

Lecture 19: Chapter 8, Section 1

Sampling Distributions: Proportions

- Typical Inference Problem
- Definition of Sampling Distribution
- 3 Approaches to Understanding Sampling Dist.
- Applying 68-95-99.7 Rule



Looking Back: *Review*

□ 4 Stages of Statistics

- Data Production (discussed in Lectures 1-4)
- Displaying and Summarizing (Lectures 5-12)
- Probability
 - Finding Probabilities (discussed in Lectures 13-14)
 - Random Variables (discussed in Lectures 15-18)
 - Sampling Distributions
 - Proportions
 - Means
- Statistical Inference



Typical Inference Problem

If sample of 100 students has 0.13 left-handed, can you believe population proportion is 0.10?

Solution Method: Assume (temporarily) that population proportion is 0.10, find **probability of sample proportion** as high as 0.13. If it's too improbable, we won't believe population proportion is 0.10.



Key to Solving Inference Problems

For a given population proportion p and sample size n , need to find **probability** of sample proportion \hat{p} in a certain range:

Need to know **sampling distribution** of \hat{p} .

Note: \hat{p} can denote a single statistic or a random variable.



Definition

Sampling distribution of sample statistic tells **probability distribution** of values taken by the statistic in repeated random samples of a given size.

Looking Back:** We summarize a probability distribution by reporting its **center, spread, shape.

Behavior of Sample Proportion (*Review*)

For random sample of size n from population with p in category of interest, sample proportion $\hat{p} = \frac{X}{n}$ has


- mean p
- standard deviation $\sqrt{\frac{p(1-p)}{n}}$
- shape approximately normal for large enough n

Looking Back: Can find normal probabilities using 68-95-99.7 Rule, etc.



Rules of Thumb (*Review*)

- Population at least 10 times sample size n
(formula for standard deviation of \hat{p}
approximately correct even if sampled
without replacement)
- np and $n(1-p)$ both at least 10
(guarantees \hat{p} approximately normal)



Understanding Dist. of Sample Proportion

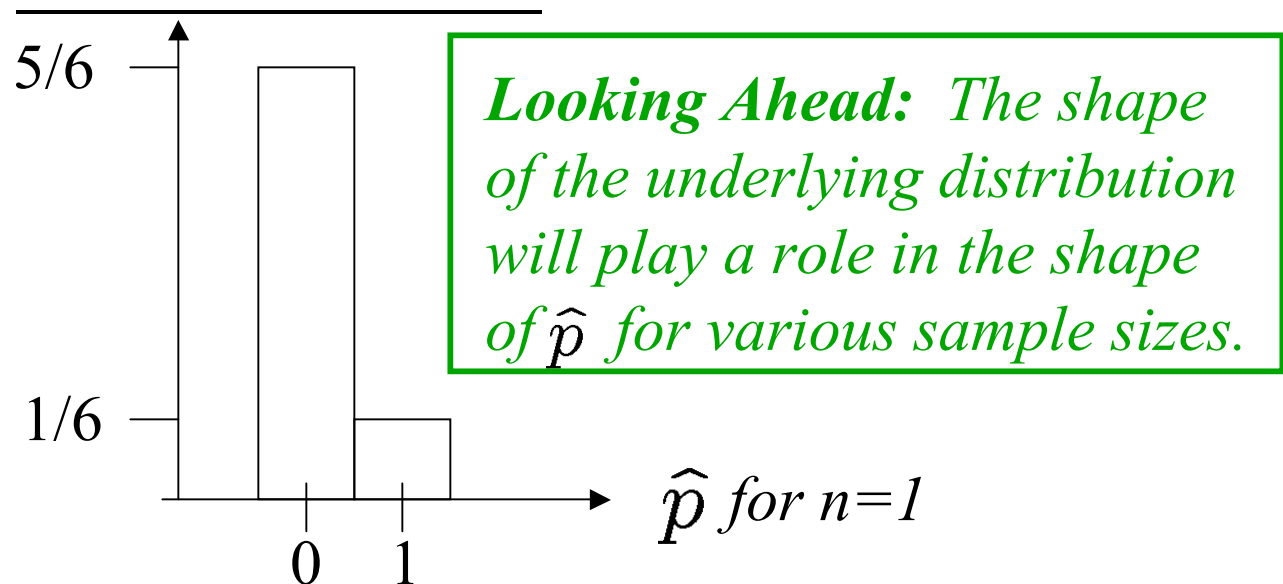
3 Approaches:

1. Intuition
2. Hands-on Experimentation
3. Theoretical Results

*Looking Ahead: We'll find that our **intuition** is consistent with **experimental** results, and both are confirmed by mathematical **theory**.*

Example: Shape of Underlying Distribution ($n=1$)

- **Background:** Population proportion of blue M&M's is $p=1/6=0.17$.
- **Question:** How does the probability histogram for sample proportions appear for samples of size 1?
- **Response:**





Example: *Sample Proportion as Random Variable*

- **Background:** Population proportion of blue M&Ms is 0.17.
- **Questions:**
 - Is the underlying variable categorical or quantitative?
 - Consider the behavior of sample proportion \hat{p} for repeated random samples of a given size. What type of variable is sample proportion?
 - What 3 aspects of the distribution of sample proportion should we report to summarize its behavior?
- **Responses:**
 - Underlying variable _____
 - _____
 - Summarize with _____, _____, _____

Example: *Center, Spread of Sample Proportion*

- **Background:** Population proportion of blue M&M's is $p=1/6=0.17$.
- **Question:** What can we say about **center and spread** of \hat{p} for repeated random samples of size $n = 25$ (a teaspoon)?
- **Response:**
 - **Center:** Some \hat{p} 's more than ____, others less; should balance out so mean of \hat{p} 's is $p =$ _____.
 - **Spread** of \hat{p} 's: s.d. depends on ____.
 - For $n=6$, could easily get \hat{p} anywhere from ____ to ____.
 - For $n=25$, spread of \hat{p} will be ____ than it is for $n = 6$.



Example: *Intuit Shape of Sample Proportion*

- **Background:** Population proportion of blue M&M's is $p=1/6=0.17$.
 - **Question:** What can we say about the **shape** of \hat{p} for repeated random samples of size $n = 25$ (a teaspoon)?
 - **Response:**
 \hat{p} close to _____ most common, far from _____ in either direction increasingly less likely →
-

Example: *Sample Proportion for Larger n*

- **Background:** Population proportion of blue M&M's is $p=1/6=0.17$.
- **Question:** What can we say about center, spread, shape of \hat{p} for repeated random samples of size $n = 75$ (a Tablespoon)?
- **Response:**
 - **Center:** mean of \hat{p} 's should be $p = \underline{\hspace{2cm}}$ (for any n).
 - **Spread** of \hat{p} 's: compared to $n=25$, spread for $n=75$ is $\underline{\hspace{2cm}}$
 - **Shape:** \hat{p} 's clumped near 0.17, taper at tails $\rightarrow \underline{\hspace{2cm}}$

*Looking Ahead: Sample size does **not** affect center but plays an important role in spread and shape of the distribution of sample proportion (also of sample mean).*



Understanding Sample Proportion

3 Approaches:

1. Intuition
2. Hands-on Experimentation
3. Theoretical Results

*Looking Ahead: We'll find that our intuition is consistent with **experimental** results, and both are confirmed by mathematical **theory**.*



Central Limit Theorem

Approximate **normality** of sample statistic for repeated random samples of a large enough size is cornerstone of inference theory.

- Makes intuitive sense.
- Can be verified with experimentation.
- Proof requires higher-level mathematics; result called **Central Limit Theorem**.



Center of Sample Proportion (*Implications*)

For random sample of size n from population with p in category of interest, sample proportion $\hat{p} = \frac{X}{n}$ has

- **mean** p

→ \hat{p} is *unbiased estimator* of p

(*sample must be random*)

Spread of Sample Proportion (*Implications*)

For random sample of size n from **population** with p in category of interest, sample proportion $\hat{p} = \frac{X}{n}$ has

- mean p

- standard deviation $\sqrt{\frac{p(1-p)}{n}}$ ← n in denominator

→ \hat{p} has less spread for larger samples

(population size must be at least $10n$)

Shape of Sample Proportion (*Implications*)

For random sample of size n from population with p in category of interest, sample proportion $\hat{p} = \frac{X}{n}$ has

- mean p
- standard deviation $\sqrt{\frac{p(1-p)}{n}}$
- shape approx. normal for large enough n
→ can find **probability** that sample proportion takes value in given interval



Example: *Behavior of Sample Proportion*

- **Background:** Population proportion of blue M&M's is $p=0.17$.
 - **Question:** For repeated random samples of $n=25$, how does \hat{p} behave?
 - **Response:** For $n=25$, \hat{p} has
 - **Center:** mean _____
 - **Spread:** standard deviation _____
 - **Shape:** not really normal because _____
-

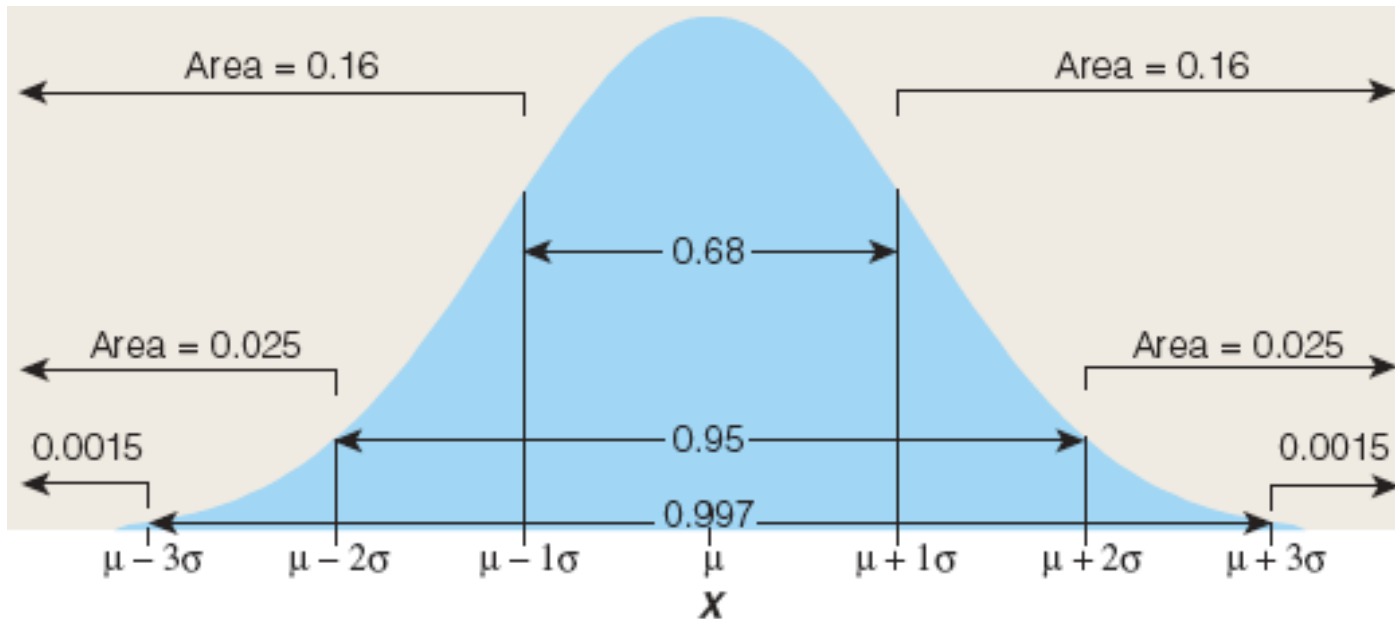
Example: *Sample Proportion for Larger n*

- **Background:** Population proportion of blue M&M's is $p=0.17$.
 - **Question:** For repeated random samples of $n=75$, how does \hat{p} behave?
 - **Response:** For $n=75$, \hat{p} has
 - **Center:** mean _____
 - **Spread:** standard deviation _____
 - **Shape:** approximately normal because _____
-

68-95-99.7 Rule for Normal R.V. (*Review*)

Sample at random from normal population; for sampled value X (a R.V.), probability is

- 68% that X is within 1 standard deviation of mean
- 95% that X is within 2 standard deviations of mean
- 99.7% that X is within 3 standard deviations of mean



68-95-99.7 Rule for Sample Proportion

For sample proportions \hat{p} taken at random from a large population with underlying p , probability is

- 68% that \hat{p} is within $1 \sqrt{\frac{p(1-p)}{n}}$ of p
- 95% that \hat{p} is within $2 \sqrt{\frac{p(1-p)}{n}}$ of p
- 99.7% that \hat{p} is within $3 \sqrt{\frac{p(1-p)}{n}}$ of p

Example: Sample Proportion for $n=75$, $p=0.17$

- **Background:** Population proportion of blue M&Ms is $p=0.17$. For random samples of $n=75$, \hat{p} approx. normal with mean 0.17, s.d. $\sqrt{\frac{0.17(1-0.17)}{75}} = 0.043$

- **Question:**

What does 68-95-99.7 Rule tell us about behavior of \hat{p} ?

- **Response:** The probability is approximately
 - 0.68 that \hat{p} is within _____ of _____: in (0.13, 0.21)
 - 0.95 that \hat{p} is within _____ of _____: in (0.08, 0.26)
 - 0.997 that \hat{p} is within _____ of _____: in (0.04, 0.30)

Looking Back: We don't use the Rule for $n=25$ because _____



90-95-98-99 Rule (*Review*)

For standard normal Z , the probability is

- 0.90 that Z takes a value in interval $(-1.645, +1.645)$
- 0.95 that Z takes a value in interval $(-1.960, +1.960)$
- 0.98 that Z takes a value in interval $(-2.326, +2.326)$
- 0.99 that Z takes a value in interval $(-2.576, +2.576)$

Example: Sample Proportion for $n=75$, $p=0.17$

- **Background:** Population proportion of blue M&Ms is $p=0.17$. For random samples of $n=75$, \hat{p} approx. normal with mean 0.17, s.d. $\sqrt{\frac{0.17(1-0.17)}{75}} = 0.043$
- **Question:**

What does 90-95-98-99 Rule tell about behavior of \hat{p} ?

- **Response:** The probability is approximately
 - 0.90 that \hat{p} is within _____ (0.043) of 0.17: in (0.10,0.24)
 - 0.95 that \hat{p} is within _____ (0.043) of 0.17: in (0.09,0.25)
 - 0.98 that \hat{p} is within _____ (0.043) of 0.17: in (0.07,0.27)
 - 0.99 that \hat{p} is within _____ (0.043) of 0.17: in (0.06,0.28)



Typical Inference Problem (*Review*)

If sample of 100 students has 0.13 left-handed, can you believe population proportion is 0.10?

Solution Method: Assume (temporarily) that population proportion is 0.10, find **probability of sample proportion** as high as 0.13. If it's too improbable, we won't believe population proportion is 0.10.

Example: *Testing Assumption About p*

- **Background:** We asked, “*If sample of 100 students has 0.13 left-handed, can you believe population proportion is 0.10?*”
- **Questions:**
 - What are the mean, standard deviation, and shape of \hat{p} ?
 - Is 0.13 improbably high under the circumstances?
 - Can we believe $p = 0.10$?
- **Response:**
 - For $p=0.10$ and $n=100$, \hat{p} has mean _____, s.d. _____ ; shape approx. normal since _____.
 - According to Rule, the probability is _____ that \hat{p} would take a value of 0.13 (1 s.d. above mean) or more.
 - Since this isn’t so improbable, _____.



Lecture Summary

(Distribution of Sample Proportion)

- Typical inference problem
- Sampling distribution; definition
- 3 approaches to understanding sampling dist.
 - Intuition
 - Hands-on experiment
 - Theory
- Center, spread, shape of sampling distribution
 - Central Limit Theorem
- Role of sample size
- Applying 68-95-99.7 Rule