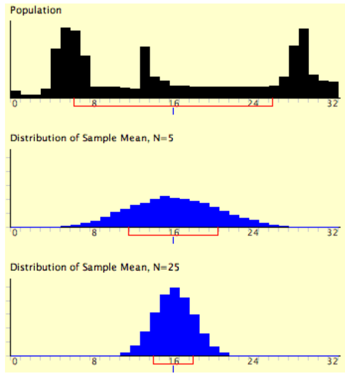


Chapter 18 – Sampling Distribution Models

How Many Heads?

- Flip a Coin 5 Times
  - Or use Minitab to simulate flipping the coin 5 times – see Chapter 14 Notes, “Coin Toss” Slide
  - Record the number of heads you saw in 5 flips: _________
  - Record the **proportion** of times you got heads: _________ (*)
- Flip a Coin 20 More Times
  - Total number of heads in the combined 25 flips: _________
  - Recompute the sample proportion (out of 25): __________ (*)
- Flip a Coin 25 More Times
  - Total number of heads in the combined 50 flips: _________
  - Recompute the sample proportion (out of 50): __________ (*)
- Enter the proportions (*) on the computer at the front…

Proportion in 5 Coin Flips

- We’ll make a dotplot of everyone’s 5 flip data
- Note: This is a distribution of sample proportions
  - The sample proportions are now the data…
- Describe the distribution
  - Shape?
  - Center?
  - Spread?

Proportion in 25 Flips

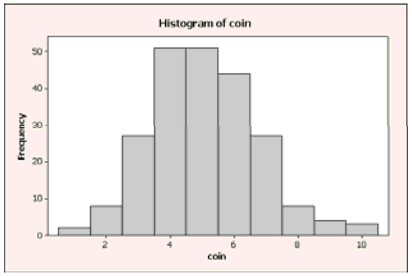
- We’ll make a new dotplot of the 25 flip data
- Describe the distribution and compare it to the 5 flip data
  - Shape?
  - Center?
  - Spread?
- Why do you think the center and spread for the new distribution are this way?

Proportion in 50 Flips

- We'll make a new dotplot of the 50 flip data
- Describe the distribution and compare it to the 25 and 5 flip data
  - Shape?
  - Center?
  - Spread?
- Why do you think the center and spread for the new distribution are this way?

But We've Seen This Before…

- Recall the Math 140 Survey Data
  - http://www.canyons.edu/faculty/morrowa/140/datasets
  - The question for "coin" was "Flip a coin 10 times. How many times did you get tails?"



Why Is This Happening?!?!?

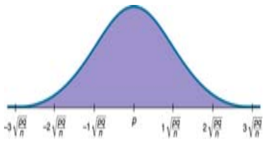
- What model describes counting the <u>number</u> of heads in a *sample* of n flips of a coin?
- What model describes the <u>proportion</u> of heads in a *sample* of n flips of a coin?
  - To convert to a proportion, we must divide out n.
  - What is the effect of a rescale?
    - On shape? Center? Spread?

What 'normal' is it?

- A **<u>sampling distribution model</u>** shows us the behavior of the statistic over all possible samples for the same size n.
- The **<u>sampling variability</u>**, or **<u>sampling error</u>**, is the variability we expect to see from one random sample to another.
- **<u>Sampling Distribution Model for a Proportion</u>**
  - Under the conditions below, the sampling distribution of $\hat{p}$ is modeled by

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

  - Conditions
    - Independence
      1. Randomization Condition
      2. 10% Condition
    - Sample Size Large Enough
      1. Success/Failure Condition

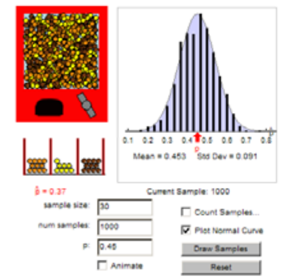


Example: Smoking

- Public health statistics indicate that 26.4% of American adults smoke cigarettes.
  - We randomly pick 100 American adults. Let $\hat{p}$ represent the proportion of American adults who smoke.
  1. What is an appropriate model for the distribution of $\hat{p}$?
     a) Specify the name of the distribution, the mean, and the SD
     b) Verify that the conditions are met
  2. Approximately what is the probability that one quarter of the sample will smoke?
  3. Between what two numbers are 68% of all estimates?
     - Recall: Normal… 68% ~ _____ SDs

Exploring Effects of n and p

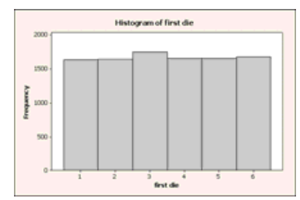
- On Your Own (Later)
  - Explore effects of different n and p
  - http://statweb.calpoly.edu/chance/applets/Reeses/ReesesPieces.html
  - Sample Size = n
  - Num Samples = LARGE
  - p = proportion of orange
- If taking a large number of samples, UNCHECK the box for Animate



Quantitative Data: Exploring Dice

- Let's use Minitab to simulating rolling dice…
  - Create a theoretical die in C1 by typing the numbers 1 to 6.
- We start with 1 die.
  - Simulate 10,000 rolls of the die
    - Calc > Random Data > Sample from Columns
    - Rows to Generate: 10000
    - From Columns: C1
    - Store in: C2
    - Sample with Replacement
  - Make a histogram of the 10,000 samples
  - Calculate the mean and SD of the 10,000 samples

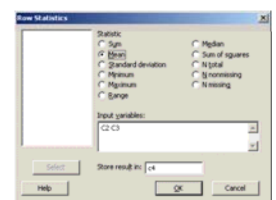
Just Checking…



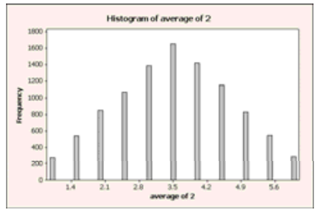
Descriptive Statistics: first die

Variable	Mean	StDev
first die	3.5097	1.7011

- Now Generate 10,000 more samples in C3
  - This gives us a total of 10,000 samples of 2 rolls of dice
- Next calculate the averages of each pair of rolls and store the result in C4
  - Calc > Row Statistics
  - Statistic: Mean
  - Input variables: C2 C3
  - Store in: C4
- Plot Histogram for C4
- Find mean, SD for C4



Just Checking…

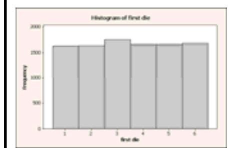


Descriptive Statistics: average of 2

Variable	Mean	StDev
average of 2	3.5133	1.2045

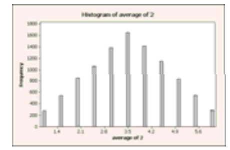
- Delete the Averages
- Put 10,000 samples of dice in each of C4, C5, and C6
- Using Row Statistics, find the average of each row across C2-C6. Store these in C7.
- Plot a histogram of the new averages
- Find the mean and SD for the averages

Just Checking:



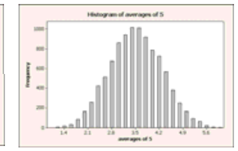
first die

Variable	Mean	StDev
first die	3.5097	1.7011



average of 2

Variable	Mean	StDev
average of 2	3.5133	1.2045



averages of 5

Variable	Mean	StDev
averages of 5	3.5064	0.7679

- As we increase the number of dice, what changes
  - Shape?
  - Center?
  - Spread?

The Central Limit Theorem

- **Sampling Distribution Model for a Mean**
  - Under the conditions below, the sampling distribution of $\bar{y}$ is modeled by

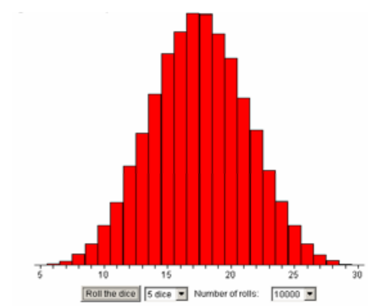
$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

  - Conditions
    - Independence
      1. Randomization Condition
      2. 10% Condition
    - Sample Size Large Enough

- This result is called the **Central Limit Theorem** and is one of the most important results from probability theory.
  - Note: The distribution of the original data is (almost) irrelevant.

On Your Own:
Another Demonstration of CLT

- http://www.stat.sc.edu/~west/javahtml/CLT.html



Example: A population of soft drink cans has amounts of liquid following a normal distribution with mean 12.0oz and standard deviation 0.2 oz.

1. What fraction of cans are between 11.9 and 12.1 oz?

2. What is the probability that the mean amount of liquid for 16 cans is between 11.9 and 12.1 oz?

Example: Given the population of men has normally distributed weights with a mean of 172 lb and a standard deviation of 29 lb,

1. If one man is randomly selected, find the probability that his weight is greater than 167 lb.

2. If 12 different men are randomly selected, find the probability that their mean weight is greater than 167 lb.

- Note: It is much easier for an individual to deviate from the mean than it is for a group of 12 to deviate from the mean.

Recap

- $\hat{p}$
  - The mean of $\hat{p}$ is E($\hat{p}$) = ______
  - The SD of $\hat{p}$ is SD($\hat{p}$) = _______
  - For large n, the distribution of $\hat{p}$ is approximately _______
- $\bar{y}$
  - The mean of $\bar{y}$ is E($\bar{y}$) = ______
  - The SD of $\bar{y}$ is SD($\bar{y}$) = _______
  - For large n, the distribution of $\bar{y}$ is approximately _______
- What happens to the SD as n increases?

Class Work

- To get credit, it is your responsibility to get checked off.
  1. Chapter 18 Handout
     - Checking solutions? No pens in the front!

Homework

- Textbook/Routine Homework
  - Due Next Week (25% chance of collection)
  1. Read Chapter 18
  2. Pg 479-485: #1, 3, 5, 7, 29, 31, 37, 47, 48, 49

Answers to Evens #s
C18 #48
(a) 0.048, (b) 0.863,
(c) 0.0019, (d) 0

- Project/Exploration Homework
  - None at the moment