popper : choose a.

MATH 3339
Statistics for the Sciences
Chapter 6

Wendy Wang
wwang60@central.uh.edu

Lecture 12 - 3339

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Outline

Sampling Distribution

Sampling Distribution for the sample mean $\bar{X}$

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

choose(5, 2)
=10

Pairs	Sample mean, $\bar{x}$
(3,5)	4 $= \frac{3+5}{2}$
(3,9)	6 $= \frac{3+9}{2}$
(3,11)	7
(3, 14)	8.5
(5, 9)	7
(5, 11)	8
(5, 14)	9.5
(9, 11)	10
(9, 14)	11.5
(11, 14)	12.5 $= \frac{11+14}{2}$

$\mu = \frac{3+5+\cdots+14}{5}$

$= 8.4$

$\bar{x} = c(4, 6, 7, 8.5, \ldots 12.5)$

$\mu_{\bar{x}} = 8.4$

sd($\bar{x}$)

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, $\sigma_{\bar{x}}$?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Sampling Distribution of size 3

choose(5, 3) = 10

What about the sampling distribution of size 3 from the family of five?

Sets	$\bar{x}$
(3, 5, 9)	5.6667
(3, 5, 11)	6.3333
(3, 5, 14)	7.3333
(3, 9, 11)	7.6667
(3, 9, 14)	8.6667
(3, 11, 14)	9.3333
(5, 9, 11)	8.3333
(5, 9, 14)	9.3333
(5, 11, 14)	10
(9, 11, 14)	11.3333

$= \frac{3+5+9}{3}$

$= \frac{5+11+14}{3}$

$\mu_{\bar{x}} = \frac{5.6667 + \cdots 11.333}{10}$

$= 8.4$

$sd(\bar{x}) =$

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

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

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.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Sampling Distribution of $\bar{X}$

large population; μ, σ

sample: n $\bar{x}$

- Suppose that $\bar{X}$ is the sample mean of a simple random sample of size n from a large population with mean μ and standard deviation σ.
- $\bar{X}$ 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 μ. Also called the **expected value.**

  $E(\bar{x}) = \mu$

  $\mu_{\bar{x}} = \mu$
- The spread of the sample means (standard deviation of the sample means) $\sigma_{\bar{X}}$ is $\sigma/\sqrt{n}$.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Shape of the Sample Mean Distribution

μ, σ comp from the population

$$\bar{X} \sim N\left(\mu_{\bar{X}} = \mu,\ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\right)$$

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

"CLT"

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.

$n > 30 \qquad \bar{X} \overset{\text{approx}}{\sim} N$

Example: Amount of Pepsi

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

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.**

$\bar{X} \overset{\text{approx}}{\sim} N(\mu_{\bar{X}} = 12, \ \sigma_{\bar{X}} = 0.009)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

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**, X of at least 12.01 oz.

$P(\bar{X} \geq 12.01)$

- 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: $P(\bar{X} \geq 12.01)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

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: $P(\bar{X} \geq 12.01)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

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: $\underline{P(\bar{X} \geq 12.01)}$ $\quad \bar{X} \sim N(12, 0.015)$

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

```
> 1-pnorm(12.01,12,.015)
[1] 0.2524925
```

$= 1 - \text{pnorm}(12.01, 12, 0.015)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Notes about finding probabilities for $\bar{X}$

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

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Orange Juice $\to \bar{X} \sim N(\mu = 4.7,\ \sigma = 0.40)$

$n = 4$

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 $\bar{X}$.

   $\bar{X}$ will be Normal

2. What is the mean of the sampling distribution of $\bar{X}$.

   $\mu_{\bar{X}} = \mu = 4.7$

   $\bar{X} \sim N(4.7, 0.20)$

3. What is the standard deviation of the sampling distribution of $\bar{X}$.

   $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{.40}{\sqrt{4}} = 0.20$

4. What is the probability that the sample mean of the 4 oranges will be at 4.5 or less?

   $P(\bar{X} \leq 4.5) = \text{pnorm}(4.5, 4.7, 0.20)$

Sample Proportions

$\hat{p}$

- The population proportion is p a parameter. In some cases we do not know the population proportion, thus we use the sample proportion, $\hat{p}$ to estimate p.
- The sample proportion is calculated by: $\hat{p} = \frac{X}{n}$
- X = the number of observations of interest in the sample or the number of "successes" in the sample.
- n = the sample size or number of observations.
- Note that $X \sim \text{Bin}(n, p)$ and can be approximated by the Normal distribution with $\mu_X = E(X) = np$ and $\sigma_X = SD(X) = \sqrt{np(1-p)}$ as long as $np \geq 10$ and $np(1-p) \geq 10$.
- Now we want to know how is $\hat{p} = \frac{X}{n}$ distributed. Thus we want to know $\mu_{\hat{p}} = E(\hat{p})$ and $\sigma_{\hat{p}} = SD(\hat{p})$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example

$p = 0.34$

- According to the National Retail Federation, 34% of taxpayers used computer software to do their taxes.

$n = 50$

- A sample of 50 taxpayers was selected, what do we expect the sample proportion $\hat{p}$ to be?

$\hat{p} = \frac{\#\text{ of use computer}}{50}$

- If we take other samples will the sample proportions always be the same value?

- If not what would $\hat{p}$ be off by?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Shape of the distribution of $\hat{p}$

We can use the **Normal distribution** as long as

- $np \geq 10$ the number of successes are at least 10
- $n(1-p) \geq 10$ the number of failures are at least 10.

Center of the distribution of $\hat{p}$

- The center is the mean (expected value): $\mu_{\hat{p}} = p$ the proportion of success.
- $\hat{p} = \frac{X}{n}$ where X is the number of **successes** out of n observations. Thus X has a binomial distribution with parameters n and p.
- The mean of X is:

$$\mu_X = E(X) = np$$

- Thus the mean of $\hat{p}$ is:

$$\mu_{\hat{p}} = E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{\mu_X}{n} = \frac{np}{n} = p$$

Spread of the distribution of $\hat{p}$

- The spread is the standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.
- The variance of X is:

$$\sigma_X^2 = Var(X) = np(1-p)$$

- The variance of $\hat{p}$ is:

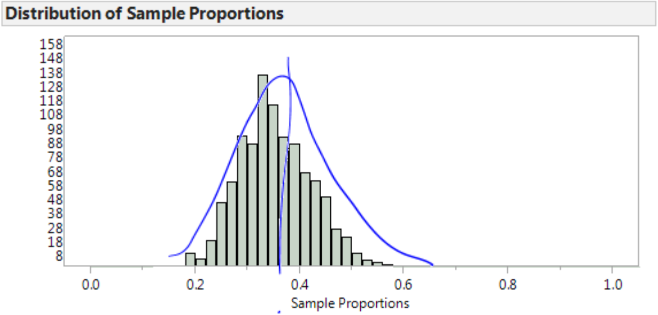
$$\sigma_{\hat{p}}^2 = Var(\hat{p}) = Var\left(\frac{X}{n}\right) = \frac{Var(X)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

$$\hat{p} \sim N\left(\mu_{\hat{p}} = p,\ \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}\right)$$

if $np \geq 10$ and $n(1-p) \geq 10$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Sample Distribution of $n = 50$.

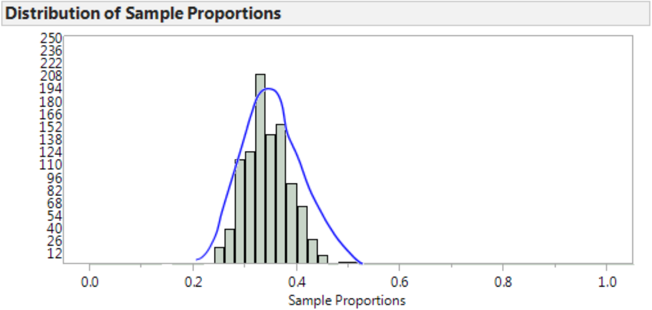


Summary Statistics

Mean of Sample Proportions	0.34244
Std Dev of Sample Proportions	0.06987
No.of Samples	1000

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Sample Distribution of $n = 125$



Summary Statistics

Mean of Sample Proportions	0.34082
Std Dev of Sample Proportions	0.04375
No.of Samples	1000

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Assumptions

- The sampled values must be random and independent of each other. This can be tested by **10% Condition**: The sample size must be no larger than 10% of the population.
- The sample size, n must be large enough. This can be be tested by **Success / Failure Condition**: The sample size has to be big enough so that both np and $n(1-p)$ at least 10.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example for distribution of $\hat{p}$

$p = 0.34$

According to the National Retail Federation, 34% of taxpayers used computer software to do their taxes. A sample of 125 taxpayers was selected. What is the distribution of $\hat{p}$, the sample proportion of the 125 taxpayers that used computer software to do their taxes?

$n = 125$

1. Check if we can use the Normal distribution.
   - $p = 0.34$, $n = 125$
   - $np = 125(0.34) = 42.5$ > 10
   - $n(1-p) = 125(1-0.34) = 125(0.66) = 82.5$ > 10
   - Both np and $n(1-p)$ are greater than 10 so we can use the Normal distribution.

$\hat{p} \sim N(\mu_{\hat{p}}, \sigma_{\hat{p}})$

2. The mean is: $\mu_{\hat{p}} = p = 0.34$. If we take a sample we "expect" 34% to have used computer software to do their taxes.
3. The standard deviation is:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.34(1-0.34)}{125}} = 0.0424$$

Example continued

$\rightarrow \hat{p} \sim N(0.34, 0.0424)$

According to the National Retail Federation, 34% of taxpayers used computer software to do their taxes. A sample of 125 taxpayers was selected. What is the probability that between 28% and 40% of the taxpayers from the sample of 125 used computer software to do their taxes?

1. We want: $P(0.28 < \hat{p} < 0.40)$

$= P(\hat{p} < 0.40) - P(\hat{p} < 0.28)$

$= pnorm(0.40, 0.34, 0.0424)$

$- pnorm(0.28, 0.34, 0.0424)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Facebook Example

The Social Media and Personal Responsibility Survey in 2010 found the 69% of parents are "friends" with their children on Facebook. A random sample of 140 parents was selected and we determined the proportion of parents from this sample, $\hat{p}$ that are "friends" with their children on Facebook.

1. What is the shape of the sampling distribution of $\hat{p}$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Facebook Example

The Social Media and Personal Responsibility Survey in 2010 found the 69% of parents are "friends" with their children on Facebook. A random sample of 140 parents was selected and we determined the proportion of parents from this sample, $\hat{p}$ that are "friends" with their children on Facebook.

2. What is the mean of the sampling distribution of $\hat{p}$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Facebook Example

The Social Media and Personal Responsibility Survey in 2010 found the 69% of parents are "friends" with their children on Facebook. A random sample of 140 parents was selected and we determined the proportion of parents from this sample, $\hat{p}$ that are "friends" with their children on Facebook.

3. What is the standard deviation of the sampling distribution of $\hat{p}$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Facebook Example

The Social Media and Personal Responsibility Survey in 2010 found the 69% of parents are "friends" with their children on Facebook. A random sample of 140 parents was selected and we determined the proportion of parents from this sample, $\hat{p}$ that are "friends" with their children on Facebook.

4. What is the probability that the sample proportion of 140 parents is greater than 72%?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Voting Questions

$np > 10$

① check assumption. $n(1-p) > 10$

Suppose that 52% voted for a certain candidate. We take a random sample of 1500 likely voters. Determine the following probabilities.

1. What is the probability that the sample proportion is less than 0.50?

$$\hat{p} \sim N\left(0.52, \sqrt{\frac{0.52*(1-0.52)}{1500}}\right)$$

$$P(\hat{p} < 0.50) = \text{pnorm}\left(0.50, 0.52, \sqrt{\frac{0.52*(1-0.52)}{1500}}\right)$$

2. What is the probability that the sample proportion is within 3% of the population proportion?

$$P(p-0.03 < \hat{p} < p+0.03)$$

$$= P(0.52-0.03 < \hat{p} < 0.52+0.03)$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS