MATH 3339 - 03 15951
Statistics for the Sciences
Sampling Distribution for the sample mean

Wendy Wang
wwang60@central.uh.edu

Lecture 25 - 3339
Outline

1. Approximating the Binomial Distribution
2. Sums of Random Variables
3. Sampling Distributions
4. Sampling Distribution of $\bar{X}$
5. Finding Probabilities for $\bar{X}$
Popper Set Up

- Fill in all of the proper bubbles.
- Make sure your ID number is correct.
- Make sure the filled in circles are very dark.
- This is popper number 21.
Approximation for Binomial

Suppose a random variable $X$ has a binomial distribution with $p = 0.1$. The following is a histogram with $n = 10$. 

$n = 10, p = 0.1$
Approximation for Binomial

Suppose a random variable $X$ has a binomial distribution with $p = 0.1$. The following is a histogram with $n = 20$. 

$n = 20, p = 0.1$
Approximation for Binomial

Suppose a random variable $X$ has a binomial distribution with $p = 0.1$. The following is a histogram with $n = 50$. 

$n = 50, p = 0.1$
Approximation for Binomial

Suppose a random variable $X$ has a binomial distribution with $p = 0.1$. The following is a histogram with $n = 100$.

$$X \sim \text{Binom}(n = 100, p = 0.1)$$

$$\mu_X = n \cdot p = 10 \times 0.1 = 1$$

$$\sigma_X = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{10 \cdot 0.1 \cdot 0.9} = 3$$
Theorem 5.3

Let $X$ be a binomial random variable based on $n$ trials with success probability $p$. Then if the binomial probability histogram is not too skewed, $X$ has an approximate Normal distribution with $\mu = np$ and $\sigma = \sqrt{np(1-p)}$. In particular, for $x = a$ possible value of $X$,

$$P(X \leq x) = \text{Binom}(x; n, p) \approx \ (\text{area under the normal curve to the left of } x + 0.5) \approx \Phi \left( \frac{x + 0.5 - np}{\sqrt{np(1-p)}} \right)$$

In practice, the approximate is adequate provided that both $np \geq 10$ and $n(1-p) \geq 10$. 
Continuity correction

- When approximating the binomial with the normal distribution, a better approximation is obtained by applying the **continuity correction**.

- This means to adjust the inequalities describing events to avoid points of discontinuity of the binomial distribution, i.e., the possible values of the discrete random variable $Y$. Since $Y$ is an integer, the events $(Y > 18)$ and $(Y > 18.5)$ are actually the same.
Example of Normal Approximation

Suppose that your mail-order company advertises that it ships 90% of its orders within three working days. Suppose you take a simple random sample of 100 orders:

1. What is the probability that 86 or fewer of the orders are shipped on time?

\[ P(X \leq 86) = \text{pbinom}(86, 100, 0.9) \]

\[ P(X \leq 86) = \text{pnorm}(86.5, 90, \sqrt{100 \cdot 0.9 \cdot 0.1}) \]

2. What is the probability that more than 95 of the orders are shipped on time?

\[ P(X > 95) = 1 - P(X \leq 95) = 1 - \text{pbinom}(95, 100, 0.9) \]

or

\[ = 1 - \text{pnorm}(95.5, 100, 0.9, \sqrt{100 \cdot 0.9 \cdot 0.1}) \]
Chapter 6: Joint Distributions and Sampling Distributions
Recall $E(X + Y)$

- If $X$ and $Y$ are two different random variables, then the expected value (mean) of the sums of the pairs of the random variable is the same as the sum of their means:

  $\mu_{X+Y} = E(X + Y) = E(X) + E(Y) = \mu_X + \mu_Y$.

  This is called the addition rule for means.

- The expected value (mean) of the difference of the pairs of the random variable is the same as the difference of their means:

  $\mu_{X-Y} = E(X - Y) = E(X) - E(Y) = \mu_X - \mu_Y$. 

Recall $\text{VAR}(X + Y)$

If $X$ and $Y$ are independent random variables,

\[ \sigma^2_{X+Y} = \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) = \sigma^2_X + \sigma^2_Y \]

and

\[ \sigma^2_{X-Y} = \text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) = \sigma^2_X + \sigma^2_Y \]
If X & Y are dependent

If X and Y are dependent random variables then

\[ \sigma^2_{X+Y} = \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{cov}(X, Y) = \sigma^2_X + \sigma^2_Y + 2\text{cov}(X, Y) \]

and

\[ \sigma^2_{X-Y} = \text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{cov}(X, Y) = \sigma^2_X + \sigma^2_Y - 2\text{cov}(X, Y) \]

Note: if X and Y are independent
then \( \text{cov}(X, Y) = 0 \)
Example

Suppose we have two independent random variables, $X$ and $Y$ where $\mu_X = 10$, $\sigma_X = 2$, $\mu_Y = 10$ and $\sigma_Y = 2$.

a. Determine: $\mu_{X+Y}$ and $\sigma_{X+Y}$

$$
\mu_{X+Y} = \mu_X + \mu_Y = 10 + 10 = 20
$$

$$
\sigma_{X+Y} = \sqrt{\text{Var}(X+Y)} = \sqrt{\text{Var}(X) + \text{Var}(Y)} = \sqrt{2^2 + 2^2} = \sqrt{8} = 2\sqrt{2}
$$

b. Suppose we want the mean of $X$ and $Y$, what would be the expected value of the mean?

$$
E\left( \frac{X+Y}{2} \right) = E\left( \frac{X}{2} \right) + E\left( \frac{Y}{2} \right)
$$

$$
= \frac{1}{2} E(X) + \frac{1}{2} E(Y)
$$

$$
= \frac{1}{2} (10) + \frac{1}{2} (10) = 10
$$
Consider one family as a population of five children. We are looking at the ages of these five children: 3, 5, 9, 11, 14.

1. Determine the population mean, \( \mu \), age of these children.
   \[ \mu = \frac{3 + 5 + 9 + 11 + 14}{5} = \frac{42}{5} = 8.4 \]
   a. 9   b. 10   c. 8.4   d. 11

2. Determine the population standard deviation, \( \sigma \), of these children.
   \[ \sigma = \sqrt{\frac{1}{N} \sum (X - \mu)^2} \]
   a. 10   b. 4   c. 8.4   0

3. Suppose we take a sample of 2 children from this population. What would we expect the sample mean, \( \bar{x} \) from the 2 children to be?
   a. 2   b. 8.4   c. 4   d. 16

#4, choose d  #5, choose e
Sampling Distribution

Sampling Distribution for the sample mean
**Sampling Distribution of size 2**

From the five children, we want to list out all possible pairs of size 2 and determine their mean. Ages are: 3, 5, 9, 11, 14

<table>
<thead>
<tr>
<th>Pairs</th>
<th>Sample mean, $\bar{x}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3,5)</td>
<td>4 = ( \frac{3+5}{2} )</td>
</tr>
<tr>
<td>(3,9)</td>
<td>6</td>
</tr>
<tr>
<td>(3,11)</td>
<td>7</td>
</tr>
<tr>
<td>(3,14)</td>
<td>8.5</td>
</tr>
<tr>
<td>(5,9)</td>
<td>7</td>
</tr>
<tr>
<td>(5,11)</td>
<td>8 = ( \frac{5+11}{2} )</td>
</tr>
<tr>
<td>(5,14)</td>
<td>9.5</td>
</tr>
<tr>
<td>(9,11)</td>
<td>10</td>
</tr>
<tr>
<td>(9,14)</td>
<td>11.5</td>
</tr>
<tr>
<td>(11,14)</td>
<td>12.5</td>
</tr>
</tbody>
</table>

The list above is a sampling distribution from a sample of 2 of $\bar{x}$, the possible values of the sample mean. What is the mean of the sample means, $\mu_{\bar{x}}$? What is the standard deviation of the sample means, $\sigma_{\bar{x}}$?
What about the sampling distribution of size 3 from the family of five?

<table>
<thead>
<tr>
<th>Sets</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3, 5, 14)</td>
<td>10</td>
</tr>
<tr>
<td>(3, 9, 14)</td>
<td>9.3333</td>
</tr>
<tr>
<td>(3, 9, 11)</td>
<td>8.3333</td>
</tr>
<tr>
<td>(3, 5, 11)</td>
<td>7.6667</td>
</tr>
<tr>
<td>(3, 11, 14)</td>
<td>7</td>
</tr>
<tr>
<td>(5, 11, 14)</td>
<td>6.3333</td>
</tr>
<tr>
<td>(5, 9, 14)</td>
<td>6.6667</td>
</tr>
<tr>
<td>(5, 9, 11)</td>
<td>5.6667</td>
</tr>
<tr>
<td>(9, 11, 14)</td>
<td>5.3333</td>
</tr>
<tr>
<td>(9, 9, 11)</td>
<td>4.3333</td>
</tr>
<tr>
<td>(9, 11, 9)</td>
<td>3.6667</td>
</tr>
</tbody>
</table>

What is the mean of these means, $\mu_{\bar{x}}$? What is the standard deviation of these means, $\sigma_{\bar{x}}$?

$\bar{x} = \frac{\sum x}{n} = \frac{1.6248}{10}$

$\mu_{\bar{x}} = 1.6248$
Sampling distribution

- When we describe distributions we use three characteristics:
  - Shape
  - Center
  - Spread

- To describe the sampling distribution we can use the same three characteristics.

- This can be shown through histograms or numerical values.
Sampling Distribution of $\bar{X}$

- Suppose that $\bar{X}$ is the sample mean of a simple random sample of size $n$ from a large population with mean $\mu$ and standard deviation $\sigma$.

- $\bar{X}$ is a random variable because every time we take a random sample we will not get the same sample mean $\bar{X}$. Thus we want to know the distribution of the sample means $\bar{X}$.

- The center of the sample means (mean of the sample means) $\mu_{\bar{X}}$ is $\mu$. Also called the expected value.

- The spread of the sample means (standard deviation of the sample means) $\sigma_{\bar{X}}$ is $\frac{\sigma}{\sqrt{n}}$. 
Sampling Distribution Example

Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz.

We take a sample of 25 cans and find the mean amount $\bar{X}$ in these 25 cans. What would we expect the mean to be? Would the sample mean be exactly that value? If not how far off could the sample mean be?

$\mu = 12$

$n = 25$

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.09}{\sqrt{25}} = 0.018$
Sampling Distribution Example

Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz.

- We take a sample of 100 cans and find the mean amount $\bar{X}$ in these 100 cans. What would we expect the mean to be? Would the sample mean be exactly that value? If not how far off could the sample mean be?

\[ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.09}{\sqrt{100}} = 0.009 \]
Shape of the Sample Mean Distribution

- If a population has a Normal distribution, then the sample mean $\bar{X}$ of $n$ independent observations also has a Normal distribution with mean $\mu$ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

- **Central limit theorem:** For any population, when $n$ is large ($n > 30$), the sampling distribution of the sample mean $\bar{X}$ is approximately a Normal distribution with mean $\mu$ and standard deviation $\frac{\sigma}{\sqrt{n}}$. 

$x \sim N(\mu, \sigma)$ 

$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$
Example: Amount of Pepsi

Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz. Suppose that a random sample of 4 cans are examined, describe the distribution of the sample means $\bar{X}$.

- Center: $\mu_{\bar{X}} = \mu = 12$
- Spread: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.09}{\sqrt{4}} = 0.045$
- Shape: Unknown because we do not know the original distribution and the sample size is small.
Example: Amount of Pepsi

Assume that cans of Pepsi are filled so that the actual amount have a mean \( \mu = 12 \) oz and a standard deviation \( \sigma = 0.09 \) oz. Suppose that a random sample of 4 cans are examined, describe the distribution of the sample means \( \bar{X} \).

- **Center:** \( \mu_{\bar{X}} = \mu = 12 \)
- **Spread:** \( \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.09}{\sqrt{4}} = 0.045 \)
- **Shape:** Unknown because we do not know the original distribution and the sample size is small.
Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz. Suppose that a random sample of 4 cans are examined, describe the distribution of the sample means $\bar{X}$.

- Center: $\mu_{\bar{X}} = \mu = 12$
- Spread: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.09}{\sqrt{4}} = 0.045$
- Shape: Unknown because we do not know the original distribution and the sample size is small.
Example: Amount of Pepsi

Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz. Suppose that a random sample of 4 cans are examined, describe the distribution of the sample means $\bar{X}$.

- **Center**: $\mu_{\bar{X}} = \mu = 12$
- **Spread**: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.09}{\sqrt{4}} = 0.045$
- **Shape**: Unknown because we do not know the original distribution and the sample size is small.
Example: Amount of Pepsi

Assume that cans of Pepsi are filled so that the actual amount have a mean \( \mu = 12 \) oz and a standard deviation \( \sigma = 0.09 \) oz. Suppose that a random sample of 100 cans are examined, describe the distribution of the sample means \( \bar{X} \).

- **Center:** \( \mu_{\bar{X}} = \mu = 12 \)
- **Spread:** \( \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.09}{\sqrt{100}} = 0.009 \)
- **Shape:** Normal because we have a large sample thus we can apply the **Central Limit Theorem**.
Example: Amount of Pepsi

Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz. Suppose that a random sample of 100 cans are examined, describe the distribution of the sample means $\bar{X}$.

- **Center:** $\mu_{\bar{X}} = \mu = 12$
- **Spread:** $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.09}{\sqrt{100}} = 0.009$
- **Shape:** Normal because we have a large sample thus we can apply the Central Limit Theorem.
Example: Amount of Pepsi

Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz. Suppose that a random sample of 100 cans are examined, describe the distribution of the sample means $\bar{X}$.

- **Center:** $\mu_{\bar{X}} = \mu = 12$
- **Spread:** $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.09}{\sqrt{100}} = 0.009$
- **Shape:** Normal because we have a large sample thus we can apply the **Central Limit Theorem**.
Example: Amount of Pepsi

Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz. Suppose that a random sample of 100 cans are examined, describe the distribution of the sample means $\bar{X}$.

- **Center:** $\mu_{\bar{X}} = \mu = 12$
- **Spread:** $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.09}{\sqrt{100}} = 0.009$
- **Shape:** Normal because we have a large sample thus we can apply the **Central Limit Theorem**.
Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz. Suppose that a random sample of 36 cans are examined, determine the probability that a sample of 36 cans will have a sample mean amount, $\bar{X}$ of at least 12.01 oz.

To find this probability we need to first describe the distribution:

- **Shape**: Normal because of the Central Limit Theorem
- **Center**: $E[\bar{X}] = \mu_{\bar{x}} = \mu = 12$
- **Spread**: $SD[\bar{X}] = \sigma_{\bar{x}} = \sigma/\sqrt{n} = 0.09/\sqrt{36} = 0.015$ this is the standard deviation we use.

We want to know: $P(\bar{X} \geq 12.01)$
Finding Probabilities

Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz. Suppose that a random sample of 36 cans are examined, determine the probability that a sample of 36 cans will have a sample mean amount, $\bar{X}$ of at least 12.01 oz.

- To find this probability we need to first describe the distribution:
  - Shape: Normal because of the Central Limit Theorem
  - Center: $E[\bar{X}] = \mu_{\bar{X}} = \mu = 12$
  - Spread: $SD[\bar{X}] = \sigma_{\bar{X}} = \sigma / \sqrt{n} = 0.09 / \sqrt{36} = 0.015$ this is the standard deviation we use.

- We want to know: $P(\bar{X} \geq 12.01)$
Finding Probabilities

Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz. Suppose that a random sample of 36 cans are examined, determine the probability that a sample of 36 cans will have a sample mean amount, $\bar{X}$ of at least 12.01 oz.

- To find this probability we need to first describe the distribution:
  - **Shape:** Normal because of the **Central Limit Theorem**
  - **Center:** $E[\bar{X}] = \mu_{\bar{X}} = \mu = 12$
  - **Spread:** $SD[\bar{X}] = \sigma_{\bar{X}} = \sigma/\sqrt{n} = 0.09/\sqrt{36} = 0.015$ this is the standard deviation we use.

- We want to know: $P(\bar{X} \geq 12.01)$
Finding Probabilities

Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz. Suppose that a random sample of 36 cans are examined, determine the probability that a sample of 36 cans will have a sample mean amount, $\bar{X}$ of at least 12.01 oz.

- To find this probability we need to first describe the distribution:
  - Shape: Normal because of the Central Limit Theorem
  - Center: $E[\bar{X}] = \mu_{\bar{X}} = \mu = 12$
  - Spread: $SD[\bar{X}] = \sigma_{\bar{X}} = \sigma/\sqrt{n} = 0.09/\sqrt{36} = 0.015$ this is the standard deviation we use.

- We want to know: $P(\bar{X} \geq 12.01)$

$$= 1 - P(\bar{X} < 12.01)$$

$$= 1 - \text{pnorm}(12.01, 12, 0.015)$$
Notes about finding probabilities for $\bar{X}$

- We have a sample size $n$. Thus the standard deviation changes by that value. $SD(\bar{X}) = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

- The mean stays the same. $\text{mean}(\bar{X}) = \mu_{\bar{X}} = \mu$.

- If we know that the original distribution is Normal or we have a large enough sample ($n > 30$). We can use the Normal distributions to find the probabilities.
Orange Juice

An orange juice producer buys all his oranges from a large orange grove. The amount of juice squeezed from each of these oranges is approximately normally distributed, with a mean of 4.70 ounces and a standard deviation of 0.40 ounce. Suppose we take a random sample of 4 oranges and determine the mean of this sample, $\bar{X}$.

1. What is the shape of the sampling distribution of $\bar{X}$.
   
   Normal, because population is normally distributed

2. What is the mean of the sampling distribution of $\bar{X}$.
   
   $\mu_{\bar{X}} = \mu = 4.7$

3. What is the standard deviation of the sampling distribution of $\bar{X}$.
   
   $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.4}{\sqrt{4}} = 0.2$

4. What is the probability that the sample mean of the 4 oranges will be at 4.5 or less?
   
   $P(\bar{X} \leq 4.5) = \text{pnorm}(4.5, 4.7, 0.2) = 0.1586$