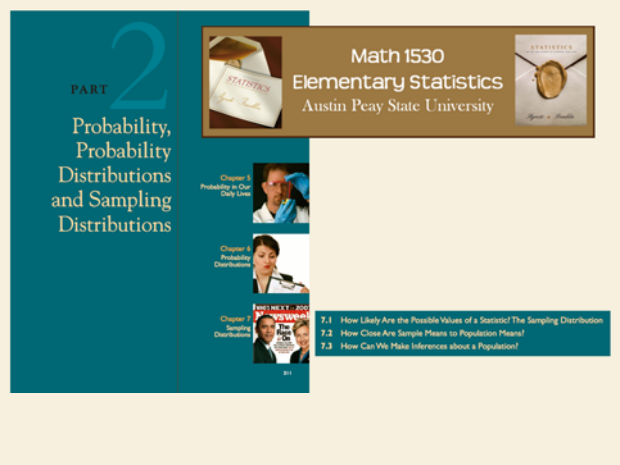
SAMPLING DISTRIBUTIONS

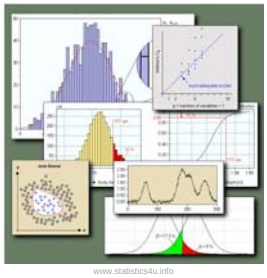
Chapter 7



7.1 How Likely Are the Possible Values of a Statistic? The Sampling Distribution

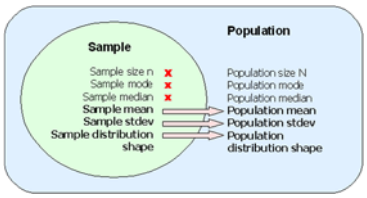
Learning Objectives

1. Statistic vs. Parameter
2. Sampling Distributions
3. μ and σ of Sampling Distribution of Proportion
4. Standard Error
5. Sampling Distribution Example
6. Population, Data, and Sampling Distributions



www.statistics4u.info

Statistic and Parameter



comfsm.fm

- **Statistic** – **Sample** summary: p-hat or xbar
- **Parameter** – **Population** summary: μ or σ
- Seldom know parameters, IRL
- Statistics estimate parameters

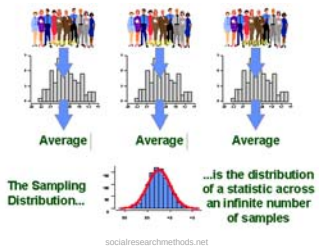
Sampling Distributions: Gray Davis

Vote to recall Governor:

- 3160 voters sampled in exit polls with 0.54, or 54%, in favor of recall
- Different random samples have *different p-hats*
- *Sampling distribution* shows all possible p-hats for a set n

humanflowerproject.com

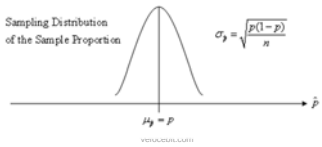
Sampling Distributions



- *Probability distributions* giving probabilities for possible statistic values
- Show variability from study to study
- Help predict how close statistic falls to the parameter it estimates

Mean and SD of Sampling Distribution for Proportion

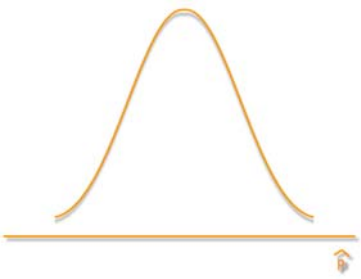
For random sample of size *n* from a population with proportion *p*, the sampling distribution of the sample proportion has:



$$\text{Mean} = p$$

$$\text{standard deviation} = \sqrt{\frac{p(1-p)}{n}}$$

The Standard Error



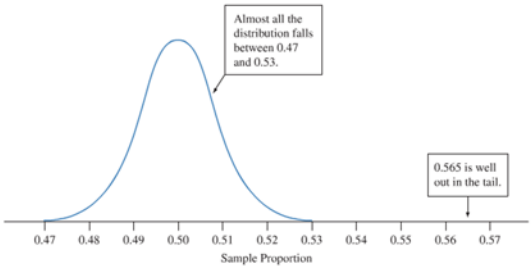
To distinguish standard deviation of a *sampling distribution* from standard deviation of ordinary probability distribution, we refer to it as a *standard error*

2006 California Election

- If p supporting reelection of Schwarzenegger was 0.50, would it have been unlikely to observe the exit-poll sample proportion of 0.565 with 2705 sampled?
- Would you be willing to predict that Schwarzenegger would win the election?

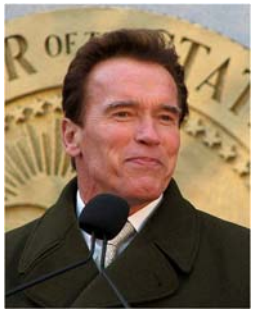
2006 California Election

2705 sampled and assuming 50% population support, estimate population proportion and standard error:



2006 California Election

What about the likelihood of having a sample proportion of 0.565?

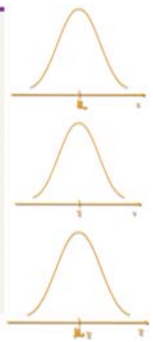


More than six standard errors from *expected value* of 0.50

Population, Data, and Sampling Distributions

Population Distribution, Data Distribution, Sampling Distribution

- **Population distribution:** This is the probability distribution from which we take the sample. Values of its *parameters*, such as the population proportion p for a categorical variable and the population mean μ for a quantitative variable, are usually unknown. They're what we'd like to learn about.
- **Data distribution:** This is the distribution of the sample data and is the distribution we actually see in practice. It's described by *statistics*, such as a sample proportion or a sample mean. With random sampling, the larger the sample size n, the more closely the data distribution resembles the population distribution.
- **Sampling distribution:** This is the probability distribution of a sample statistic, such as a sample proportion or a sample mean. With random sampling, the sampling distribution provides probabilities for all the possible values of the statistic. The sampling distribution provides the key for telling us how close a sample statistic falls to the corresponding unknown parameter. Its standard deviation is called the *standard error*.



Clinton vs. Spencer: 2006 Senatorial Seat

- Exit poll of 1336
  - 67% (895) Clinton
  - 33% (441) Spencer
- When 4.1 million votes tallied
  - 68% Clinton
  - 32% Spencer

Let X= vote outcome with x=1 for Clinton and x=0 for Spencer

nymag.com

Clinton vs. Spencer: Senatorial Seat

rbvergara.files.wordpress.com

- Population distribution: 4.1 million, 32% are 0, and 68% are 1.
- Data distribution: 1336, 33% are 0, and 67% are 1.
- Sampling distribution of sample proportion: approximately normal with p=0.68 and

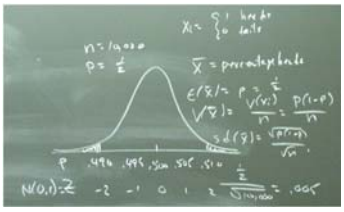
$$\sigma = \sqrt{0.68(1-0.68)/1336} = 0.013$$

- Sampling distribution bell-shaped; others discrete with values 0 and 1

7.2 How Close Are Sample Means to Population Means?

Learning Objectives

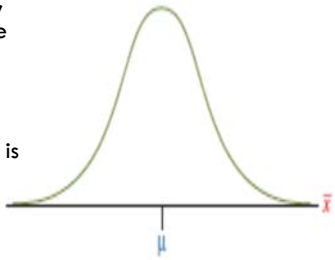
1. Sampling Distribution of Sample Mean
2. Effect of n on Standard Error
3. Central Limit Theorem (CLT)
4. Calculating Probabilities of Sample Means



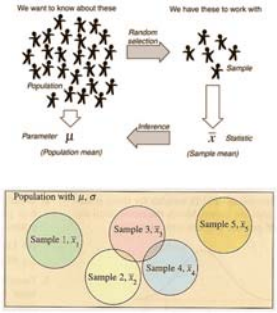
econ-courses.com

Sampling Distribution of Sample Mean

The sample mean, $\bar{x}$, is a random variable that varies from sample to sample, whereas the population mean, μ, is fixed.



Sampling Distribution of Sample Mean



rchsbowman.files.wordpress.com

- Sampling distribution of sample mean for random samples of size *n* from a population with mean μ and standard deviation σ, has:
  - Center and mean is same mean, μ
  - Spread is standard error of $\bar{x} = \sigma/\sqrt{n}$

cdn.information-management.com

Pizza Sales

Daily sales at a vary around a mean, μ = $900, with a standard deviation of σ = $300.

- What are the center and spread of the sampling distribution?

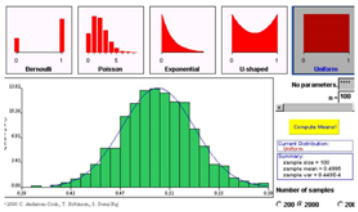
ilfondaccio.com

Effect of *n* on the Standard Error

- The standard error of the sample mean = $\frac{\sigma}{\sqrt{n}}$
- As *n* increases, denominator increases, so s.e. decreases
- With larger samples, the sample mean is more likely to be close to the population mean

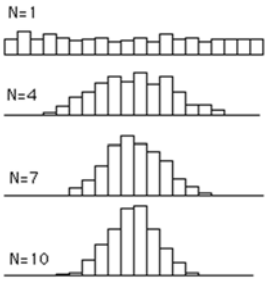
Central Limit Theorem

How does the sampling distribution of the sample mean relate with respect to shape, center, and spread to the probability distribution from which the samples were taken?



serc.carleton.edu

Central Limit Theorem (CLT)

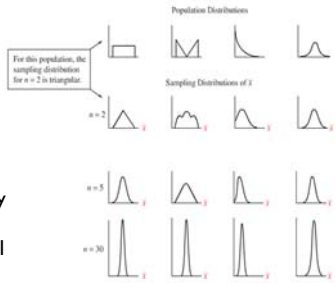


davidmlane.com

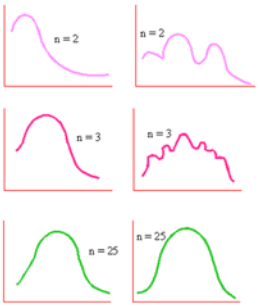
For random sampling with a large sample size *n*, sampling distribution of sample mean is approximately normal, *no matter what the shape* of the original probability distribution

Sampling Distribution of Sample Means

- More bell-shaped as *n* increases
- The more skewed, the larger *n* must be to get close to normal
- Usually close to normal when *n* is 30
- Always approximately normal for approximately normal populations



CLT: Making Inferences



pinkmonkey.com

- For large *n*, sampling distribution is approximately normal even if population distribution is not
- Enables inferences about population means regardless of shape of population distribution

Calculating Probabilities of Sample Means

- Distribution of milk bottle weights is normally distributed with a mean of 1.1 lbs and σ = 0.20
- What is the probability that the mean of a random sample of 5 bottles will be greater than 0.99 lbs?

1.bp.blogspot.com

Calculating Probabilities of Sample Means

stockmarketinvesting.com

- Closing prices of stocks have a right skewed distribution with a mean (μ) of $25 and σ= $20.
- What is the probability that the mean of a random sample of 40 stocks will be less than $20?

Calculating Probabilities of Sample Means

Repair claims have a mean of $920 and a standard deviation of $870.

- What is the probability that the average of the 100 claims is larger than $900?

carinsurancecomparison.com

Calculating Probabilities of Sample Means

displayfakefoods.com

Distribution of actual weights of 8 oz. wedges of cheddar cheese is normal with mean =8.1 oz and standard deviation of 0.1 oz

- Find *x* such that there is only a 10% chance that the average weight of a sample of five wedges will be above *x*

Calculating Probabilities of Sample Means

Distribution of 8 oz. wedges have mean = 8.1 oz. and standard deviation = 0.1 oz.

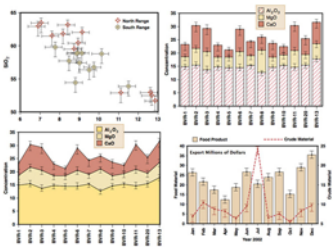
- Find *x* such that there is only a 5% chance the average weight of a sample of five wedges will be below *x*

pongcheese.co.uk

7.3 How Can We Make Inferences About a Population?

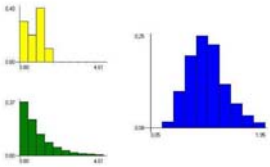
Learning Objectives

1. Using the CLT to Make Inferences
2. Standard Errors in Practice
3. Sampling Distribution for a Proportion



gigawiz.com

Using the CLT to Make Inferences



lehigh.edu

Implications of the CLT:

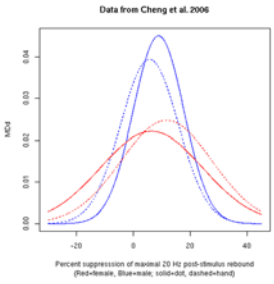
1. For large *n*, sampling distribution of $\bar{x}$ is approximately normal despite population shape
2. When approximately normal, $\bar{x}$ is within 2 standard errors of μ 95% of the time and almost certainly within 3

Standard Errors in Practice

Standard error have exact values depend on parameters:

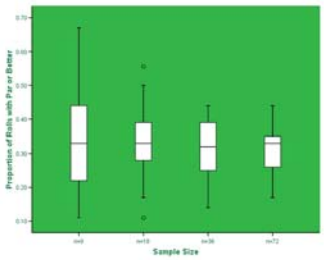
- $\sqrt{p(1-p)/n}$
- $\sigma/\sqrt{n}$

In practice, parameters are unknown so we approximate with p-hat and s



languagelog.ldc.upenn.edu

Sampling Distribution for a Proportion



amstat.org

- Binomial probability distribution with x as # of successes and y as probability
- Sample *proportion* (not #) of successes is usually reported

Sampling Distribution for a Proportion

Mean and Standard Deviation of Sampling Distribution of a Proportion

For a binomial random variable with *n* trials and probability *p* of success for each, the sampling distribution of the *proportion* of successes has

$$\text{Mean} = p \quad \text{standard error} = \sqrt{\frac{p(1-p)}{n}}.$$

To obtain these values, take the mean *np* and standard deviation $\sqrt{np(1-p)}$ for the binomial distribution of the *number* of successes and divide by *n*.

Image Sources

Statistics: The Art and Science of Learning from Data, 2nd Edition, Agresti and Franklin
The Drunkard's Walk, Mlodinow
Elementary Statistics, 7th Edition, Neil Weiss

http://cdn.information-management.com/media/editorial/dmdirect/0603/062403_dutta_2.gif
http://weblog.terrellrussell.com/wp-content/uploads/2007/10/800px-normal_distribution_pdf.png
http://www.besthealthcarergtes.com/Portals/41042/images//J042304 6.jpg
http://www.views.pk/wp-content/uploads/2007/06/blood-pressure.jpg http://media.rd.com/rd/images/rdc/books/change-one/blood-pressure-tests-of.jpg
http://www.healthinformer.net/wp-content/uploads/2008/07/blood-pressure.jpg
https://sol.mnstate.edu/virtual/hrm/DFC/images/MPJ04003600000%581%5D.jpg
http://www.betterreportcard.com/img/SAT-ACT.jpg
http://www.mschip.com/wp-content/uploads/2009/12/affordable-health-insurance2.jpg
http://3.bp.blogspot.com/_CjXtoDbokPE/R980587M4KI/AAAAAAAAAYM/udwbdWYLew/s400/binary+nails.bmp
http://sheerchaos.files.wordpress.com/2010/03/1-bp-blogspot-com.jpg
http://www.fallingfifth.com/files/comics/factorial.png
http://www.phantasmpsiresearch.com/JPGS/Rhine%20Institute%20ESP%20testing.jpg
http://www.psychicsavers.com/wp-content/uploads/2009/02/devin-1-main_full.jpg
http://www.columbusyouthexpressions.org/teamwork/images/stories/failure-success.jpg
http://3.bp.blogspot.com/_M8chV5JZMBc/RsxjjG6QrLI/AAAAAAAAABQ/NRdAb72JU08/s320/profilisgart2_large.jpg
http://www.zerocustoms.com/news/images/Dodge/2009-Dodge-Charger-police-car-1.jpg
http://3.bp.blogspot.com/_30PRmkOI4ro/SFGXTA4HePI/AAAAAAAAACdk/E6r4MjvTmtJc/s400/african-american-hairstyles1.jpg
http://zoonek2.free.fr/UNIX/48_R/g564.png
http://www.physics.umd.edu/grt/taj/104a/lightning2.jpg
http://hoppingintopuddles.files.wordpress.com/2008/03/random-numbers.jpg
http://www.economicpopulist.org/files/images/roulette_wheel.jpg
http://www.anticsland.com/images/games/large/billiards.jpg
http://psy2.ucsd.edu/~kang/child%20pictures/children-jump.jpg
http://www.vat19.com/webimages/covers/large/slot-machine-paperweight.jpg
http://rchsbowman.files.wordpress.com/2008/11/graph-of-discrete-probability-distribution1.jpg
http://cnfolio.com/images/img536expectedvalue.gif
http://www.math.armstrong.edu/statsonline/3/emprule.gif
https://www.glastonbury.org/staff/Mohart/PublishingImages/Ti-84-graphing-calculator.jpg