

Sampling Distributions

Research Process

- Idea for study - Question posed
- Information collected
 - Data
 - Experiment
 - Survey
- Information - organized/analyzed
- Conclusions drawn based on analysis of data

Data

- Sample - use to make inferences about population
 - Sample mean - use to draw conclusions about population mean
 - Dependent upon sample
 - Changes based on values in sample

Inference

- A logical process of drawing conclusions from a collection of data and relationships between data and potential conclusions
- Assumption based on an observation
- A determination arrived at by reasoning; using facts to arrive at a broader conclusion

Inference

- The reasoning involved in drawing a conclusion or making a logical judgment on the basis of circumstantial evidence and prior conclusions rather than on the basis of direct observation
- The act or process of deriving a conclusion based solely on what one already knows

Data

- Use data to make inferences regarding the population
 - Problems:
 - Samples - not the population
 - Change sample then change
 - Data
 - Statistics
 - Conclusions - variable

Inferential Statistics

- **Based on probability statements**

Inferential Statistics

- Based on probability statements
- EX: Suppose the distribution for SAT scores for students entering the University of Florida in a recent year are roughly normal with mean 1100 points and standard deviation 180 points

Inferential Statistics

- Based on probability statements
- EX: Suppose the distribution for SAT scores for students entering the University of Florida in a recent year are roughly normal with mean 1100 points and standard deviation 180 points - we can make a probability statement regarding the likelihood that a randomly selected student has an SAT score of 1400 or higher.

Random Variable

- A variable characterized by random behavior in assuming its different possible values.
- Described by its probability distribution, which specifies the possible values of a random variable together with the probability associated with each value.

Random Variable

- A variable whose value is determined by the outcome of a random experiment.

Sample Mean

- Dependent upon the sample used
- If use a different sample then may get a different sample mean

To make Probability Statements for Sample Statistics ...

- Need to know the probability distribution for the sample statistics
 - Shape
 - Center
 - Spread
 - ...

Sampling Distribution for the Mean

- Probability distribution for all possible values of the random variable \bar{X} computed from a sample of size n from a population with mean μ and standard deviation σ .

Sampling Distribution for the Mean

- General idea for determining sampling distribution for mean \bar{X} :
 - Obtain sample of size n
 - Compute sample mean \bar{X}
 - Assuming that sampling from finite population, repeat until all possible simple random samples of size n have been obtained

Nursing Home Data

<http://lib.stat.cmu.edu/DASL/Datafiles/nursinghomedat.html>

Datafile Name:

Nursing Home Data

Datafile Subjects:

[Health](#) , [Consumer](#) , [Medical](#) , [Economics](#)

Story Names:

[Nursing Home Data](#)

Reference:

These data are part of the data analyzed in Howard L. Smith, Niell F. Piland, and Nancy Fisher, "A Comparison of Financial Performance, Organizational Characteristics, and Management Strategy Among Rural and Urban Nursing Facilities, Journal of Rural Health, Winter 1992, pp 27-40.

Authorization:

free use

Description:

The data were collected by the Department of Health and Social Services of the State of New Mexico and cover 52 of the 60 licensed nursing facilities in New Mexico in 1988.

Number of cases:

52

Variable Names:

1. BED = number of beds in home
2. MCDAYS = annual medical in-patient days (hundreds)
3. TDAYS = annual total patient days (hundreds)
4. PCREV = annual total patient care revenue (\$hundreds)
5. NSAL = annual nursing salaries (\$hundreds)
6. FEXP = annual facilities expenditures (\$hundreds)
7. RURAL = rural (1) and non-rural (0) homes

Nursing Home Data

- $C(52, 5) = 2,598,960$
- $C(52, 10) = 1.582002422 \times 10^{10}$
- $C(52, 20) = 1.259946279 \times 10^{14}$

Nursing Home Data

- $C(52, 5) = 2,598,960$
- $C(52, 10) = 1.582002422 \times 10^{10}$
- $C(52, 20) = 1.259946279 \times 10^{14}$
- Question: Could we make all the possible samples???

Nursing Home Data

- $C(52, 5) = 2,598,960$
- $C(52, 10) = 1.582002422 \times 10^{10}$
- $C(52, 20) = 1.259946279 \times 10^{14}$
- Question: Could we make all the possible samples????
- Let's explore some samples and the distribution for the sample means

Law of Large Numbers

- As additional observations are added to the sample, the difference between the sample mean, \bar{X} , and the population mean, μ , approaches zero.

Sampling Distribution

- As the sample size n increases, the standard deviation for the distribution of the \bar{X} decreases.
- Note: This decrease may not be as extreme as that provided in Example 2 and Example 3 in Section 8.1.

Mean and Standard Deviation for a Sampling Distribution

- Suppose a simple random sample of size n is taken from a population that has a mean μ and standard deviation σ . The sampling distribution of \bar{X} will have mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. $\sigma_{\bar{x}}$ is called the standard error of the mean.

Sampling Distribution from Population that is Approximately Normal

- Given a random variable X , if X is normally distributed with mean μ and standard deviation σ then the sample mean \bar{X} is normally distributed with mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Central Limit Theorem

- Given a random variable X with population mean μ and (population) standard deviation σ , for a random sample of size n taken from this population, the sampling distribution of \bar{X} becomes approximately normal as the sample size n increases. The mean of the distribution is $\mu_{\bar{x}} = \mu$ and the standard deviation is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Population with mean μ and standard deviation σ	Distribution of sample means	Mean of sample means	Standard deviation of sample means
Not normal with $n < 30$	Not normal	$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
Not normal with $n \geq 30$	Approximately normal	$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
Normal	Normal for any sample size	$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

- **Suppose a population has mean 30 units and a standard deviation of 5 units. If a sample of 100 is taken from the population, what is the probability that the sample mean will be more than 31 units?**

- Suppose a population has mean 30 units and a standard deviation of 5 units. If a sample of 100 is taken from the population, what is the probability that the sample mean will be more than 31 units?
- By the Central Limit Theorem, the sampling distribution of \bar{x} becomes approximately normal as the sample size increases.

- Suppose a population has mean 30 units and a standard deviation of 5 units. If a sample of 100 is taken from the population, what is the probability that the sample mean will be more than 31 units?
- We must calculate the mean and the standard deviation for the sampling distribution.

$$\mu_{\bar{x}} = 30 \quad \sigma_{\bar{x}} = \frac{5}{\sqrt{100}} = \frac{1}{2}$$

- Suppose a population has mean 30 units and a standard deviation of 5 units. If a sample of 100 is taken from the population, what is the probability that the sample mean will be more than 31 units?
- Using these, $\mu_{\bar{x}}=30$ and $\sigma_{\bar{x}}=\frac{1}{2}$, we determine the z-score.

$$z = \frac{31 - 30}{\left(\frac{1}{2}\right)} = 2$$

- Suppose a population has mean 30 units and a standard deviation of 5 units. If a sample of 100 is taken from the population, what is the probability that the sample mean will be more than 31 units?
- Then, we determine the probability.

$$\begin{aligned}P(X > 31) &= P(z > 2) \\ &= P(z < -2) \\ &= 0.0228\end{aligned}$$

- Suppose a population has mean 30 units and a standard deviation of 5 units. If a sample of 100 is taken from the population, what is the probability that the sample mean will be more than 31 units?
- Then, we determine the probability.

$$\begin{aligned} P(X > 31) &= P(z > 2) \\ &= 1 - P(z < 2) \\ &= 1 - 0.9772 \\ &= 0.0228 \end{aligned}$$

- As reported by the U.S. National Center for Health Statistics, the mean serum high-density-lipoprotein (HDL) cholesterol for females age 20-29 years old is $\mu = 53$. If serum HDL cholesterol is normally distributed with $\sigma = 13.4$, what is the probability that a random sample of 20 females age 20-29 years old will have a mean serum cholesterol above 60?
- NOTE: The author did NOT provide the units for HDL cholesterol in the text.

- As reported by the U.S. National Center for Health Statistics, the mean serum high-density-lipoprotein (HDL) cholesterol for females age 20-29 years old is $\mu = 53$ mg/dL. If serum HDL cholesterol is normally distributed with $\sigma = 13.4$ mg/dL, what is the probability that a random sample of 20 females age 20-29 years old will have a mean serum cholesterol above 60 mg/dL?

- NOTE: This version of the problem has been corrected to include the units for HDL cholesterol.

- As reported by the U.S. National Center for Health Statistics, the mean serum high-density-lipoprotein (HDL) cholesterol for females age 20-29 years old is $\mu = 53$ mg/dL. If serum HDL cholesterol is normally distributed with $\sigma = 13.4$ mg/dL, what is the probability that a random sample of 20 females age 20-29 years old will have a mean serum cholesterol above 60 mg/dL?

- We determine the mean and standard deviation.

$$\mu_{\bar{x}} = 53$$

$$\sigma_{\bar{x}} = \frac{13.4}{\sqrt{20}}$$

- Caution: Please remember not to round values used in intermediate calculations.

- As reported by the U.S. National Center for Health Statistics, the mean serum high-density-lipoprotein (HDL) cholesterol for females age 20-29 years old is $\mu = 53$ mg/dL. If serum HDL cholesterol is normally distributed with $\sigma = 13.4$ mg/dL, what is the probability that a random sample of 20 females age 20-29 years old will have a mean serum cholesterol above 60 mg/dL?
- Since the population is approximately normal, the sampling distribution of \bar{x} is approximately normal.

- As reported by the U.S. National Center for Health Statistics, the mean serum high-density-lipoprotein (HDL) cholesterol for females age 20-29 years old is $\mu = 53$ mg/dL. If serum HDL cholesterol is normally distributed with $\sigma = 13.4$ mg/dL, what is the probability that a random sample of 20 females age 20-29 years old will have a mean serum cholesterol above 60 mg/dL?

- We determine the z-score.

$$z = \frac{60 - 53}{\left(\frac{13.4}{\sqrt{20}} \right)} = 2.34$$

- Caution: Please remember not to round values used in intermediate calculations.

- As reported by the U.S. National Center for Health Statistics, the mean serum high-density-lipoprotein (HDL) cholesterol for females age 20-29 years old is $\mu = 53$ mg/dL. If serum HDL cholesterol is normally distributed with $\sigma = 13.4$ mg/dL, what is the probability that a random sample of 20 females age 20-29 years old will have a mean serum cholesterol above 60 mg/dL?

- We determine the z-score.

$$z = \frac{60 - 53}{\left(\frac{13.4}{\sqrt{20}} \right)} = 2.34$$

- Caution: Please remember that z-scores are recorded to two decimal places in the standard normal table.

- As reported by the U.S. National Center for Health Statistics, the mean serum high-density-lipoprotein (HDL) cholesterol for females age 20-29 years old is $\mu = 53$ mg/dL. If serum HDL cholesterol is normally distributed with $\sigma = 13.4$ mg/dL, what is the probability that a random sample of 20 females age 20-29 years old will have a mean serum cholesterol above 60 mg/dL?

- We determine the z-score.

$$z = \frac{60 - 53}{\left(\frac{13.4}{\sqrt{20}} \right)} = 2.34$$

- That is, remember to round your z-score to two decimal places.

- As reported by the U.S. National Center for Health Statistics, the mean serum high-density-lipoprotein (HDL) cholesterol for females age 20-29 years old is $\mu = 53$ mg/dL. If serum HDL cholesterol is normally distributed with $\sigma = 13.4$ mg/dL, what is the probability that a random sample of 20 females age 20-29 years old will have a mean serum cholesterol above 60 mg/dL?
- Finally, we determine the probability.

$$\begin{aligned}P(X > 60) &= P(z > 2.34) \\ &= P(z < -2.34) \\ &= 0.0096\end{aligned}$$

- Caution: Please remember not to round values used in intermediate calculations.

- Suppose cars arrive at Burger King's drive-through at a rate of 20 cars per hour between 12 PM and 1 PM. A random sample of 40 one-hour time periods between 12 PM and 1 PM is selected and has 22.1 as the mean number of cars arriving.
 - Why is the sampling distribution of \bar{X} approximately normal?

- Suppose cars arrive at Burger King's drive-through at a rate of 20 cars per hour between 12 PM and 1 PM. A random sample of 40 one-hour time periods between 12 PM and 1 PM is selected and has 22.1 as the mean number of cars arriving.
 - Why is the sampling distribution of \bar{X} approximately normal?
 - Hint: What does the Central Limit Theorem tell you?

- Suppose cars arrive at Burger King's drive-through at a rate of 20 cars per hour between 12 PM and 1 PM. A random sample of 40 one-hour time periods between 12 PM and 1 PM is selected and has 22.1 as the mean number of cars arriving.
 - What is the probability that a simple random sample of 40 one-hour time periods results in a mean of at least 22.1 cars?
Note: This is a Poisson Process - take $\mu = 20$ and $\sigma = \sqrt{20}$.
 - We determine the mean and the standard deviation for the sampling distribution.

$$\mu_{\bar{x}} = 20$$

$$\begin{aligned}\sigma_{\bar{x}} &= \frac{\sqrt{20}}{\sqrt{40}} \\ &= \frac{1}{\sqrt{2}}\end{aligned}$$

- Suppose cars arrive at Burger King's drive-through at a rate of 20 cars per hour between 12 PM and 1 PM. A random sample of 40 one-hour time periods between 12 PM and 1 PM is selected and has 22.1 as the mean number of cars arriving.

- What is the probability that a simple random sample of 40 one-hour time periods results in a mean of at least 22.1 cars?

Note: This is a Poisson Process - take $\mu = 20$ and $\sigma = \sqrt{20}$.

- We determine the z-score.

$$z = \frac{22.1 - 20}{\left(\frac{1}{\sqrt{2}} \right)} = 2.97$$

- Suppose cars arrive at Burger King's drive-through at a rate of 20 cars per hour between 12 PM and 1 PM. A random sample of 40 one-hour time periods between 12 PM and 1 PM is selected and has 22.1 as the mean number of cars arriving.
 - What is the probability that a simple random sample of 40 one-hour time periods results in a mean of at least 22.1 cars?
Note: This is a Poisson Process - take $\mu = 20$ and $\sigma = \sqrt{20}$.
 - Finally, we determine the probability.

$$\begin{aligned}P(X > 22.1) &= P(z > 2.97) \\ &= P(z < -2.97) \\ &= 0.0015\end{aligned}$$

- According to the Energy Information Administration, the mean age of a refrigerator in a home owned by the occupant is 8.5 years with a standard deviation of 6.6 years. A random sample of 50 homes owned by the occupant is conducted and the average age of the refrigerators is found to be 5.7 years. What is the probability that a random sample of 50 homes owned by the occupant results in a mean age of a refrigerator of 5.7 years or more?

- According to the Energy Information Administration, the mean age of a refrigerator in a home owned by the occupant is 8.5 years with a standard deviation of 6.6 years. A random sample of 50 homes owned by the occupant is conducted and the average age of the refrigerators is found to be 5.7 years. What is the probability that a random sample of 50 homes owned by the occupant results in a mean age of a refrigerator of 5.7 years or more?
- By the Central Limit Theorem, the sampling distribution of \bar{x} becomes approximately normal as the sample size increases.

- According to the Energy Information Administration, the mean age of a refrigerator in a home owned by the occupant is 8.5 years with a standard deviation of 6.6 years. A random sample of 50 homes owned by the occupant is conducted and the average age of the refrigerators is found to be 5.7 years. What is the probability that a random sample of 50 homes owned by the occupant results in a mean age of a refrigerator of 5.7 years or more?
- We determine the mean and standard deviation for the sampling distribution.

$$\mu_{\bar{x}} = 8.5 \quad \sigma_{\bar{x}} = \frac{6.6}{\sqrt{50}}$$

- According to the Energy Information Administration, the mean age of a refrigerator in a home owned by the occupant is 8.5 years with a standard deviation of 6.6 years. A random sample of 50 homes owned by the occupant is conducted and the average age of the refrigerators is found to be 5.7 years. What is the probability that a random sample of 50 homes owned by the occupant results in a mean age of a refrigerator of 5.7 years or more?
- We determine the z-score.

$$z = \frac{5.7 - 8.5}{\left(\frac{6.6}{\sqrt{50}} \right)} = -3$$

- According to the Energy Information Administration, the mean age of a refrigerator in a home owned by the occupant is 8.5 years with a standard deviation of 6.6 years. A random sample of 50 homes owned by the occupant is conducted and the average age of the refrigerators is found to be 5.7 years. What is the probability that a random sample of 50 homes owned by the occupant results in a mean age of a refrigerator of 5.7 years or more?
- Finally, we determine the probability.

$$\begin{aligned}P(X > 5.7) &= P(z > -3) \\ &= P(z < 3) \\ &= 0.9987\end{aligned}$$

- During the daytime in winter when nobody is at home, the mean household temperature is 67.6° F with a standard deviation of 4.2° F for households whose 1997 income was between \$10,000 and \$24,999 according to the Energy Information Administration. In a random sample of 50 households whose 1997 income was between \$10,000 and \$24,999 and for which nobody is home during the daytime, it was determined that the average temperature was 68.3° F. What is the probability that a random sample of 50 households whose 1997 income was between \$10,000 and \$24,999 and for which nobody is home during the daytime results in an average household temperature of more than 68.3° F?

- During the daytime in winter when nobody is at home, the mean household temperature is 67.6°F with a standard deviation of 4.2°F for households whose 1997 income was between \$10,000 and \$24,999 according to the Energy Information Administration. In a random sample of 50 households whose 1997 income was between \$10,000 and \$24,999 and for which nobody is home during the daytime, it was determined that the average temperature was 68.3°F . What is the probability that a random sample of 50 households whose 1997 income was between \$10,000 and \$24,999 and for which nobody is home during the daytime results in an average household temperature of more than 68.3°F ?
- By the Central Limit Theorem, the sampling distribution of \bar{x} becomes approximately normal as the sample size increases.

- During the daytime in winter when nobody is at home, the mean household temperature is 67.6° F with a standard deviation of 4.2° F for households whose 1997 income was between \$10,000 and \$24,999 according to the Energy Information Administration. In a random sample of 50 households whose 1997 income was between \$10,000 and \$24,999 and for which nobody is home during the daytime, it was determined that the average temperature was 68.3° F. What is the probability that a random sample of 50 households whose 1997 income was between \$10,000 and \$24,999 and for which nobody is home during the daytime results in an average household temperature of more than 68.3° F?
- We determine the mean and standard deviation for the sampling distribution.

$$\mu_{\bar{x}} = 67.6 \quad \sigma_{\bar{x}} = \frac{4.2}{\sqrt{50}}$$

- During the daytime in winter when nobody is at home, the mean household temperature is 67.6°F with a standard deviation of 4.2°F for households whose 1997 income was between \$10,000 and \$24,999 according to the Energy Information Administration. In a random sample of 50 households whose 1997 income was between \$10,000 and \$24,999 and for which nobody is home during the daytime, it was determined that the average temperature was 68.3°F . What is the probability that a random sample of 50 households whose 1997 income was between \$10,000 and \$24,999 and for which nobody is home during the daytime results in an average household temperature of more than 68.3°F ?
- We determine the z-score.

$$z = \frac{68.3 - 67.6}{\left(\frac{4.2}{\sqrt{50}}\right)} = 1.18$$

- During the daytime in winter when nobody is at home, the mean household temperature is 67.6°F with a standard deviation of 4.2°F for households whose 1997 income was between \$10,000 and \$24,999 according to the Energy Information Administration. In a random sample of 50 households whose 1997 income was between \$10,000 and \$24,999 and for which nobody is home during the daytime, it was determined that the average temperature was 68.3°F . What is the probability that a random sample of 50 households whose 1997 income was between \$10,000 and \$24,999 and for which nobody is home during the daytime results in an average household temperature of more than 68.3°F ?
- Finally, we determine the probability.

$$\begin{aligned}P(X > 68.3) &= P(z > 1.18) \\ &= P(z < -1.18) \\ &= 0.1190\end{aligned}$$

- The process for manufacturing a ball bearing results in weights that have an approximately normal distribution with mean 0.15 g and standard deviation 0.003 g.
 - If you select 10 ball bearings, what is the probability that their mean weight is less than 0.148 g?
 - We determine the mean and standard deviation for the sampling distribution.

$$\mu_{\bar{x}} = 0.15 \quad \sigma_{\bar{x}} = \frac{0.003}{\sqrt{10}}$$

- The process for manufacturing a ball bearing results in weights that have an approximately normal distribution with mean 0.15 g and standard deviation 0.003 g.
 - If you select 10 ball bearings, what is the probability that their mean weight is less than 0.148 g?
 - We determine the z-score.

$$z = \frac{0.148 - 0.15}{\left(\frac{0.003}{\sqrt{10}} \right)} = -2.11$$

- The process for manufacturing a ball bearing results in weights that have an approximately normal distribution with mean 0.15 g and standard deviation 0.003 g.
 - If you select 10 ball bearings, what is the probability that their mean weight is less than 0.148 g?
 - Finally, we determine the probability.

$$\begin{aligned}P(X < 0.148) &= P(z < -2.11) \\ &= 0.0174\end{aligned}$$

- The process for manufacturing a ball bearing results in weights that have an approximately normal distribution with mean 0.15 g and standard deviation 0.003 g.
 - How large should the sample size be so that the probability that the mean weight for the ball bearings is no more than 0.149 g is 0.0015?

- The process for manufacturing a ball bearing results in weights that have an approximately normal distribution with mean 0.15 g and standard deviation 0.003 g.
 - How large should the sample size be so that the probability that the mean weight for the ball bearings is no more than 0.149 g is 0.0015?
 - We are told that $P(X > 0.149) = 0.0015$.
 - The z-score corresponding to 0.0015 on the standard normal table is $z = 2.96$ or $z = 2.97$.

- In order to determine n , we begin with the z-score formula appropriately modified with the mean and standard deviation for the sampling distribution.

$$z = \frac{X - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

- Then, we make a change of form since

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

and obtain

$$z = \frac{X - \mu_{\bar{x}}}{\left(\frac{\sigma}{\sqrt{n}} \right)}$$

- Next, we isolate n.

$$z\left(\frac{\sigma}{\sqrt{n}}\right) = X - \mu_{\bar{x}}$$

$$\frac{\sigma}{\sqrt{n}} = \frac{X - \mu_{\bar{x}}}{z}$$

$$\frac{\sqrt{n}}{\sigma} = \frac{z}{X - \mu_{\bar{x}}}$$

$$\sqrt{n} = \frac{z\sigma}{X - \mu_{\bar{x}}}$$

- Finally, we square both sides to obtain our new formula for n,

$$n = \left(\frac{z\sigma}{X - \mu_{\bar{x}}}\right)^2$$

- The process for manufacturing a ball bearing results in weights that have an approximately normal distribution with mean 0.15 g and standard deviation 0.003 g.
 - How large should the sample size be so that the probability that the mean weight for the ball bearings is no more than 0.149 g is 0.0015?
 - Now, we use our new formula to determine the value of n.
$$\begin{aligned}n &= \left(\frac{2.96(0.003)}{0.149 - 0.15} \right)^2 \\ &= (-8.88)^2 \\ &= 78.8544 \\ &\approx 79\end{aligned}$$

- The process for manufacturing a ball bearing results in weights that have an approximately normal distribution with mean 0.15 g and standard deviation 0.003 g.
 - How large should the sample size be so that the probability that the mean weight for the ball bearings is no more than 0.149 g is 0.0015?
 - Now, we use our new formula to determine the value of n.
$$n = \left(\frac{2.97(0.003)}{0.149 - 0.15} \right)^2$$
$$= (-8.91)^2$$
$$= 79.3881$$
$$\approx 79$$

- The process for manufacturing a ball bearing results in weights that have an approximately normal distribution with mean 0.15 g and standard deviation 0.003 g.
 - How large should the sample size be so that the probability that the mean weight for the ball bearings is no more than 0.149 g is 0.0015?
 - So, for both possible z-scores, we find that a sample size of 79 would be sufficient.

- The process for manufacturing a ball bearing results in weights that have an approximately normal distribution with mean 0.15 g and standard deviation 0.003 g.
 - If you select 10 ball bearings, what is the probability that their mean weight is more than 0.149 g?

- The process for manufacturing a ball bearing results in weights that have an approximately normal distribution with mean 0.15 g and standard deviation 0.003 g.
 - If you select 10 ball bearings, what is the probability that their mean weight is more than 0.149 g?
 - We have

$$\mu_{\bar{x}} = 0.15$$
$$\sigma_{\bar{x}} = \frac{0.003}{\sqrt{10}}$$

$$z = \frac{0.149 - 0.15}{\left(\frac{0.003}{\sqrt{10}} \right)}$$
$$= -1.05$$

- The process for manufacturing a ball bearing results in weights that have an approximately normal distribution with mean 0.15 g and standard deviation 0.003 g.
 - If you select 10 ball bearings, what is the probability that their mean weight is more than 0.149 g?
 - Finally, determining the probability, we obtain

$$\begin{aligned}P(X > 0.149) &= P(z > -1.05) \\ &= P(z < 1.05) \\ &= 0.8531\end{aligned}$$

- The process for manufacturing a ball bearing results in weights that have an approximately normal distribution with mean 0.15 g and standard deviation 0.003 g.
 - How large should the sample size be so that the probability that the mean weight for the ball bearings is no more than 0.148 g is 0.0025?
 - We determine the z-score such that $P(X > 0.149) = 0.0015$ to be -2.81.

- The process for manufacturing a ball bearing results in weights that have an approximately normal distribution with mean 0.15 g and standard deviation 0.003 g.
 - How large should the sample size be so that the probability that the mean weight for the ball bearings is no more than 0.148 g is 0.0025?
 - We then determine the value of n.

$$\begin{aligned}
 n &= \left(\frac{-2.81(0.003)}{0.148 - 0.15} \right)^2 \\
 &= (4.215)^2 \\
 &= 17.766225 \\
 &\approx 18
 \end{aligned}$$

to find that a sample size of 18 is sufficient.

Sample Proportion

- Consider a sample of size n obtained from a population for which each element/individual either has or does not have a particular characteristic. The sample proportion, denoted p , is given by

$$p = \frac{x}{n}$$

where x is the number of elements/individuals in the sample that have the characteristic.

Sample Proportion

- The sample proportion, p , is a statistic that estimates the population proportion, P .

$$p = \frac{x}{n}$$

number in sample with characteristic
number in sample

- Please carefully note the use of the uppercase P and lowercase p .

Sample Proportion of P

- For a simple random sample of size n for which the sample size is less than 5% of the population size, that is, $n \leq 0.05N$,
 - the sampling distribution of P is approximately normal provided $np(1 - p) \geq 10$.

Sample Proportion of P

- For a simple random sample of size n for which the sample size is less than 5% of the population size, that is, $n \leq 0.05N$,
 - the mean of the sampling distribution of P is $\mu_p = p$.

Sample Proportion of P

- For a simple random sample of size n for which the sample size is less than 5% of the population size, that is, $n \leq 0.05N$,
 - the mean of the sampling distribution of P is $\mu_p = p$.

Sample Proportion of P

- For a simple random sample of size n for which the sample size is less than 5% of the population size, that is, $n \leq 0.05N$,
 - the standard deviation of the sampling distribution of P is

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}$$

Sample Proportion of P

- Having the sample size be less than 5% of the population size (that is, $n \leq 0.05N$) is necessary so that the result obtained from one individual is independent of the result obtained from the rest of the individuals.

Sample Proportion of P

- It is important to note that the mean of the sampling distribution for P is $\mu_p = p$ and the standard deviation of the sampling distribution of P is

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}$$

if $np(1 - p) \geq 10$ or $np(1 - p) \not\geq 10$

Sample Proportion of P

- With this information and the values of μ_p and σ_p , we determine probabilities and the sample size in a similar manner to that illustrated in the examples for the distribution for the sample mean.

- **Suppose a simple random sample of 100 households is obtained from a town with 5000 households. Suppose 30% of households in this town plant a garden in the spring.**
- **Describe the sampling distribution of P .**

- Suppose a simple random sample of 100 households is obtained from a town with 5000 households. Suppose 30% of households in this town plant a garden in the spring.
- Describe the sampling distribution of P .
- $N = 5000$ and $n = 100$
- $0.05N = 0.05(5000)$
= 250
> 100

- Suppose a simple random sample of 100 households is obtained from a town with 5000 households. Suppose 30% of households in this town plant a garden in the spring.
- Describe the sampling distribution of P .
- $N = 5000$ and $n = 100$
- $n \leq 0.05N$

- Suppose a simple random sample of 100 households is obtained from a town with 5000 households. Suppose 30% of households in this town plant a garden in the spring.
- Describe the sampling distribution of P .
- $N = 5000$ and $n = 100$
- $n \leq 0.05N$
 - The sample size is no more than 5% of the population size.

- Suppose a simple random sample of 100 households is obtained from a town with 5000 households. Suppose 30% of households in this town plant a garden in the spring.
- Describe the sampling distribution of P .
- $N = 5000$ and $n = 100$
- $p = 0.3$ and $1 - p = 0.7$

- Suppose a simple random sample of 100 households is obtained from a town with 5000 households. Suppose 30% of households in this town plant a garden in the spring.
- Describe the sampling distribution of P .
- $N = 5000$ and $n = 100$
- $p = 0.3$ and $1 - p = 0.7$
- $np(1 - p) = 5000(0.3)(0.7)$
 $= 21$
 > 10

- Suppose a simple random sample of 100 households is obtained from a town with 5000 households. Suppose 30% of households in this town plant a garden in the spring.
- Describe the sampling distribution of P .
- $N = 5000$ and $n = 100$
- $p = 0.3$ and $1 - p = 0.7$
- $np(1 - p) \geq 10$

- Suppose a simple random sample of 100 households is obtained from a town with 5000 households. Suppose 30% of households in this town plant a garden in the spring.
- Describe the sampling distribution of P .
- Since the sample size is no more than the population size ($n \leq 0.05N$) and since $np(1 - p) \geq 10$, the sampling distribution of \hat{p} is approximately normal.

- Suppose a simple random sample of 100 households is obtained from a town with 5000 households. Suppose 30% of households in this town plant a garden in the spring.
- What is the probability that more than 37 households in the sample plant a garden?

- $\mu_{\hat{p}} = p$
 $= 0.3$

$$\begin{aligned}\sigma_{\hat{p}} &= \sqrt{\frac{p(1-p)}{n}} \\ &= \sqrt{\frac{0.3(1-0.3)}{100}}\end{aligned}$$

- Suppose a simple random sample of 100 households is obtained from a town with 5000 households. Suppose 30% of households in this town plant a garden in the spring.
- What is the probability that more than 37 households in the sample plant a garden?

- $$\begin{aligned}\mu_{\hat{p}} &= p \\ &= 0.3\end{aligned}$$

$$\begin{aligned}\sigma_{\hat{p}} &= \sqrt{\frac{0.3(0.7)}{100}} \\ &= \sqrt{\frac{0.21}{100}} \\ &= \sqrt{0.0021}\end{aligned}$$

- Suppose a simple random sample of 100 households is obtained from a town with 5000 households. Suppose 30% of households in this town plant a garden in the spring.
- What is the probability that more than 37 households in the sample plant a garden?

$$\begin{aligned}P\left(\hat{p} > \frac{37}{100}\right) &= P\left(z > \frac{0.37 - 0.3}{\sqrt{0.0021}}\right) \\&= P(z > 1.527525232) \\&= P(z > 1.53) \\&= P(z < -1.53) \\&= 0.0630\end{aligned}$$

- Suppose a simple random sample of 100 households is obtained from a town with 5000 households. Suppose 30% of households in this town plant a garden in the spring.
- What is the probability that 18 or fewer households in the sample plant a garden?

$$\begin{aligned}P\left(\hat{p} \leq \frac{18}{100}\right) &= P\left(z \leq \frac{0.18 - 0.3}{\sqrt{0.0021}}\right) \\ &= P(z \leq -2.618614683) \\ &= P(z \leq -2.62) \\ &= 0.0044\end{aligned}$$

