Sampling Distribution of the Sample Mean

- Law of Large Numbers
- Central Limit Theorem
- Mean, Standard Error, and Shape
- Sampling Distribution Probabilities
- Looking Ahead to Inference

University of Pittsburgh

Motivation: Sampling Distribution

- **Scenario:** SAT scores are normally distributed with population mean $\mu = 1000$ and population standard deviation $\sigma = 217$. Ask 1 random college freshman what their SAT score was.
- **Question:** Should we be surprised if their score was greater than 1100?
- **Answer:** _____
  - 1100 is ________________ standard deviation _______ the mean
  - Scores in this range are ________________
  - **Calculation:**

$$P(X > 1100) = P\left(Z > \frac{1100 - 1000}{217}\right) = P(Z > \_\_\_\_) = \_\_\_\_\_\_$$

Motivation: Sampling Distribution

- **Scenario:** SAT scores are normally distributed with population mean $\mu = 1000$ and population standard deviation $\sigma = 217$. Ask 9 random college freshmen what their SAT score was.
- **Question:** Should we be surprised if the sample mean SAT score was greater than 1100?
- **Answer:** ___________
  - Requires each of the 9 to do ______ well with ______ outliers
  - **Example Data:** 880, 910, 970, 1000, 1070, 1130, 1270, 1320, 1350
- **Question:** Why can't we find this exact probability yet?
- **Answer:** Sampled ________________ observation
  - Don't yet know how to handle working with the ________ of a sample

Motivation: Sampling Distribution

4

- **Scenario:** SAT scores are normally distributed with population mean $\mu = 1000$ and population standard deviation $\sigma = 217$. Ask 100 random college freshmen what their SAT score was.
- **Question:** Should we be surprised if the sample mean SAT score was greater than 1100?
- **Answer:** ____
  - Requires a ______________ of the sample to score ____________ with few people performing ______ average
    - Would be unusual because we would expect ___ to score above 1000 and ___ to score below in a large sample
- **Takeaway:** It becomes _______________ for ___ to deviate from ___ as the sample size ____________.

Law of Large Numbers

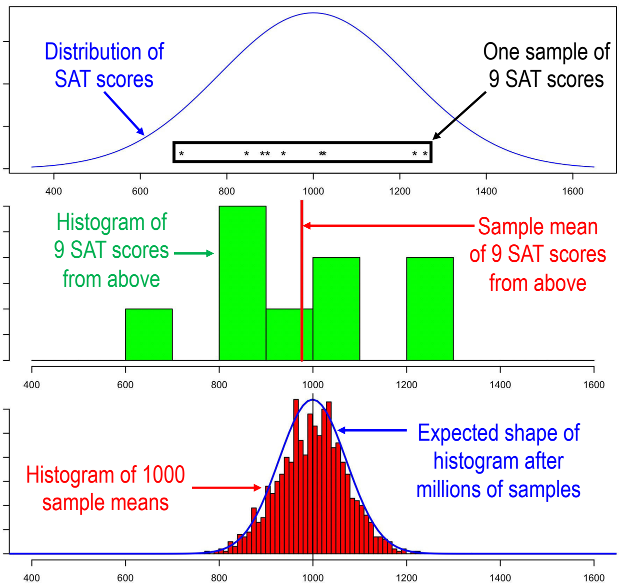
5

- **Question:** Why does it become harder for the sample mean to deviate from the population mean as the sample size gets larger?
- **Answer:** The _________________________
  - Sample n observations independently from any population with mean μ
  - Calculate the sample mean $\bar{x}$
  - As n increases, the sample mean $\bar{x}$ __________ to the population mean μ
- **Question:** Why does this occur?
- **Answer:**
  - In small samples, one outlier ____________________ the mean up or down
  - In large samples, an unusually small observation will often be __________ by an unusually large observation

Simulation: Creating a Sampling Distribution

6

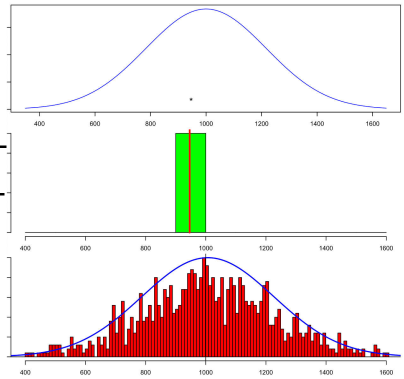
- **Scenario:** Take a random sample of SAT scores of a given sample size, calculate the sample mean, and plot this new sample mean in a histogram. Repeat this 1000 times.
  - ***Know:*** *Sample mean converges to the population mean as the sample size increases.*



- **Goal:** Understand the ______ and ________ of this new distribution.

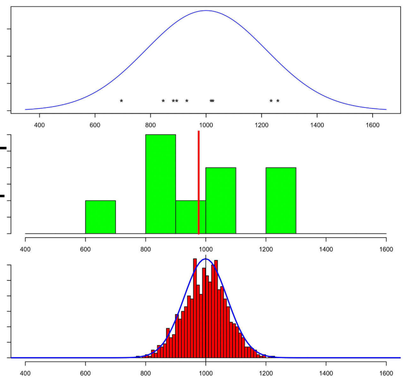
SIMULATION: CREATING A SAMPLING DISTRIBUTION

- **Scenario:** SAT scores follow a normal distribution with population mean $\mu = 1000$ and population standard deviation $\sigma = 217$. Take 1000 samples of size **1**, calculate and plot the sample mean.
- **Question:** What are the mean, standard deviation, and shape of the resulting histogram of sample means?
- **Answer:**
  - Average Sample Mean: __________
  - Standard Deviation of Sample Means: _________
  - Shape of Sample Means: _____________________



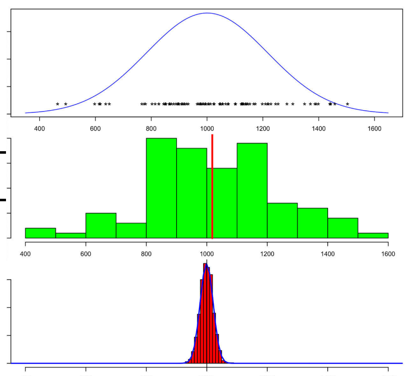
SIMULATION: CREATING A SAMPLING DISTRIBUTION

- **Scenario:** SAT scores follow a normal distribution with population mean $\mu = 1000$ and population standard deviation $\sigma = 217$. Take 1000 samples of size **9**, calculate and plot the sample mean.
- **Question:** What are the mean, standard deviation, and shape of the resulting histogram of sample means?
- **Answer:**
  - Average Sample Mean: __________
  - Standard Deviation of Sample Means: _________
  - Shape of Sample Means: _____________________



SIMULATION: CREATING A SAMPLING DISTRIBUTION

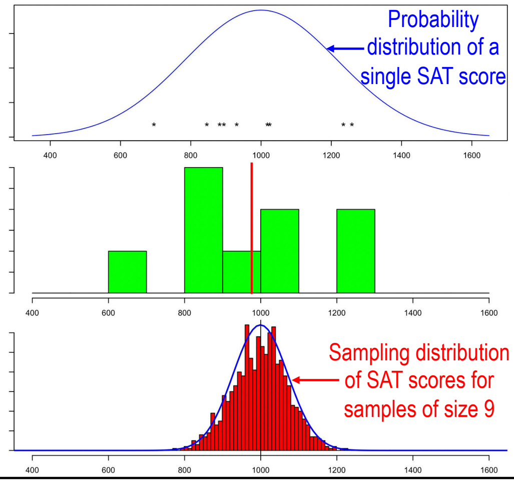
- **Scenario:** SAT scores follow a normal distribution with population mean $\mu = 1000$ and population standard deviation $\sigma = 217$. Take 1000 samples of size **100**, calculate and plot the sample mean.
- **Question:** What are the mean, standard deviation, and shape of the resulting histogram of sample means?
- **Answer:**
  - Average Sample Mean: __________
  - Standard Deviation of Sample Means: _________
  - Shape of Sample Means: _____________________



- **Question:** What are the red distributions?
- **Answer:** ___________________________

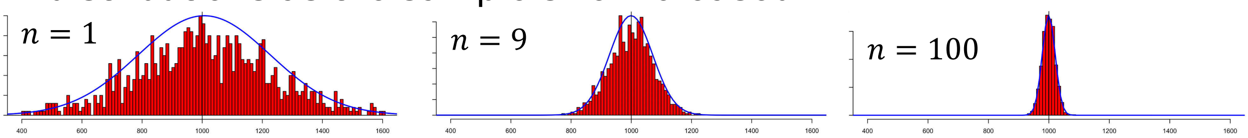
SAMPLING DISTRIBUTION

- **Sampling distribution:** For quantitative data, the distribution of all possible sample means for a given mean μ, standard deviation σ, and sample size n
  - Defined by three components:
    1. **Mean**
    2. **Standard error:** Standard deviation of the sampling distribution; measures how spread out the sample means tend to be
    3. **Shape**
  - In other words, a **sampling distribution** describes where we can expect a sample mean to fall in relation to the population mean.



MOTIVATION: SAMPLING DISTRIBUTION CHANGES

- **Question:** What do you notice about the changes in the sampling distributions as the sample size increased?

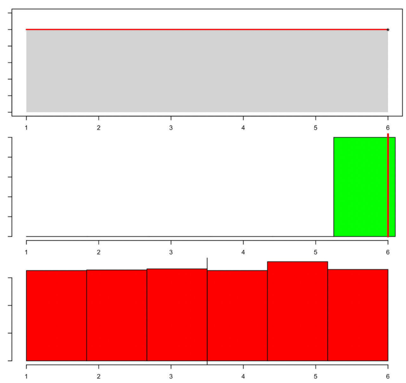


- **Answer:**
  - **Mean:** All were _______________ to the population mean (1000)
  - **Standard Error:** Became _________
  - **Shape:** ______________________________

- ***Takeaway:*** *When sampling from a normal population, the mean of the sampling distribution equals the _______________ mean and the shape will be _________, but the spread will _________ as n increases.*

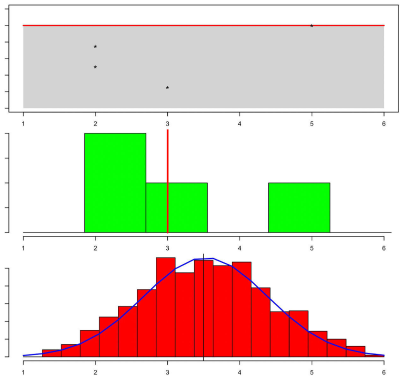
MOTIVATION: CENTRAL LIMIT THEOREM

- **Scenario:** A fair six-sided die has a population mean of $\mu = 3.5$ and a standard deviation of $\sigma = 1.708$. Roll **1** fair six-sided die, calculate and plot the sample mean. Repeat this 1000 times.
- **Question:** What are the mean, standard error, and shape of the resulting sampling distribution?
- **Answer:**
  - **Average Sample Mean:** ________
  - **Standard Error:** ________
  - **Shape:** __________



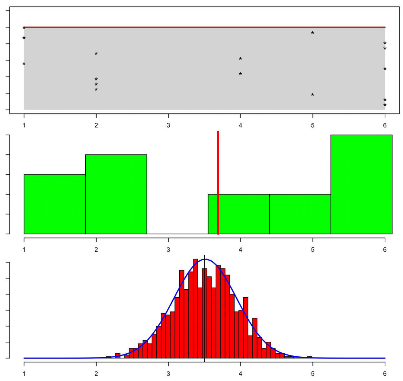
MOTIVATION: CENTRAL LIMIT THEOREM

- **Scenario:** A fair six-sided die has a population mean of $\mu = 3.5$ and a standard deviation of $\sigma = 1.708$. Roll **4** fair six-sided dice, calculate and plot the sample mean. Repeat this 1000 times.
- **Question:** What are the mean, standard error, and shape of the resulting sampling distribution?
- **Answer:**
  - **Average Sample Mean:** _______
  - **Standard Error:** _______
  - **Shape:** ___________________
    - Sampling distribution still is not "______" enough to be considered normal given how many samples were taken
    - Looks too much like a _________



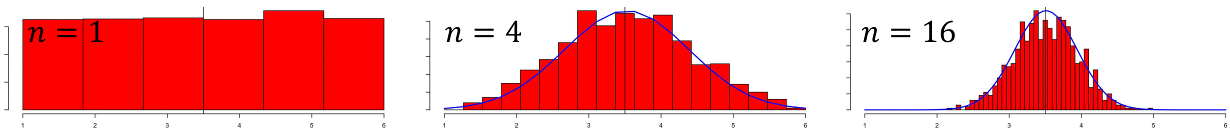
MOTIVATION: CENTRAL LIMIT THEOREM

- **Scenario:** A fair six-sided die has a population mean of $\mu = 3.5$ and a standard deviation of $\sigma = 1.708$. Roll **16** fair six-sided dice, calculate and plot the sample mean. Repeat this 1000 times.
- **Question:** What are the mean, standard error, and shape of the resulting sampling distribution?
- **Answer:**
  - **Average Sample Mean:** _______
  - **Standard Error:** _______
  - **Shape:** ___________________
    - Curve takes on more of a __________



MOTIVATION: SAMPLING DISTRIBUTION CHANGES

- **Question:** What do you notice about the changes in the sampling distributions as the sample size increased?



- **Answer:**
  - **Mean:** All were _____________ to the population mean (3.50)
  - **Standard Error:** Became ________
  - **Shape:** Became ________________

- ***Takeaway:*** *When sampling from a **slightly skewed** or **uniform** population, the mean of the sampling distribution always equals the ___________ mean, but the shape will become ________ and the spread will _________ as n increases.*

Central Limit Theorem

- **Central Limit Theorem:** The sampling distribution of the sample mean approaches a normal distribution as the sample size increases regardless of the shape of the original population.

- ***Rules of Thumb:*** *Take a random sample of size n from some distribution X of quantitative data. Then the distribution of the sample mean $\bar{X}$ is normal if any of the following are true:*
  1. *The original distribution of X is normal (unimodal and approximately symmetric), regardless of the sample size.*
  2. *The distribution of X is either uniform* **or** *unimodal and only slightly skewed and $n \geq 15$.*
  3. *The distribution of X is severely skewed and $n \geq 30$.*

Mean and Standard Error of Sample Mean

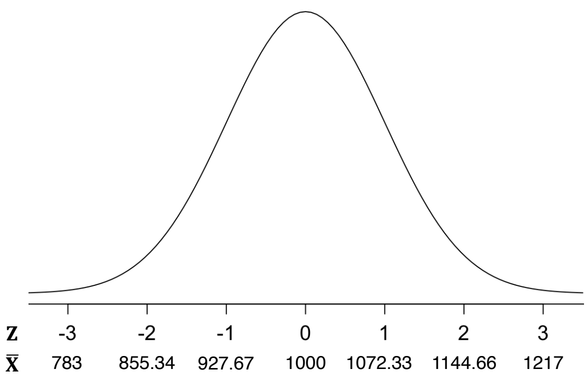
- Sample n observations from **any** population with quantitative data that has mean μ and standard deviation σ. Then:
  1. **Mean:** $\mu_{\bar{X}} = \mu$
     - Mean of the sampling distribution of $\bar{X}$ equals the population mean
  2. **Standard Error:** $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
     - Standard error equals the standard deviation of the original population divided by the square root of the sample size

- If the Central Limit Theorem holds, (i.e. the shape of the sampling distribution is normal), then the sample mean is standardized as:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Example: Determining Sampling Distribution

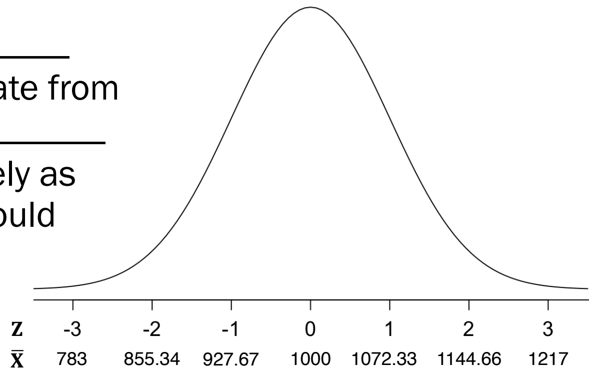
- **Scenario:** SAT scores are normally distributed with population mean $\mu = 1000$ and population standard deviation $\sigma = 217$. Ask 9 random college freshmen what their SAT score was.
- **Question:** What is the sampling distribution of the sample mean?
- **Answer:**
  1. **Mean:** ___________
  2. **Standard Error:** ________________
  3. **Shape:** _________
     - Rule of Thumb ___ applies: __________ distribution of SAT scores is normal so the sampling distribution is _______



Note: Simulation yielded a mean of 999.176 and standard error of 71.895.

Example: Understanding Sampling Distribution

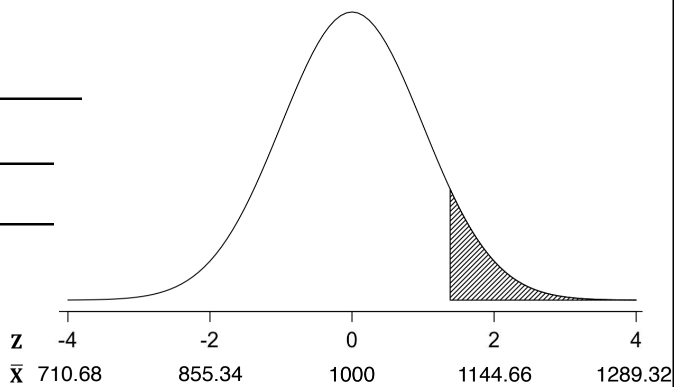
- **Scenario:** SAT scores are normally distributed with population mean $\mu = 1000$ and population standard deviation $\sigma = 217$. Ask 9 random college freshmen what their SAT score was.
- **Question:** What does the sampling distribution reveal?
- **Answer:**
  - Expect the average SAT score to be _____
  - Sample means of size 9 tend to deviate from 1000 by about _______________________
  - A sample mean of ____ is equally likely as a sample mean of _____, but both would be ______________
    - ______________ from the mean



Example: Probability Calculation

- **Scenario:** SAT scores are normally distributed with population mean $\mu = 1000$ and population standard deviation $\sigma = 217$.
- **Question:** What is the probability that the sample mean SAT score of 9 random college freshmen is greater than 1100?
- **Answer:**

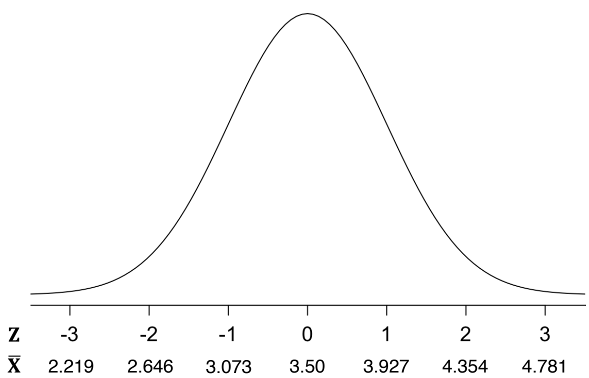
$$P(\bar{X} > 1100) = \_\_\_\_\_\_\_\_\_\_$$
$$= \_\_\_\_\_\_\_\_\_\_$$
$$= \_\_\_\_\_\_\_\_\_\_$$
$$= \_\_\_\_\_\_\_$$
$$= \_\_\_\_\_$$



Example: Determining Sampling Distribution

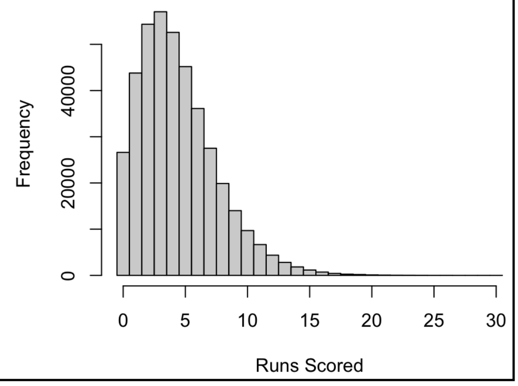
- **Scenario:** A fair six-sided die has a population mean of $\mu = 3.5$ and a standard deviation of $\sigma = 1.708$. Roll 16 fair six-sided dice and average the results.
- **Question:** What is the sampling distribution of the sample mean?
- **Answer:**
  1. Mean: __________
  2. Standard Error: _______________
  3. Shape: ________
     - Rule of Thumb ___ applies: Original distribution of die rolls is ________, but the sample size is ___________

Note: Simulation yielded a mean of 3.51 and standard error of 0.433.



Example: Determining Sampling Distribution

- **Scenario:** Probability distribution of runs scored for all games in MLB history shown. Population mean is $\mu = 4.41$ with population standard deviation $\sigma = 3.13$. Randomly sample 25 games and average the runs scored.
- **Question:** What is the sampling distribution of the sample mean?
- **Answer:**
  1. Mean: __________
  2. Standard Error: _______________
  3. Shape: _____________
     - _____ of the rules of thumb apply: Original distribution is ________ skewed and sample size is __________



Example: Impossible Probabilities

- **Question:** What is the probability that the average number of runs scored in a sample of size 25 is less than 3?
- **Answer:** ___________________ the probability
  - While we know the mean and standard deviation, we do not know the ______ (____________) of the sampling distribution
- **Question:** What would need to happen to be able to compute probabilities for runs scored?
- **Answer:** Need to sample ___________________
  - Allows us to apply Rule of Thumb ___: Shape of sampling distribution is ________________ of the shape of original population

Looking Ahead to Inference

- **Scenario:** Suppose we want to learn about the average SAT score for only incoming freshmen at Pitt. This will likely be higher than the national mean of 1000.
- **Question:** How can we determine if Pitt students score higher than 1000 on average if we don't know their population mean?
- **Answer:**
  - Make a ______ about the true population mean
  - Take a _____________ of incoming Pitt freshman
  - Create a __________ and find summary statistics (mean, standard deviation) to determine if the shape of the sample mean is _______
  - Compare the _____________ with our guess for the population mean to see if our guess is __________