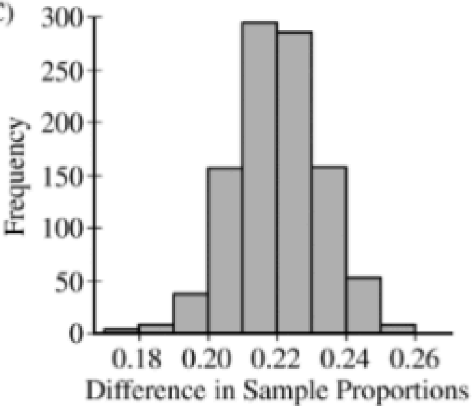
(A)



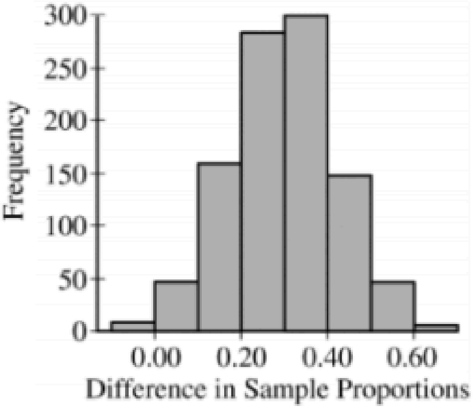
(B)



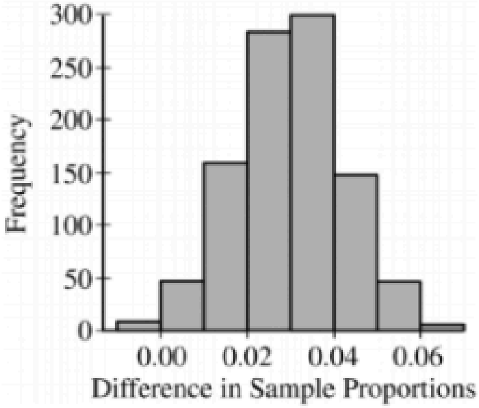
(C)



(D)



(E)



Unit 09 Lesson 03: Determine if a Sampling Distribution for a Proportion Meets the Normal Condition

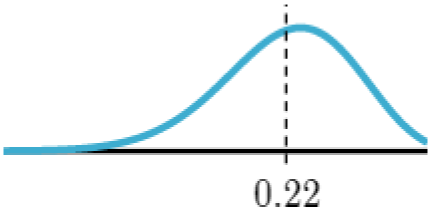
What has to happen so that I can estimate the probability of a sample occurring given a sample proportion?

- Determine if a sampling distribution for a proportion will be normal… enough… aka approximately normal

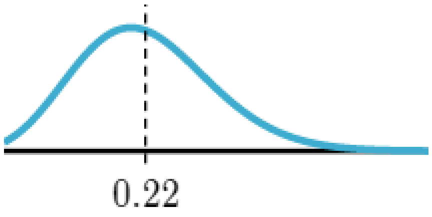
1. [Source: Khan Academy] According to the admissions director of a certain college, approximately 11% of the 250 freshmen admitted had applied using the early decision option. Suppose that we took random samples of 20 freshmen from this population and computed the proportion $\hat{p}$ of freshmen in each sample who had applied using the early decision option. We can assume the admissions director's claim is true. What will be the shape of the sampling distribution of $\hat{p}$?

2. [Source: Khan Academy] A college has a limited enrollment summer course that only accepts 50 students, selected at random from the more than 2000 qualified students who apply each year. Of these applicants, 22% have completed 2 or fewer years of college. Suppose the registrar calculates the annual proportion $\hat{p}$ of students accepted into the course who have completed 2 or fewer years of college. Which of the following distributions is the best approximation of the sampling distribution of $\hat{p}$?

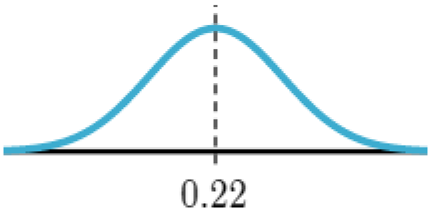
(A)



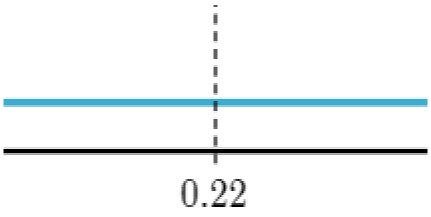
(B)



(C)



(D)



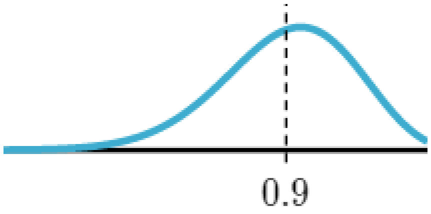
3. [Source: Khan Academy] A quality control inspector routinely takes random samples of 200 cans of fruit cocktail produced in a packing plant and calculates the proportion $\hat{p}$ of cans from each sample with at least 3 cherries. Suppose that 98% of fruit cocktail cans produced in that packing plant contain at least 3 cherries. What will be the shape of the sampling distribution of $\hat{p}$?

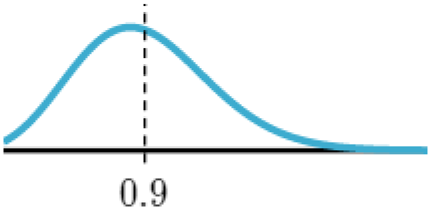
**~~~U09L03 Homework~~~** 1. Sampling distributions: The normal condition for sample proportions

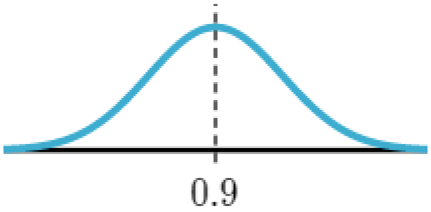
~~~U09L03 Classwork~~~

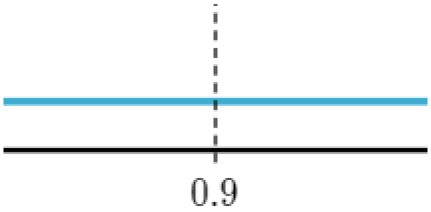
1. [Source: Khan Academy] According to a Nielsen survey, 92% of consumers around the world say that they trust recommendations from friends and family most, above all other forms of advertising. Suppose we took random samples of n = 100 consumers from this population and computed the proportion of consumers in each sample who trust recommendations from friends and family most. What will be the shape of the sampling distribution of the proportions of consumers who trust recommendations from friends and family most?

2. [Source: Khan Academy] A certain city's Department of Transportation (DOT) repairs about 80,000 potholes per year. The DOT completes 90% of the repairs on the reported potholes within two weeks of the report. Suppose the DOT annually takes a random sample of 75 of the reported potholes and calculates the proportion $\hat{p}$ of potholes repaired within two weeks. Which of the following distributions is the best approximation of the sampling distribution of $\hat{p}$?

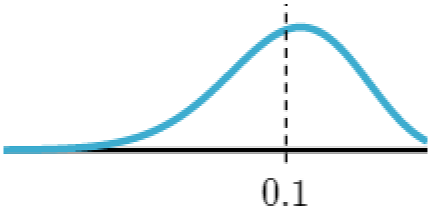
A


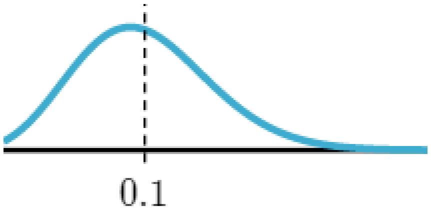
B


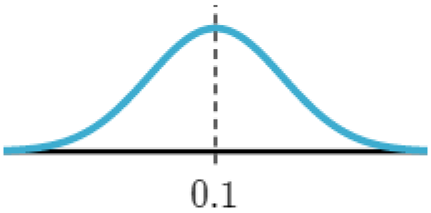
C


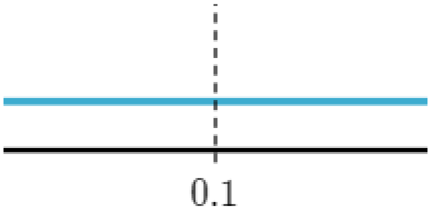
D


3. [Source: Khan Academy] According to a Gallup poll in 2006, about 10% of Americans said they were "very afraid" of flying on an airplane. Suppose that we took random samples of n = 40 people from this population and computed the proportion of people in each sample who were very afraid of flying. Which of the following distributions is the best approximation of the sampling distribution of the proportion of people who were very afraid of flying?

A


B


C


D


Unit 09 Lesson 04: Find Probabilities with Sample Proportions

How likely is it for a certain sample mean to occur? This seems like it would be the most important idea we've learned.

- Calculate probabilities for a given true proportion and sample proportion to occur

1. [Source: Khan Academy] The United Kingdom Forestry Commission reports that 43% of the 3.16 million hectares of woodland area in the United Kingdom had certification identifying them as "sustainably managed" in 2016. Suppose an employee took a simple random sample of 400 of the hectares and saw that the records showed that 47% of the sampled hectares had that certification in 2016. Assuming that the Forestry Commission's report is accurate, what is the approximate probability that more than 47% of the sample would have had the certification in 2016?

2. An article written for a magazine claims that 78% of the magazine's subscribers report eating healthily the previous day. Suppose we select a simple random sample of 675 of the magazine's approximately 50,000 subscribers to check the accuracy of this claim. Assuming the article's 78% claim is correct, what is the approximate probability that less than 80% of the sample would report eating healthily the previous day?

**~~~U09L04 Homework~~~** 1. Sampling distributions: Finding probabilities with sample proportions

~~~U09L04 Classwork~~~

1. [Source: Khan Academy] Suppose that 15% of the 1750 students at a school have experienced extreme levels of stress during the past month. A high school newspaper doesn't know this figure, but they are curious what it is, so they decide to ask a simple random sample of 160 students if they have experienced extreme levels of stress during the past month. Subsequently, they find that 10% of the sample replied "yes" to the question. Assuming the true proportion is 15%, what is the approximate probability that more than 10% of the sample would report that they experienced extreme levels of stress during the past month?

2. [Source: Khan Academy] The school board claims that 85% of the over 1,600 families with children in their district send their children to public school. Skeptical of this claim, a private school administrator takes a simple random sample of 120 families with children in the district and asks them whether they send their children to public school. Assuming the school board's claim is correct, what is the probability that the administrator's results are within 5 percentage points of the school board's 85% claim?

3. **AP**: According to government data, a 20% of people between ages 24-30 have attained a certain education level. A simple random sample of 400 people in the United States between ages 24-30 were selected for a study on income levels. If the government data are correct, which of the following best approximates the probability that at least 30% of the people in the sample have attained that education level?
(Note: z represents a standard normal random variable.)

(A) $$P\left(z > \frac{0.30 - 0.20}{\sqrt{\frac{(0.50)(0.50)}{400}}}\right)$$

(B) $$P\left(z > \frac{0.30 - 0.20}{\sqrt{\frac{(0.20)(0.80)}{400}}}\right)$$

(C) $$P\left(z > \frac{0.30 - 0.20}{\sqrt{\frac{(0.30)(0.70)}{400}}}\right)$$

(D) $$P\left(z > \frac{0.20 - 0.30}{\sqrt{\frac{(0.20)(0.80)}{400}}}\right)$$

(E) $$P\left(z > \frac{0.20 - 0.30}{\sqrt{\frac{(0.30)(0.70)}{400}}}\right)$$

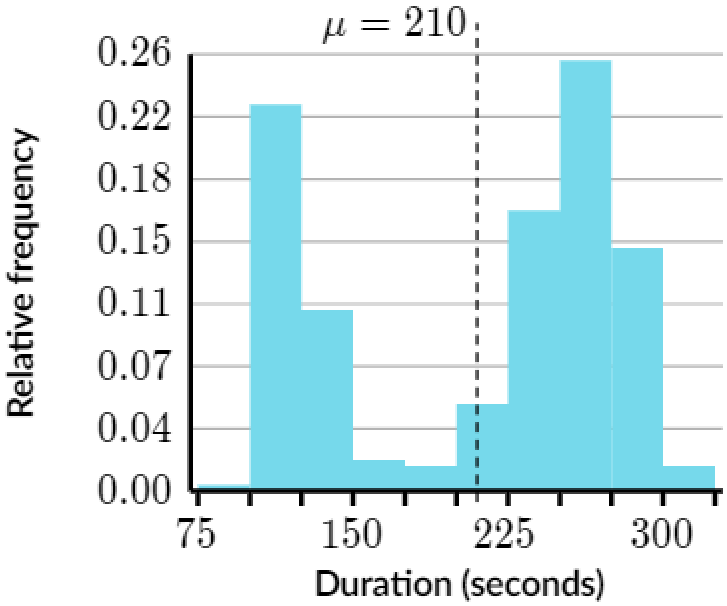
4. [Source: Khan Academy] A local agricultural cooperative claims that 55% of about 60,000 adults in a county believe that gardening should be part of the school curriculum. However, when you take a simple random sample of 300 of the adults in the county, only 50% say that they believe that gardening should be part of the school curriculum. Assuming that the agricultural cooperative's claim is accurate, what is the approximate probability that less than 50% of the sample would say that they believe that gardening should be part of the school curriculum?

Unit 09 Lesson 05: Determine if a Sampling Distribution for a Mean Meets the Normal Condition

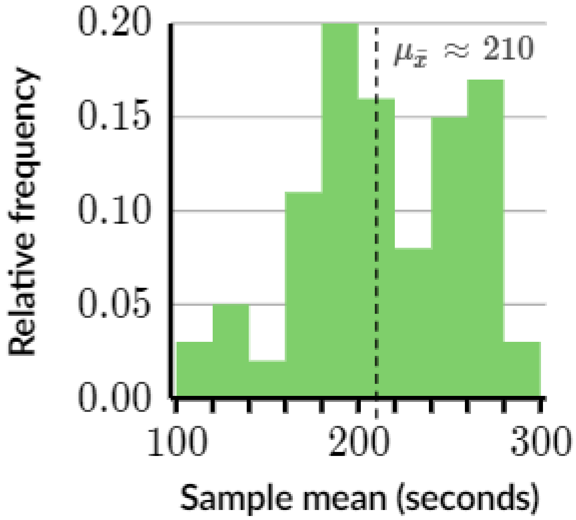
What has to happen so that I can estimate the probability of a sample occurring given a sample mean?

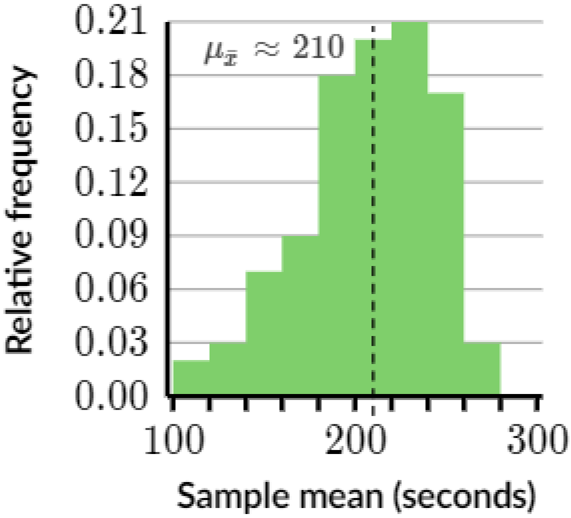
- Determine if a sampling distribution for a mean will be normal… enough… aka approximately normal

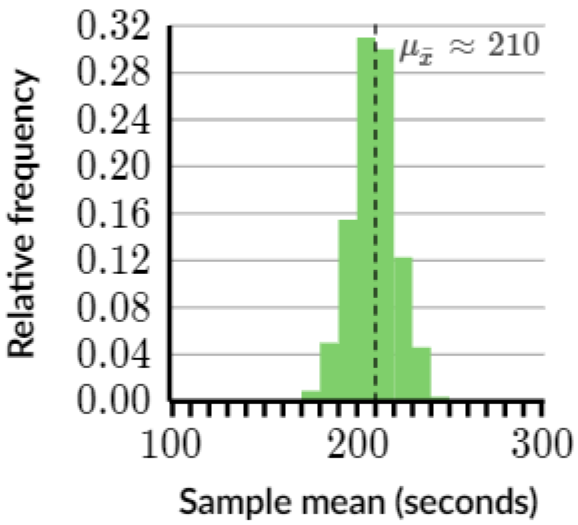
1. [Source: Khan Academy] Geologists recorded the duration (in seconds) of over 500 recent eruptions of the Old Faithful geyser in Yellowstone National Park. Here's a histogram showing the distribution of their data:

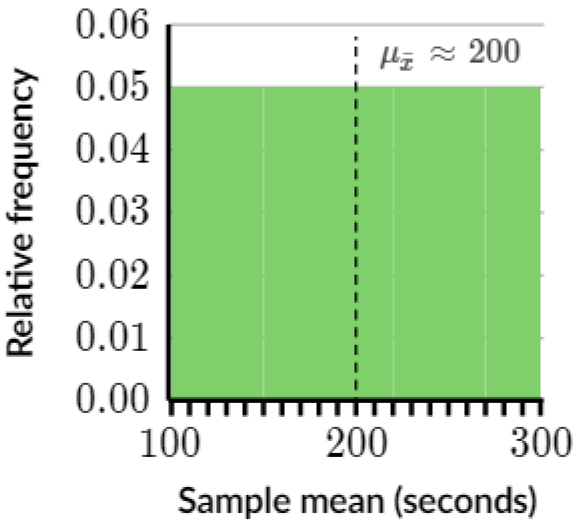


Suppose that we were to take random samples of 30 eruptions from this population and calculate $\bar{x}$ as the sample mean duration of the eruptions in each sample. Which graph shows the most reasonable approximation of the sampling distribution of $\bar{x}$?









2. [Source: Khan Academy] Patients recovering from a stroke have their grip strengths measured in each hand to monitor their progress. A certain population of over 100 male patients have grip strengths in their dominant hands with a mean and standard deviation of approximately 41kg and 9kg, respectively. Suppose that we take random samples of 4 male patients from this population and calculate $\bar{x}$ as the sample mean grip strength from each group of patients. What will be the shape of the sampling distribution of $\bar{x}$?
   (A) Skewed to the left
   (B) Skewed to the right
   (C) Approximately normal
   (D) Unknown; we don't have enough information to determine the shape.

3. [Source: Khan Academy] A group of teachers decided that they would each bring a similar jar of candies to their classrooms so their students could practice estimation. The students were told that whoever had the closest guess in the class would win the candy, and so they all guessed how many candies were in their classroom's jar. The distribution of individual guesses was strongly skewed to the right with a mean of 115 candies and a standard deviation of 35 candies. Suppose we took random samples of 6 students and calculated $\bar{x}$ as the sample mean of their guesses. We can assume that the students in each sample are independent. What will be the shape of the sampling distribution of $\bar{x}$?
   (A) Skewed to the left
   (B) Skewed to the right
   (C) Approximately normal
   (D) Unknown; we don't have enough information to determine the shape

**~~~U09L05 Homework~~~** 1. Sampling distributions: Sampling means and the central limit theorem

~~~U09L05 Classwork~~~

1. **AP**: There were 13,041 cars sold in a western city in the year 2010. The distribution of the sales prices of these cars was strongly right-skewed, with a mean of \$21,004 and a standard deviation of \$3,218. If all possible simple random samples of size 100 are drawn from this population and the mean is computed for each of these samples, which of the following describes the sampling distribution of the sample mean?
   (A) Approximately normal with mean \$21,004 and a standard deviation of \$322.
   (B) Approximately normal with mean \$21,004 and a standard deviation of \$3,218.
   (C) Approximately normal with mean \$21,004 and a standard deviation of \$161.
   (D) Strongly right-skewed with mean \$21,004 and a standard deviation of \$3,218.
   (E) Strongly right-skewed with mean \$21,004 and a standard deviation of \$32,180.

2. **AP**: Market researchers examined customer records and found the distribution of bottles of soda purchased per year by customers is skewed to the left with a mean 11 bottles and a standard deviation of 7 bottles. A random sample of 49 customer records was selected, and the sample mean number of bottles was recorded. Suppose the process of selecting a random sample of 49 customer records and recording the sample mean number of gallons was repeated for a total of 100 samples. Which of the following is the best description of a dotplot created from the 100 sample means?
   (A) The dotplot is skewed to the left with a mean of 11 bottles and a standard deviation of 7 bottles.
   (B) The dotplot is skewed to the left with a mean of 11 bottles and a standard deviation of 1 bottle.
   (C) The dotplot is skewed to the left with a mean of 11 bottles and a standard deviation of 0.7 bottles.
   (D) The dotplot is approximately normal with a mean of 11 bottles and a standard deviation of 1 bottle.
   (E) The dotplot is approximately normal with a mean of 11 bottles and a standard deviation of 0.7 bottles.

3. **AP**: Let X be a random variable that has a skewed distribution with mean $\mu = 12$ and standard deviation $\sigma = 12$. Based on random samples of size 225, the sampling distribution of $\bar{x}$ is
   (A) highly skewed with mean 12 and standard deviation 12.
   (B) highly skewed with mean 12 and standard deviation 6.
   (C) highly skewed with mean 12 and standard deviation 0.8.
   (D) approximately normal with mean 12 and standard deviation 12.
   (E) approximately normal with mean 12 and standard deviation 0.8.

Unit 09 Lesson 06: Find Probabilities with Sample Means

How likely is it for a certain sample mean to occur? This seems like it would be the most important idea we've learned. Again, but for means instead of proportions.

- Calculate probabilities for a given true mean and sample mean to occur, this time for means instead of proportions

1. [Source: Khan Academy] A company sells eggs whose individual weights are normally distributed with a mean of 70g and a standard deviation of 2g. Suppose that these eggs are sold in packages that each contain 4 eggs that represent an SRS from the population. What is the probability that the mean weight of 4 eggs in a package $\bar{x}$ is less than 68.5g?

2. [Source: Khan Academy] Mathieu grows specialty tomatoes that are much larger than typical tomatoes. The distribution of their weights is strongly skewed to the left with a mean of 232g and a standard deviation of 12g. Suppose we were to calculate the mean weight from a random sample of 16 of Mathieu's tomatoes. We can assume independence between tomatoes in the sample. What is the probability that the mean weight from the sample $\bar{x}$ of 16 tomatoes is within 6g of the population mean?

3. **AP**: A market research agency wants to do a study on movie viewing habits. 1000 randomly selected adults will be asked the question: How many films did you see last year? The sample mean will be computed. Let μ denote the population mean response to the question if everyone in the population is to be asked the question. Is the sample mean $\bar{x}$ unbiased for estimating μ?
   (A) Yes, because for random samples the mean (expected value) of the sample mean $\bar{x}$ is equal to the population mean μ.
   (B) Yes, because with a sample size of 1,000 the standard deviation of the sample mean $\bar{x}$ is small.
   (C) Yes, because the wording of the question is not biased.
   (D) No, because the sample mean $\bar{x}$ does not always equal the population mean μ.
   (E) No, because the number of films seen is not normally distributed so the mean (expected value) of the sample mean $\bar{x}$ does not equal the population mean μ.

**~~~U09L06 Homework~~~** 1. Sampling distributions: Finding probabilities with sample means

~~~U09L06 Classwork~~~

1. **AP**: A recent study was conducted to investigate the average typing speed on a computer keyboard. The reported mean was 40 words per minute with a standard deviation of 14 words per minute. Suppose the reported values are the true mean and standard deviation for the population of subjects in the study. If a random sample of 121 subjects is selected from the population, what is the approximate probability that the mean of the sample will be more than 42 words per minute?
   (A) 0.4443
   (B) 0.5557
   (C) 0.0582
   (D) 0.9418
   (E) 0.4164

2. [Source: Khan Academy] A pizza chain monitors the total weight of pepperoni that goes on its deluxe pepperoni pizzas to make sure customers are satisfied and product isn't being wasted. Suppose that for pizzas in this population, these weights are strongly skewed to the left with a mean of 250g and a standard deviation of 8g. Management takes a random sample of 64 of these pizzas and calculates the mean weight of the pepperoni on the pizzas. Assume that the pizzas in the sample are independent. What is the probability that the mean weight of the pepperoni from the sample of 64 pizzas $\bar{x}$ is within 1g of the true mean?

3. **AP**: Farmers have variable crop yields based on weather patterns. The distribution of crop yields across all farms in a region last year has a mean of 1.8 and a standard deviation of 0.8. Let z represent the standard normal distribution. If $\bar{x}$ represents the mean crop yield last year for a random sample of 81 farms, which of the following calculations will give the approximate probability that $\bar{x}$ is less than 2?

(A) $P\left(z < \frac{2 - 1.8}{\left(\frac{0.8}{\sqrt{81}}\right)}\right)$

(B) $P\left(z < \frac{2 - 1.8}{0.8}\right)$

(C) $P(z < 2)$

(D) $P\left(z < \frac{1.8 - 2}{\left(\frac{0.8}{\sqrt{81}}\right)}\right)$

(E) $P\left(z < \frac{1.8 - 2}{0.8}\right)$

4. **AP**: In a report to the government, a large company that makes cereal bars stated that the distribution of calories in each bar is approximately normal with a mean of 200 calories and a standard deviation of 9 calories. The state health inspector decided to check a sample of 100 of the company's bars to test the company's claim concerning the mean calories and standard deviation of the calories in its cereal bars.

   (a) Assume that the company's claim is true. Describe the distribution of the sample mean weight for random samples, each consisting of 100 bars.

   (b) The health inspector selects a random sample of 100 cereal bars and finds that the mean calories of those bars is 202 calories. What is the probability that a random sample of 100 of the company's cereal bars would have a mean of 202 calories or more if the company's claim is true?

   (c) A second health inspector is assigned to check the company's cereal bars, but on a different day. However the second health inspector believes that the instructions to carry out a random sample are too complicated and too time consuming. Instead he inspects the first 100 bars made by the factory that day and finds that the mean number of calories 197. Why is the lack of random selection in using the first 100 cereal bars a potential problem?

4. **AP**: Two financial analysts for the Wall Street Journal, Alice and Bob, want to know how many Bitcoins their average subscriber owns. A survey of all the names and genders for the 463,521 subscribers they had in the last year is available. Alice selected a simple random sample of 100 customers. Knowing from their list that 70% of their customers are male, Bob selected a simple random sample of 70 males and an independent simple random sample of 30 females. Both asked all of the subscribers in their sample how many Bitcoins they own, if any.

   (a) Describe a method Alice could have used to select a simple random sample of 100 customers.

Alice and Bob conducted their studies as described. Alice used the sample mean $\bar{X}$ as a point estimator for μ. Bob used $\bar{X}_{\text{overall}} = (0.3)\bar{X}_{\text{female}} + (0.7)\bar{X}_{\text{male}}$ as a point estimator for μ, where $\bar{X}_{\text{female}}$ is the mean of the sample of 30 females and $\bar{X}_{\text{male}}$ is the mean of the sample of 70 males.

Summary statistics for Alice's data are shown in the table below.

Variable	N	Mean	Standard Deviation
Number of Bitcoins	100	4.71	2.78

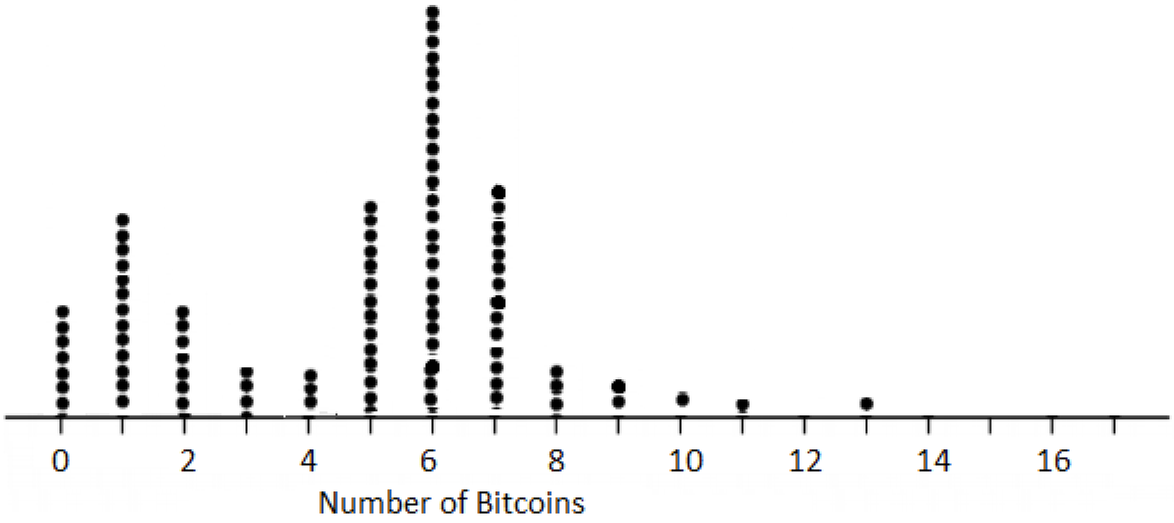
   (b) Based on the summary statistics, calculate the estimated standard deviation of the sampling distribution (sometimes called the standard error) of Alice's point estimator $\bar{X}$.

Summary statistics for Bob's data are shown in the table below.

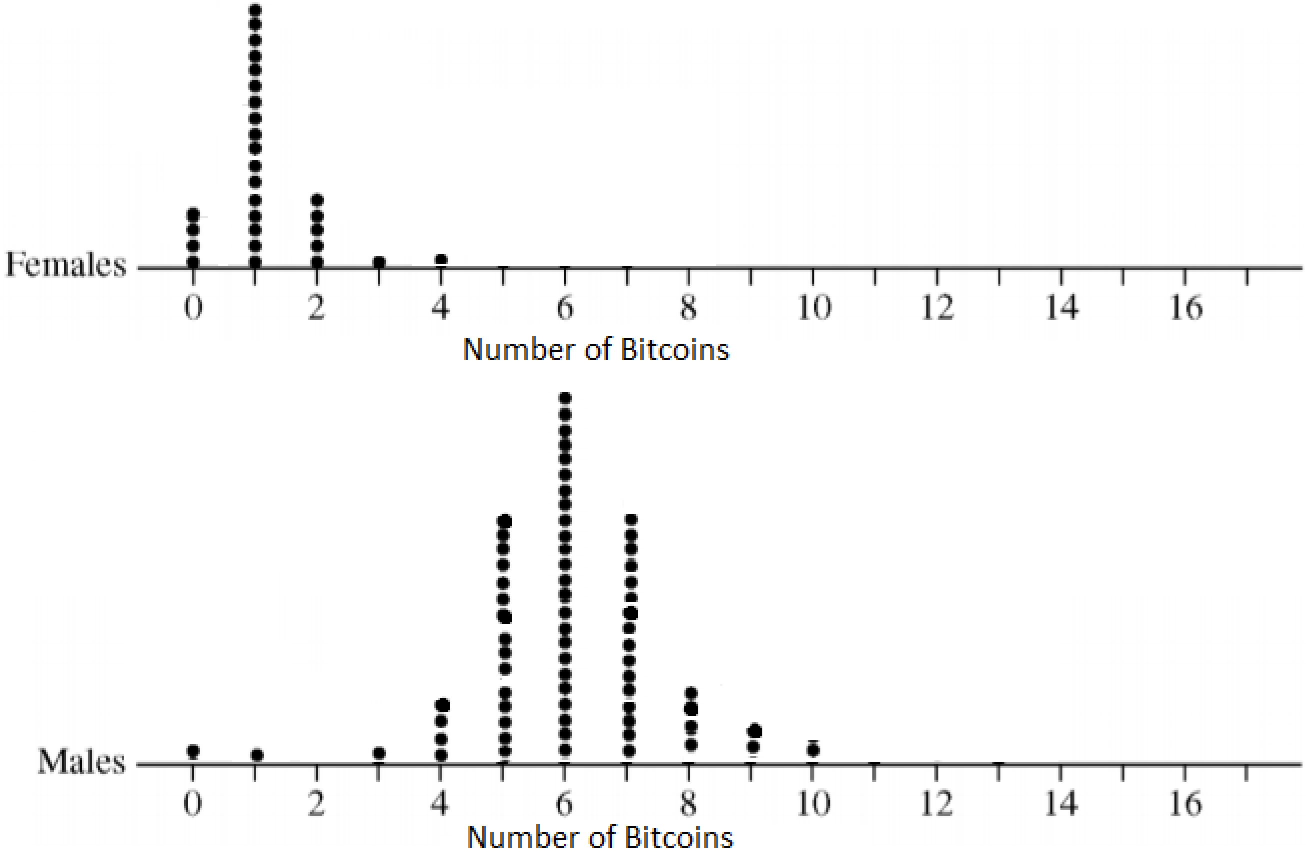
Variable	Gender	N	Mean	Standard Deviation
Number of Bitcoins	Female	30	1.21	0.88
	Male	70	5.96	1.56

   (c) Based on the summary statistics, calculate the estimated standard deviation of the sampling distribution of Bob's point estimator $\bar{X}_{\text{overall}} = (0.3)\bar{X}_{\text{female}} + (0.7)\bar{X}_{\text{male}}$.

A dotplot of Alice's sample data is given below



Comparative dotplots of Bob's sample data are given below



(d) Using the dotplots above, explain why Bob's point estimator has a smaller estimated standard deviation than the estimated standard deviation of Alice's point estimator.

Unit 09 Lesson 07: Review for Test on Sampling Distributions

Can I handle any situation that involves bias in sampling distributions, and probabilities in sampling distributions?

- Ensure you've mastered the concepts and skills of sampling distributions
- Ensure you've retained mastery of previous units

Khan Academy Sampling Distributions Unit Test