Economics 103 – Statistics for Economists

Francis J. DiTraglia

University of Pennsylvania

Lecture #1 – Introduction

Overview – Population vs. Sample, Probability vs. Statistics

Polling – Sampling vs. Non-sampling Error, Random Sampling

Causality – Observational vs. Experimental Data, RCTs

Racial Discrimination in the Labor Market

Source: Bureau of Labor Statistics

	Oct. 2017	Nov. 2017	Dec. 2017
White:	3.2	3.2	3.4
Black/African American:	7.4	7.3	6.6

Table: Unemployment rate in percentage points for men aged 20 and over in the last quarter of 2017.

The unemployment rate for African Americans has historically been much higher than for whites. What can this information by itself tell us about racial discrimination in the labor market?

This Course: Use Sample to Learn About Population

Population

Complete set of all items that interest investigator

Sample

Observed subset, or portion, of a population

Sample Size

of items in the sample, typically denoted n

Examples...

In Particular: Use Statistic to Learn about Parameter

Parameter

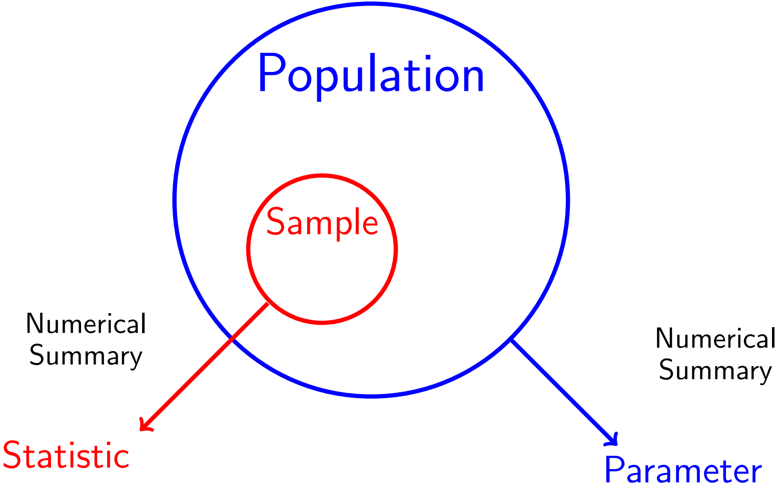
Numerical measure that describes specific characteristic of a population.

Statistic

Numerical measure that describes specific characteristic of sample.

Examples...

Essential Distinction You Must Remember!



This Course

1. Descriptive Statistics: summarize data
   - Summary Statistics
   - Graphics
2. Probability: Population $\rightarrow$ Sample
   - deductive: "safe" argument
     - All ravens are black. Mordecai is a raven, so Mordecai is black.
3. Inferential Statistics: Sample $\rightarrow$ Population
   - inductive: "risky" argument
     - I've only every seen black ravens, so all ravens must be black.

Sampling and Nonsampling Error

In statistics we use samples to learn about populations, but samples almost never be *exactly* like the population they are drawn from.

1. Sampling Error
   - *Random* differences between sample and population
   - Cancel out on average
   - Decreases as sample size grows
2. Nonsampling Error
   - *Systematic* differences between sample and population
   - Does *not* cancel out on average
   - Does *not* decrease as sample size grows


NEW COLORED MAP OF POLAND IN THIS ISSUE
Showing the Territorial Changes Wrought by the War
The Literary Digest
(Title Reg. U.S. Pat. Off.)
New York FUNK & WAGNALLS COMPANY London
PUBLIC OPINION New York combined with The LITERARY DIGEST
Vol. 68, No. 8. Whole No. 1609
FEBRUARY 19, 1921

Literary Digest – 1936 Presidential Election Poll

FDR versus Kansas Gov. Alf Landon

Huge Sample

Sent out over 10 million ballots; 2.4 million replies! (Compared to less than 45 million votes cast in actual election)

Prediction

Landslide for Landon: *Landonslide*, if you will.

Spectacularly Mistaken!

FDR versus Kansas Gov. Alf Landon

	Roosevelt	Landon
Literary Digest Prediction:	41%	57%
Actual Result:	61%	37%

What Went Wrong? *Non-sampling Error (aka Bias)*

Source: Squire (1988)

Biased Sample

Some units more likely to be sampled than others.

- Ballots mailed those on auto reg. list and in phone books.

Non-response Bias

Even if sample is unbiased, can't force people to reply.

- Among those who recieved a ballot, Landon supporters were more likely to reply.

In this case, neither effect *alone* was enough to throw off the result but together they did.

Randomize to Get an Unbiased Sample

Simple Random Sample

Each member of population is chosen strictly by chance, so that: (1) selection of one individual doesn't influence selection of any other, (2) each individual is just as likely to be chosen, (3) every possible sample of size n has the same chance of selection.

What about non-response bias? – we'll come back to this...

“Negative Views of Trump’s Transition”

Source: Pew Research Center

Ahead of Donald Trump’s scheduled press conference in New York City on Wednesday, the public continues to give the president-elect low marks for how he is handling the transition process. . . The latest national survey by Pew Research Center, conducted Jan. 4-9 among 1,502 adults, finds that 39% approve of the job President-elect Trump has done so far explaining his policies and plans for the future to the American people, while a larger share (55%) say they disapprove.

Quantifying Sampling Error

95% Confidence Interval for Poll Based on Random Sample

Margin of Error a.k.a. ME

We report $P \pm \text{ME}$ where $\text{ME} \approx 2\sqrt{P(1-P)/n}$

Quantifying Sampling Error

95% Confidence Interval for Poll Based on Random Sample

Margin of Error a.k.a. ME

We report $P \pm \text{ME}$ where $\text{ME} \approx 2\sqrt{P(1-P)/n}$

Trump Transition Approval Rate

$P = 0.39$ and $n = 1502$ so $\text{ME} \approx 0.013$. We'd report 39% plus or minus 1.3% if the poll were based on a simple random sample...

Quantifying Sampling Error

95% Confidence Interval for Poll Based on Random Sample

Margin of Error a.k.a. ME

We report $P \pm \text{ME}$ where $\text{ME} \approx 2\sqrt{P(1-P)/n}$

Trump Transition Approval Rate

$P = 0.39$ and $n = 1502$ so $\text{ME} \approx 0.013$. We'd report 39% plus or minus 1.3% if the poll were based on a simple random sample...

But Pew Reports an ME of 2.9% – more than twice as large as the one we calculated! What's going on here?!

Non-response bias is a huge problem...

Source: Pew Research Center

Surveys Face Growing Difficulty Reaching, Persuading Potential Respondents

	1997	2000	2003	2006	2009	2012
	%	%	%	%	%	%
Contact rate (percent of households in which an adult was reached)	90	77	79	73	72	62
Cooperation rate (percent of households contacted that yielded an interview)	43	40	34	31	21	14
Response rate (percent of households sampled that yielded an interview)	36	28	25	21	15	9

PEW RESEARCH CENTER 2012 Methodology Study. Rates computed according to American Association for Public Opinion Research (AAPOR) standard definitions for CON2, COOP3 and RR3. Rates are typical for surveys conducted in each year.

Methodology – "Negative Views of Trump's Transition"

Source: Pew Research Center

The combined landline and cell phone sample are weighted using an iterative technique that matches gender, age, education, race, Hispanic origin and nativity and region to parameters from the 2015 Census Bureaus American Community Survey and population density to parameters from the Decennial Census. The sample also is weighted to match current patterns of telephone status (landline only, cell phone only, or both landline and cell phone), based on extrapolations from the 2016 National Health Interview Survey. The weighting procedure also accounts for the fact that respondents with both landline and cell phones have a greater probability of being included in the combined sample and adjusts for household size among respondents with a landline phone. The margins of error reported and statistical tests of significance are adjusted to account for the surveys design effect, a measure of how much efficiency is lost from the weighting procedures.

Simple Example of Weighting a Survey

Post-stratification

- Women make up 49.6% of the population but suppose they are less likely to respond to your survey than men.
- If women have different opinions of Trump, this will skew the survey.
- Calculate Trump approval rate separately for men P_M vs. women P_W.
- Report $0.496 \times P_W + 0.504 \times P_M$, not the raw approval rate P.

Simple Example of Weighting a Survey

Post-stratification

- Women make up 49.6% of the population but suppose they are less likely to respond to your survey than men.
- If women have different opinions of Trump, this will skew the survey.
- Calculate Trump approval rate separately for men P_M vs. women P_W.
- Report $0.496 \times P_W + 0.504 \times P_M$, not the raw approval rate P.

Caveats

- Post-stratification isn't a magic bullet: you have to figure out what factors could skew your poll to adjust for them.
- Calculating the ME is more complicated. Since this is an intro class we'll focus on simple random samples.

Survey to find effect of Polio Vaccine

Ask random sample of parents if they vaccinated their kids or not and if the kids later developed polio. Compare those who were vaccinated to those who weren't.

Would this procedure:

(a) Overstate effectiveness of vaccine

(b) Correctly identify effectiveness of vaccine

(c) Understate effectiveness of vaccine

Confounding

Parents who vaccinate their kids may differ systematically from those who don't in *other ways* that impact child's chance of contracting polio!

Wealth is related to vaccination *and* whether child grows up in a hygenic environment.

Confounder

Factor that influences both outcomes and whether subjects are treated or not. Masks true effect of treatment.

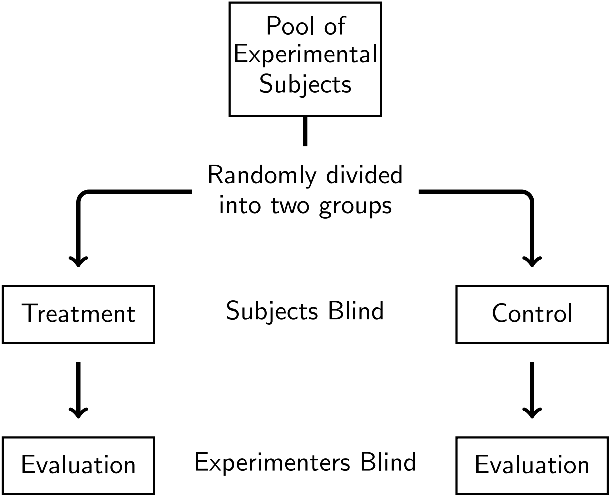
Experiment Using Random Assignment: Randomized Experiment

Treatment Group Gets Vaccine, Control Group Doesn't

Essential Point!

Random assignment *neutralizes* effect of all confounding factors: since groups are initially equal, on average, any difference that emerges must be the treatment effect.

Placebo Effect and Randomized Double Blind Experiment


Pool of Experimental Subjects
Randomly divided into two groups
Treatment
Subjects Blind
Control
Evaluation
Experimenters Blind
Evaluation

Gold Standard: Randomized, Double-blind Experiment

Randomized blind experiments ensure that on average the two groups are initially equal, and continue to be treated equally. Thus a fair comparison is possible.

Randomized, double-blind experiments are considered the "gold standard" for untangling causation.

Sugar Doesn't Make Kids Hyper

http://www.youtube.com/watch?v=mkr9YsmrPAI

Randomization is not always possible, practical, or ethical.

Observational Data

Data that do not come from a randomized experiment.

It much more challenging to untangle cause and effect using observational data because of confounders. But sometimes it's all we have.

Racial Bias in the Labor Market

Bertrand & Mullainathan (2004, American Economic Review)

When faced with observably similar African-American and White applicants, do they [employers] favor the White one? Some argue yes, citing either employer prejudice or employer perception that race signals lower productivity. Others argue that differential treatment by race is a relic of the past . . . Data limitations make it difficult to empirically test these views. Since researchers possess far less data than employers do, White and African-American workers that appear similar to researchers may look very different to employers. So any racial difference in labor market outcomes could just as easily be attributed to differences that are observable to employers but unobservable to researchers.

Racial Bias in the Labor Market: continued ...

Bertrand & Mullainathan (2004, American Economic Review)

To circumvent this difficulty, we conduct a field experiment ... We send resumes in response to help-wanted ads in Chicago and Boston newspapers and measure call-back for interview for each sent resume. We experimentally manipulate the perception of race via the name of the ficticious job applicant. We randomly assign very White-sounding names (such as Emily Walsh or Greg Baker) to half the resumes and very African-American-soundsing names (such as Lakisha Washington or Jamal Jones) to the other half.

Racial Bias in the Labor Market: continued ...

Bertrand & Mullainathan (2004, American Economic Review)

Sample	White Names	African-American Names
All sent resumes	9.7	6.5
Females	9.9	6.6
Males	8.9	5.8

Table: % Callback by racial soundingness of names.

Later this semester: if there were no racial bias in callbacks, what is the chance that we would observe such large differences?

Lecture #2 – Summary Statistics Part I

Class Survey

Types of Variables

Frequency, Relative Frequency, & Histograms

Measures of Central Tendency

Measures of Variability / Spread

Class Survey

- Collect some data to analyze later in the semester.
- None of the questions are sensitive and your name will not be linked to your responses. I will post an anonymized version of the dataset on my website.
- The survey is *strictly voluntary* – if you don't want to participate, you don't have to.

Multiple Choice Entry – What is your biological sex?

(a) Male

(b) Female

Numeric Entry – How Many Credits?

How many credits are you taking this semester? Please respond using your remote.

Multiple Choice – Class Standing

What is your class standing at Penn?

(a) Freshman

(b) Sophomore

(c) Junior

(d) Senior

Multiple Choice – What is Your Eye Color?

Please enter your eye color using your remote.

(a) Black

(b) Blue

(c) Brown

(d) Green

(e) Gray

(f) Green

(g) Hazel

(h) Other

How Right-Handed are You?

The sheet in front of you contains a handedness inventory. Please complete it and calculate your handedness score:

$$\frac{\text{Right} - \text{Left}}{\text{Right} + \text{Left}}$$

When finished, enter your score using your remote.

What is your Height in Inches?

Using your remote, please enter your height in inches, rounded to the nearest inch:

$4\text{ft} = 48\text{in}$

$5\text{ft} = 60\text{in}$

$6\text{ft} = 72\text{in}$

$7\text{ft} = 84\text{in}$

What is your Hand Span (in cm)?

On the sheet in front of you is a ruler. Please use it to measure the span of your right hand in centimeters, to the nearest 1/2 cm.

> *Hand Span: the distance from thumb to little finger when your fingers are spread apart*

When ready, enter your measurement using your remote.

We chose (by computer) a random number between 0 and 100. The number selected and assigned to you is written on the slip of paper in front of you. Please do not show your number to anyone else or look at anyone else's number.

Please enter your number now using your remote.

Call your random number X. Do you think that the percentage of countries, among all those in the United Nations, that are in Africa is higher or lower than X?

(a) Higher

(b) Lower

Please answer using your remote.

What is your best estimate of the percentage of countries, among all those that are in the United Nations, that are in Africa?

Please enter your answer using your remote.

Types of Variables

Categorical

Qualitative, assigns each unit to category, number either meaningless or indicates order only

- **Nominal** no order to the categories
- **Ordinal** categories with natural order

Types of Variables

Categorical

Qualitative, assigns each unit to category, number either meaningless or indicates order only

- **Nominal** no order to the categories
- **Ordinal** categories with natural order

Numerical

Quantitative, number meaningful

- **Discrete** Value from discrete set (often count data)
- **Continuous** Value could conceptually be any real number within some range (though *measurements* have finite precision)

Types of Variables – Examples

Categorical (called *Factor* in R)

Nominal eye color, sex

Ordinal course evaluations ($0 =$ Poor, $1 =$ Fair, ...)

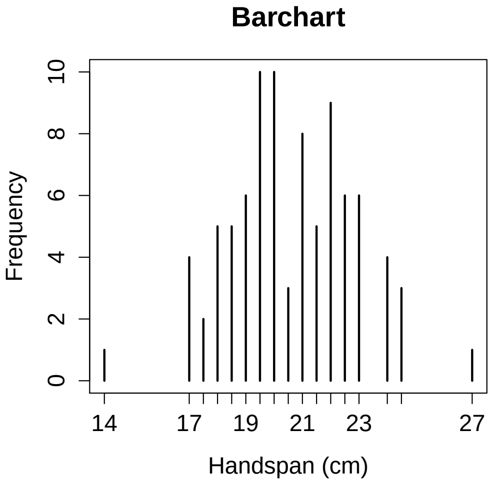
Numerical

Discrete credits you are taking this semester

Continuous height, handspan, handedness (from survey)

Handspan - Frequency and Relative Frequency

cm	Freq.	Rel. Freq.
14.0	1	0.01
17.0	4	0.05
17.5	2	0.02
18.0	5	0.06
18.5	5	0.06
19.0	6	0.07
19.5	10	0.11
20.0	10	0.11
20.5	3	0.03
21.0	8	0.09
21.5	5	0.06
22.0	9	0.10
22.5	6	0.07
23.0	6	0.07
24.0	4	0.05
24.5	3	0.03
27.0	1	0.01
	$n = 89$	1.00



Handspan - Summarize Barchart by "Smoothing"

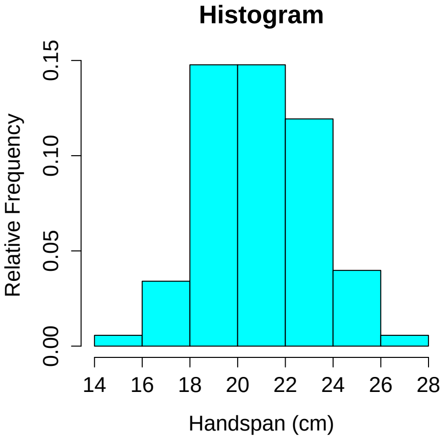
cm	Freq.	Rel. Freq.
14.0	1	0.01
17.0	4	0.05
17.5	2	0.02
18.0	5	0.06
18.5	5	0.06
19.0	6	0.07
19.5	10	0.11
20.0	10	0.11
20.5	3	0.03
21.0	8	0.09
21.5	5	0.06
22.0	9	0.10
22.5	6	0.07
23.0	6	0.07
24.0	4	0.05
24.5	3	0.03
27.0	1	0.01
	$n = 88$	1.00

Group data into non-overlapping bins of equal width:

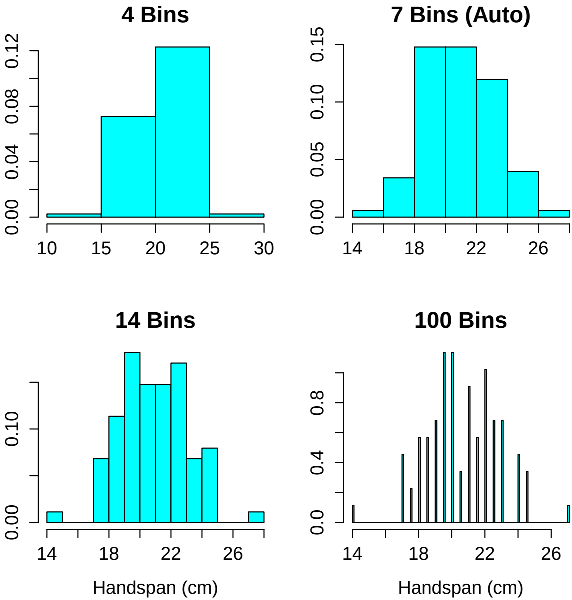
Bins	Freq.	Rel. Freq.
$[14, 16)$	1	0.01
$[16, 18)$	6	0.07
$[18, 20)$	26	0.30
$[20, 22)$	26	0.30
$[22, 24)$	21	0.24
$[24, 26)$	7	0.08
$[26, 28)$	1	0.01
	$n = 88$	1.00

Histogram – Density Estimate by Smoothing Barchart

Bins	Freq.	Rel. Freq.
$[14, 16)$	1	0.01
$[16, 18)$	6	0.07
$[18, 20)$	26	0.30
$[20, 22)$	26	0.30
$[22, 24)$	21	0.24
$[24, 26)$	7	0.08
$[26, 28)$	1	0.01
	$n = 88$	1.00



Number of Bins Controls Degree of Smoothing



Histograms are *Really* Important

Why Histogram?

Summarize numerical data, especially continuous (few repeats)

Too Many Bins – Undersmoothing

No longer a summary (lose the shape of distribution)

Too Few Bins – Oversmoothing

Miss important detail

Don't confuse with barchart!

Summary Statistic: Numerical Summary of Sample

1. Measures of Central Tendency
   - Mean
   - Median
2. Measures of Spread
   - Variance
   - Standard Deviation
   - Range
   - Interquartile Range (IQR)
3. Measures of Symmetry
   - Skewness
4. Measures of relationship between variables
   - Covariance
   - Correlation
   - Regression

Questions to Ask Yourself about Each Summary Statistic

1. What does it measure?
2. What are its units compared to those of the data?
3. (How) do its units change if those of the data change?
4. What are the benefits and drawbacks of this statistic?

Some of the information regarding items 2 and 3 is on the homework rather than in the slides because working it out for yourself is a good way to check your understanding.

What is an Outlier?

Outlier

A very unusual observation relative to the other observations in the dataset (i.e. very small or very big).

Measures of Central Tendency

Suppose we have a dataset with observations $x_1, x_2, \ldots, x_n$

Sample Mean

- $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- Only for numeric data
- Sensitive to asymmetry and outliers

Measures of Central Tendency

Suppose we have a dataset with observations $x_1, x_2, \ldots, x_n$

Sample Mean

- $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- Only for numeric data
- Sensitive to asymmetry and outliers

Sample Median

- Middle observation if n is odd, otherwise the mean of the two observations closest to the middle.
- Applicable to numerical or ordinal data
- Insensitive to outliers and skewness

Mean is Sensitive to Outliers, Median Isn't

First Dataset: 1 2 3 4 5

Mean = 3, Median = 3

Mean is Sensitive to Outliers, Median Isn't

First Dataset: 1 2 3 4 5

Mean = 3, Median = 3

Second Dataset: 1 2 3 4 4990

Mean = 1000, Median = 3

Mean is Sensitive to Outliers, Median Isn't

First Dataset: 1 2 3 4 5

Mean = 3, Median = 3

Second Dataset: 1 2 3 4 4990

Mean = 1000, Median = 3

When Does the Median Change?

Ranks would have to change so that 3 is no longer in the middle.

Percentage of UN Countries that are in Africa

You Were a Subject in a Randomized Experiment!

- There were only two numbers in the bag: 10 and 65
- Randomly assigned to Low group (10) or High group (65)

Percentage of UN Countries that are in Africa

You Were a Subject in a Randomized Experiment!

- There were only two numbers in the bag: 10 and 65
- Randomly assigned to Low group (10) or High group (65)

Anchoring Heuristic (Kahneman and Tversky, 1974)

Subjects' estimates of an unknown quantity are influenced by an irrelevant previously supplied starting point.

Are Penn students subject to to this cognitive bias?

Last Semester's Class

	Mean	Median
Low ($n = 43$)	17.1	17
High ($n = 46$)	30.7	30

Last Semester's Class

	Mean	Median
Low ($n = 43$)	17.1	17
High ($n = 46$)	30.7	30

Kahneman and Tversky (1974)

Low Group (shown 10) $\rightarrow$ median answer of 25

High Group (shown 65) $\rightarrow$ median answer of 45

(Kahneman shared 2002 Economics Nobel Prize with Vernon Smith.)

Percentiles (aka Quantiles) – Generalization of Median

$\boxed{\text{Approx. } P\% \text{ of the data are at or below the } P^{th} \text{ percentile.}}$

Percentiles (aka Quantiles)

P^{th} Percentile $=$ Value in $(P/100) \cdot (n+1)^{th}$ Ordered Position

Percentiles (aka Quantiles) – Generalization of Median

Approx. $P\%$ of the data are at or below the P^{th} percentile.

Percentiles (aka Quantiles)

P^{th} Percentile = Value in $(P/100) \cdot (n+1)^{th}$ Ordered Position

Quartiles

Q1 = 25th Percentile

Q2 = Median (i.e. 50th Percentile)

Q3 = 75th Percentile

An Example: $n = 12$

60 63 65 67 70 72 75 75 80 82 84 85

$$Q_1 = \text{value in the } 0.25(n+1)^{th} \text{ ordered position}$$

An Example: n = 12

60 63 65 67 70 72 75 75 80 82 84 85

$$\begin{aligned} Q_1 &= \text{value in the } 0.25(n+1)^{th} \text{ ordered position} \\ &= \text{value in the } 3.25^{th} \text{ ordered position} \end{aligned}$$

An Example: n = 12

60 63 65 67 70 72 75 75 80 82 84 85

$$
\begin{aligned}
\mathrm{Q}_1 &= \text{value in the } 0.25(n+1)^{th} \text{ ordered position} \\
&= \text{value in the } 3.25^{th} \text{ ordered position} \\
&= 65 + 0.25 * (67 - 65)
\end{aligned}
$$

An Example: n = 12

60 63 65 67 70 72 75 75 80 82 84 85

$$
\begin{aligned}
Q_1 &= \text{value in the } 0.25(n+1)^{th} \text{ ordered position} \\
&= \text{value in the } 3.25^{th} \text{ ordered position} \\
&= 65 + 0.25 * (67 - 65) \\
&= 65.5
\end{aligned}
$$

Quantiles of Student Debt

Source: Avery & Turner (2012)

Table 4

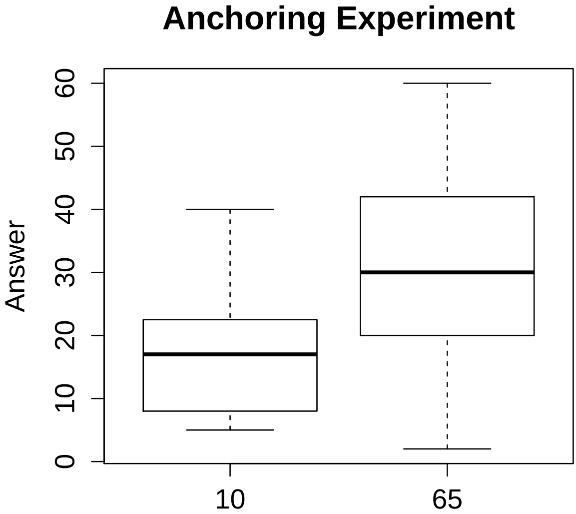
Borrowing Distribution after Six Years, by Degree Type and First Institution

	Type of institution of first enrollment			
	Public 4-year	*Private nonprofit 4-year*	*Private for-profit 4-year*	*Public 2-year*
All students beginning in 2004				
% Borrowing	61%	68%	89%	41%
Percentile of borrowers				
10^{th}	$0	$0	$0	$0
25^{th}	$0	$0	$6,376	$0
50^{th}	$6,000	$11,500	$13,961	$0
75^{th}	$19,000	$24,750	$28,863	$6,625
90^{th}	$30,000	$40,000	$45,000	$18,000
Mean	**$11,706**	**$16,606**	**$19,726**	**$5,586**
BA recipients				
BA completion	61.5%	70.7%	14.8%	13%
% Borrowing	59%	66%	92%	69%
Percentile of borrowers				
10^{th}	$0	$0	$12,000	$0
25^{th}	$0	$0	$30,000	$0
50^{th}	$7,500	$15,500	$45,000	$11,971
75^{th}	$20,000	$27,000	$50,000	$23,265
90^{th}	$32,405	$45,000	$100,000	$40,000
Mean	**$12,922**	**$18,700**	**$45,042**	**$15,960**

Source: Authors' tabulations based on the Beginning Postsecondary Survey 2004:2009.

Boxplots and the Five-Number Summary

Minimum $<$ Q1 $<$ Median $<$ Q3 $<$ Maximum



Measures of Variability/Spread – 1

Range

- Range = Maximum Observation - Minimum Observation
- *Very* sensitive to outliers.
- Displayed in boxplot.

Interquartile Range (IQR)

- IQR= $Q_3 - Q_1$
- IQR = Range of middle 50% of the data.
- Insensitive to outliers.
- Displayed in boxplot.

Measures of Variability/Spread – 2

Variance

- $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
- Essentially the average squared distance from the mean.
- (We'll talk about $n-1$ versus n later in the semester)
- Sensitive to both skewness and outliers.

Standard Deviation

- $s = \sqrt{s^2}$
- Same information as variance but more convenient since it has the same units as the data

Measures of Spread for Anchoring Experiment

Past Semester's Data

Treatment:	X = 10	X = 65
Range	35	58
IQR	14.5	21
S.D.	9.3	15.9
Var.	86.1	253.5

Lecture #3 – Summary Statistics Part II

Why squares in the definition of variance?

Outliers, Skewness, & Symmetry

Sample versus Population, Empirical Rule

Centering, Standardizing, & Z-Scores

Relating Two Variables: Cross-tabs, Covariance, & Correlation

Why Squares?

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

What's Wrong With This?

$$\frac{1}{n-1}\sum_{i=1}^{N}(x_i - \bar{x}) \quad =$$

Why Squares?

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

What's Wrong With This?

$$\frac{1}{n-1}\sum_{i=1}^{N}(x_i - \bar{x}) \quad = \quad \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\bar{x}\right] =$$

Why Squares?

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

What's Wrong With This?

$$\frac{1}{n-1}\sum_{i=1}^{N}(x_i - \bar{x}) \quad = \quad \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x}\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i - n\bar{x}\right]$$

$$=$$

Why Squares?

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

What's Wrong With This?

$$\begin{aligned}
\frac{1}{n-1}\sum_{i=1}^{N}(x_i - \bar{x}) &= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\bar{x}\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i - n\bar{x}\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i - n\cdot\frac{1}{n}\sum_{i=1}^{n} x_i\right] \\
&=
\end{aligned}$$

Why Squares?

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

What's Wrong With This?

$$\begin{aligned}
\frac{1}{n-1}\sum_{i=1}^{N}(x_i - \bar{x}) &= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\bar{x}\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i - n\bar{x}\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i - n\cdot\frac{1}{n}\sum_{i=1}^{n} x_i\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i\right] =
\end{aligned}$$

Why Squares?

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

What's Wrong With This?

$$\begin{aligned}
\frac{1}{n-1}\sum_{i=1}^{N}(x_i - \bar{x}) &= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\bar{x}\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i - n\bar{x}\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i - n\cdot\frac{1}{n}\sum_{i=1}^{n} x_i\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i\right] = 0
\end{aligned}$$

Variance is Sensitive to Skewness and Outliers

And so is Standard Deviation!

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Outliers

Differentiate with respect to $(x_i - \bar{x}) \Rightarrow$ the farther an observation is from the mean, the *larger* its effect on the variance.

Skewness

Variance measures average squared distance from center, taking mean as the center, but the mean is sensitive to skewness!

Skewness – A Measure of Symmetry

$$\boxed{\text{Skewness} = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}}$$

Skewness – A Measure of Symmetry

$$\boxed{\text{Skewness} = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}}$$

What do the values indicate?

Zero $\Rightarrow$ symmetry, positive right-skewed, negative left-skewed.

Skewness – A Measure of Symmetry

$$\boxed{\text{Skewness} = \frac{1}{n} \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{s^3}}$$

What do the values indicate?

Zero $\Rightarrow$ symmetry, positive right-skewed, negative left-skewed.

Why cubed?

To get the desired sign.

Skewness – A Measure of Symmetry

$$\boxed{\text{Skewness} = \frac{1}{n} \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{s^3}}$$

What do the values indicate?

Zero $\Rightarrow$ symmetry, positive right-skewed, negative left-skewed.

Why cubed?

To get the desired sign.

Why divide by s^3?

So that skewness is unitless

Skewness – A Measure of Symmetry

$$\boxed{\text{Skewness} = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}}$$

What do the values indicate?

Zero $\Rightarrow$ symmetry, positive right-skewed, negative left-skewed.

Why cubed?

To get the desired sign.

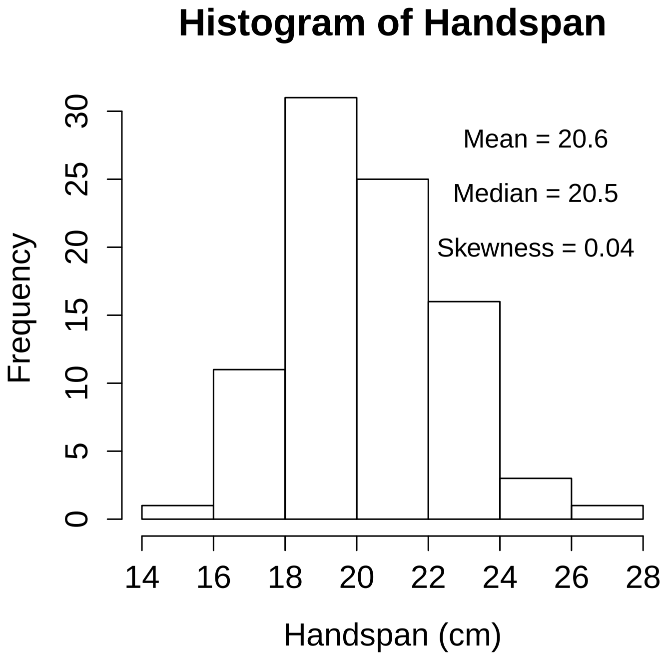
Why divide by s^3?

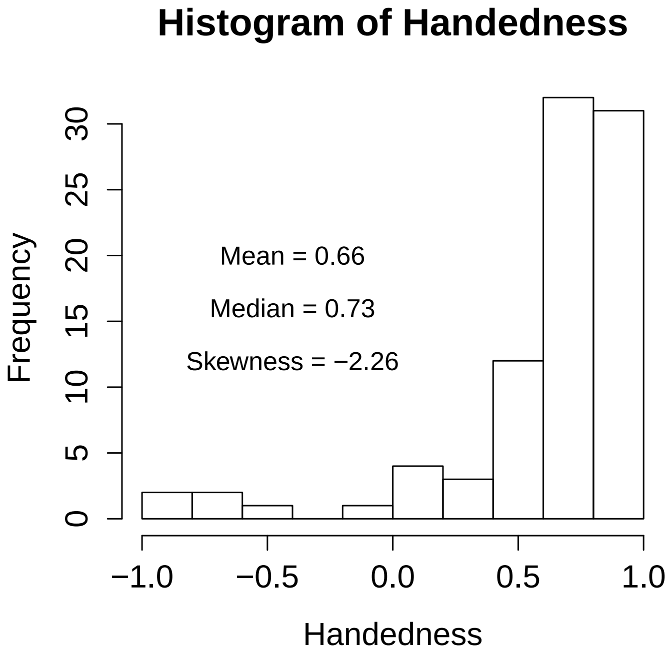
So that skewness is unitless

Rule of Thumb

Typically (but not always), right-skewed $\Rightarrow$ mean $>$ median

left-skewed $\Rightarrow$ mean $<$ median


Histogram of Handspan
Mean = 20.6
Median = 20.5
Skewness = 0.04
Frequency
0
5
10
15
20
25
30
14
16
18
20
22
24
26
28
Handspan (cm)



Histogram of Handedness
Mean = 0.66
Median = 0.73
Skewness = −2.26
Frequency
0
5
10
15
20
25
30
−1.0
−0.5
0.0
0.5
1.0
Handedness

Essential Distinction: Sample vs. Population

For now, you can think of the population as a list of N objects:

Population: $x_1, x_2, \ldots, x_N$

from which we draw a sample of size $n < N$ objects:

Sample: $x_1, x_2, \ldots, x_n$

Essential Distinction: Sample vs. Population

For now, you can think of the population as a list of N objects:

Population: $x_1, x_2, \ldots, x_N$

from which we draw a sample of size $n < N$ objects:

Sample: $x_1, x_2, \ldots, x_n$

Important Point:

Later in the course we'll be more formal by considering probability models that represent the *act of sampling* from a population rather than thinking of a population as a list of objects. Once we do this we will no longer use the notation N as the population will be *conceptually infinite*.

Essential Distinction: Parameter vs. Statistic

N individuals in the Population, n individuals in the Sample:

	Parameter (Population)	**Statistic** (Sample)
Mean	$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Var.	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
S.D.	$\sigma = \sqrt{\sigma^2}$	$s = \sqrt{s^2}$

Key Point

We use a sample $x_1, \ldots, x_n$ to calculate statistics (e.g. $\bar{x}$, s^2, s) that serve as estimates of the corresponding population parameters (e.g. μ, σ^2, σ).

Why Do Sample Variance and Std. Dev. Divide by $n-1$?

Pop. Var.	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$	Sample Var.	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
Pop. S.D.	$\sigma = \sqrt{\sigma^2}$	Sample S.D.	$s = \sqrt{s^2}$

There is an important reason for this, but explaining it requires some concepts we haven't learned yet.

Why Mean and Variance (and Std. Dev.)?

Empirical Rule

For large populations that are approximately bell-shaped, std. dev. tells where most observations will be relative to the mean:

- $\approx$ 68% of observations are in the interval $\mu \pm \sigma$
- $\approx$ 95% of observations are in the interval $\mu \pm 2\sigma$
- Almost all of observations are in the interval $\mu \pm 3\sigma$

Why Mean and Variance (and Std. Dev.)?

Empirical Rule

For large populations that are approximately bell-shaped, std. dev. tells where most observations will be relative to the mean:

- $\approx$ 68% of observations are in the interval $\mu \pm \sigma$
- $\approx$ 95% of observations are in the interval $\mu \pm 2\sigma$
- Almost all of observations are in the interval $\mu \pm 3\sigma$

Therefore

We will be interested in $\bar{x}$ as an estimate of μ and s as an estimate of σ since these population parameters are so informative.

Which is more "extreme?"

(a) Handspan of 27cm

(b) Height of 78in

Centering: Subtract the Mean

Handspan	Height
$27\text{cm} - 20.6\text{cm} = 6.4\text{cm}$	$78\text{in} - 67.6\text{in} = 10.4\text{in}$

Standardizing: Divide by S.D.

Handspan	Height
$27\text{cm} - 20.6\text{cm} = 6.4\text{cm}$	$78\text{in} - 67.6\text{in} = 10.4\text{in}$
$6.4\text{cm}/2.2\text{cm} \approx 2.9$	$10.4\text{in}/4.5\text{in} \approx 2.3$

Standardizing: Divide by S.D.

Handspan	Height
$27\text{cm} - 20.6\text{cm} = 6.4\text{cm}$	$78\text{in} - 67.6\text{in} = 10.4\text{in}$
$6.4\text{cm}/2.2\text{cm} \approx 2.9$	$10.4\text{in}/4.5\text{in} \approx 2.3$

The units have disappeared!

Z-scores: How many standard deviations from the mean?

Best for Symmetric Distribution, No Outliers (Why?)

$$z_i = \frac{x_i - \bar{x}}{s}$$

Z-scores: How many standard deviations from the mean?

Best for Symmetric Distribution, No Outliers (Why?)

$$z_i = \frac{x_i - \bar{x}}{s}$$

Unitless

Allows comparison of variables with different units.

Z-scores: How many standard deviations from the mean?

Best for Symmetric Distribution, No Outliers (Why?)

$$z_i = \frac{x_i - \bar{x}}{s}$$

Unitless

Allows comparison of variables with different units.

Detecting Outliers

Measures how "extreme" one observation is relative to the others.

Z-scores: How many standard deviations from the mean?

Best for Symmetric Distribution, No Outliers (Why?)

$$z_i = \frac{x_i - \bar{x}}{s}$$

Unitless

Allows comparison of variables with different units.

Detecting Outliers

Measures how "extreme" one observation is relative to the others.

Linear Transformation

What is the sample mean of the z-scores?

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$$

What is the sample mean of the z-scores?

$$\bar{z} \;=\; \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i - \bar{x}}{s} =$$

What is the sample mean of the z-scores?

$$
\begin{aligned}
\bar{z} &= \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i - \bar{x}}{s} = \frac{1}{n \cdot s}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\bar{x}\right] \\
&=
\end{aligned}
$$

What is the sample mean of the z-scores?

$$\begin{aligned}\bar{z} &= \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i - \bar{x}}{s} = \frac{1}{n \cdot s}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\bar{x}\right] \\ &= \frac{1}{n \cdot s}\left[\sum_{i=1}^{n} x_i - n\bar{x}\right] =\end{aligned}$$

What is the sample mean of the z-scores?

$$
\begin{aligned}
\bar{z} &= \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i-\bar{x}}{s} = \frac{1}{n\cdot s}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\bar{x}\right] \\
&= \frac{1}{n\cdot s}\left[\sum_{i=1}^{n} x_i - n\bar{x}\right] = \frac{1}{n\cdot s}\left[\sum_{i=1}^{n} x_i - n\cdot\frac{1}{n}\sum_{i=1}^{n} x_i\right] \\
&=
\end{aligned}
$$

What is the sample mean of the z-scores?

$$\begin{aligned}
\bar{z} &= \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i - \bar{x}}{s} = \frac{1}{n \cdot s}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\bar{x}\right] \\
&= \frac{1}{n \cdot s}\left[\sum_{i=1}^{n} x_i - n\bar{x}\right] = \frac{1}{n \cdot s}\left[\sum_{i=1}^{n} x_i - n \cdot \frac{1}{n}\sum_{i=1}^{n} x_i\right] \\
&= \frac{1}{n \cdot s}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i\right] =
\end{aligned}$$

What is the sample mean of the z-scores?

$$
\begin{aligned}
\bar{z} &= \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i - \bar{x}}{s} = \frac{1}{n \cdot s}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\bar{x}\right] \\
&= \frac{1}{n \cdot s}\left[\sum_{i=1}^{n} x_i - n\bar{x}\right] = \frac{1}{n \cdot s}\left[\sum_{i=1}^{n} x_i - n \cdot \frac{1}{n}\sum_{i=1}^{n} x_i\right] \\
&= \frac{1}{n \cdot s}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i\right] = 0
\end{aligned}
$$

What is the variance of the z-scores?

$$s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})^2 =$$

What is the variance of the z-scores?

$$s_z^2 \;=\; \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})^2 = \frac{1}{n-1}\sum_{i=1}^{n} z_i^2 =$$

What is the variance of the z-scores?

$$s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})^2 = \frac{1}{n-1}\sum_{i=1}^{n} z_i^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)^2$$

$$=$$

What is the variance of the z-scores?

$$
\begin{aligned}
s_z^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})^2 = \frac{1}{n-1}\sum_{i=1}^{n} z_i^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)^2 \\
&= \frac{1}{s_x^2}\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right] =
\end{aligned}
$$

What is the variance of the z-scores?

$$
\begin{aligned}
s_z^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(z_i-\bar{z})^2 = \frac{1}{n-1}\sum_{i=1}^{n} z_i^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)^2 \\
&= \frac{1}{s_x^2}\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right] = \frac{s_x^2}{s_x^2} =
\end{aligned}
$$

What is the variance of the z-scores?

$$
\begin{aligned}
s_z^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})^2 = \frac{1}{n-1}\sum_{i=1}^{n} z_i^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)^2 \\
&= \frac{1}{s_x^2}\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right] = \frac{s_x^2}{s_x^2} = 1
\end{aligned}
$$

What is the variance of the z-scores?

$$
\begin{aligned}
s_z^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})^2 = \frac{1}{n-1}\sum_{i=1}^{n} z_i^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)^2 \\
&= \frac{1}{s_x^2}\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right] = \frac{s_x^2}{s_x^2} = 1
\end{aligned}
$$

So what is the *standard deviation* of the z-scores?

Population Z-scores and the Empirical Rule: $\mu \pm 2\sigma$

If we knew the population mean μ and standard deviation σ we could create a *population version* of a z-score. This leads to an important way of rewriting the Empirical Rule:

Population Z-scores and the Empirical Rule: $\mu \pm 2\sigma$

If we knew the population mean μ and standard deviation σ we could create a *population version* of a z-score. This leads to an important way of rewriting the Empirical Rule:

Bell-shaped population $\Rightarrow$ approx. 95% of observations x_i satisfy

$$\mu - 2\sigma \leq x_i \leq \mu + 2\sigma$$

Population Z-scores and the Empirical Rule: $\mu \pm 2\sigma$

If we knew the population mean μ and standard deviation σ we could create a *population version* of a z-score. This leads to an important way of rewriting the Empirical Rule:

Bell-shaped population $\Rightarrow$ approx. 95% of observations x_i satisfy

$$\mu - 2\sigma \leq x_i \leq \mu + 2\sigma$$

$$-2\sigma \leq x_i - \mu \leq 2\sigma$$

Population Z-scores and the Empirical Rule: $\mu \pm 2\sigma$

If we knew the population mean μ and standard deviation σ we could create a *population version* of a z-score. This leads to an important way of rewriting the Empirical Rule:

Bell-shaped population $\Rightarrow$ approx. 95% of observations x_i satisfy

$$\begin{aligned} \mu - 2\sigma \leq x_i &\leq \mu + 2\sigma \\ -2\sigma \leq x_i - \mu &\leq 2\sigma \\ -2 \leq \frac{x_i - \mu}{\sigma} &\leq 2 \end{aligned}$$

Relationships Between Variables

Crosstabs – Show Relationship between Categorical Vars.

(aka Contingency Tables)

Eye Color	*Sex*		
	Male	Female	Total
Black	5	2	7
Blue	6	4	10
Brown	26	31	57
Copper	1	0	1
Dark Brown	0	1	1
Green	4	1	5
Hazel	2	2	4
Maroon	1	0	1
Total	45	41	86

Example with Crosstab in *Percents*

Who Supported the Vietnam War?

In January 1971 the Gallup poll asked: "A proposal has been made in Congress to require the U.S. government to bring home all U.S. troops before the end of this year. Would you like to have your congressman vote for or against this proposal?"

Guess the results, for respondents in each education category, and fill out this table (the two numbers in each column should add up to 100%):

	Adults with:			
	Grade school education	High school education	College education	Total adults
% for withdrawal of U.S. troops (doves)				73%
% against withdrawal of U.S. troops (hawks)				27%
Total	100%	100%	100%	100%

Who Were the Doves?

Which group do you think was most strongly in favor of the withdrawal of US troops from Vietnam?

(a) Adults with only a Grade School Education

(b) Adults with a High School Education

(c) Adults with a College Education

Please respond with your remote.

Who Were the Hawks?

Which group do you think was most strongly opposed to the withdrawal of US troops from Vietnam?

(a) Adults with only a Grade School Education

(b) Adults with a High School Education

(c) Adults with a College Education

Please respond with your remote.

From The Economist –"Lexington," October 4th, 2001

"Back in the Vietnam days, the anti-war movement spread from the intelligentsia into the rest of the population, eventually paralyzing the country's will to fight."

Who *Really* Supported the Vietnam War

Gallup Poll, January 1971

	Adults with: Grade school education	High school education	College education	Total adults
% for withdrawal of U.S. troops (doves)	80%	75%	60%	73%
% against withdrawal of U.S. troops (hawks)	20%	25%	40%	27%
Total	100%	100%	100%	100%

What about numeric data?

Covariance and Correlation: Linear Dependence Measures

Two Samples of Numeric Data

$x_1, \ldots, x_n$ and $y_1, \ldots, y_n$

Dependence

Do x and y both tend to be large (or small) at the same time?

Key Point

Use the idea of centering and standardizing to decide what “big” or “small” means in this context.

Notation

$$
\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\
\bar{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i \\
s_x &= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} \\
s_y &= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2}
\end{aligned}
$$

Covariance

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

- Centers each observation around its mean and multiplies.
- Zero $\Rightarrow$ no linear dependence
- Positive $\Rightarrow$ positive linear dependence
- Negative $\Rightarrow$ negative linear dependence
- Population parameter: σ_{xy}
- Units?

Correlation

$$r_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right) = \frac{s_{xy}}{s_x s_y}$$

- Centers *and* standardizes each observation
- Bounded between -1 and 1
- Zero $\Rightarrow$ no linear dependence
- Positive $\Rightarrow$ positive linear dependence
- Negative $\Rightarrow$ negative linear dependence
- Population parameter: ρ_{xy}
- Unitless

We'll have more to say about correlation and covariance when we discuss linear regression.

Essential Distinction: Parameter vs. Statistic

And Population vs. Sample

N individuals in the Population, n individuals in the Sample:

	Parameter (Population)	**Statistic** (Sample)
Mean	$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Var.	$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$	$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
S.D.	$\sigma_x = \sqrt{\sigma_x^2}$	$s_x = \sqrt{s^2}$
Cov.	$\sigma_{xy} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{N}$	$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$
Corr.	$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$	$r = \frac{s_{xy}}{s_x s_y}$

Lecture #4 – Linear Regression I

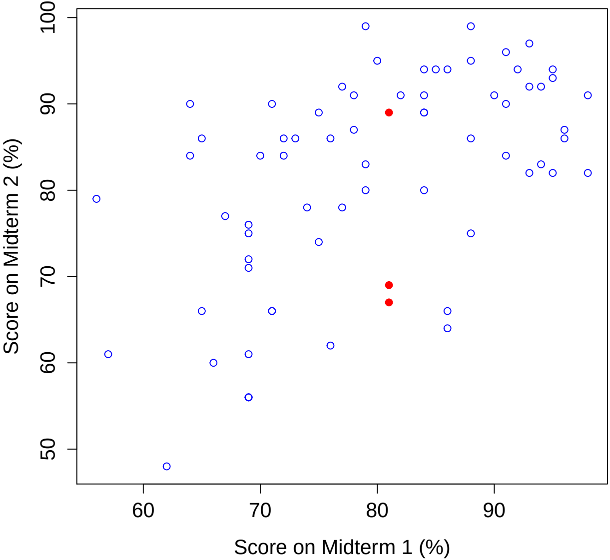
Overview / Intuition for Linear Regression

Deriving the Regression Equations

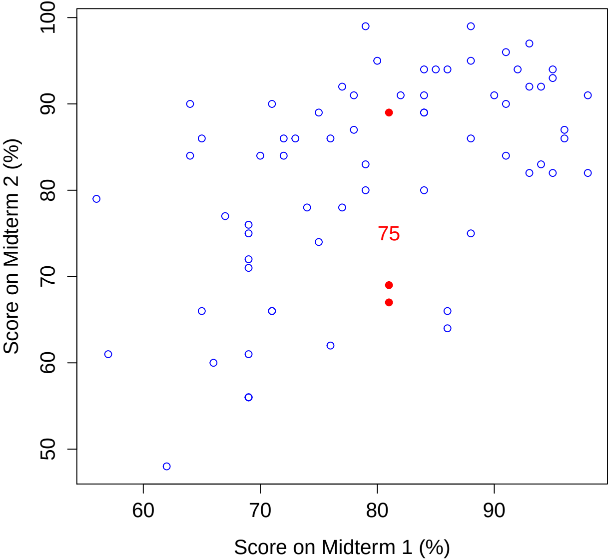
Relating Regression, Covariance and Correlation

Regression to the Mean

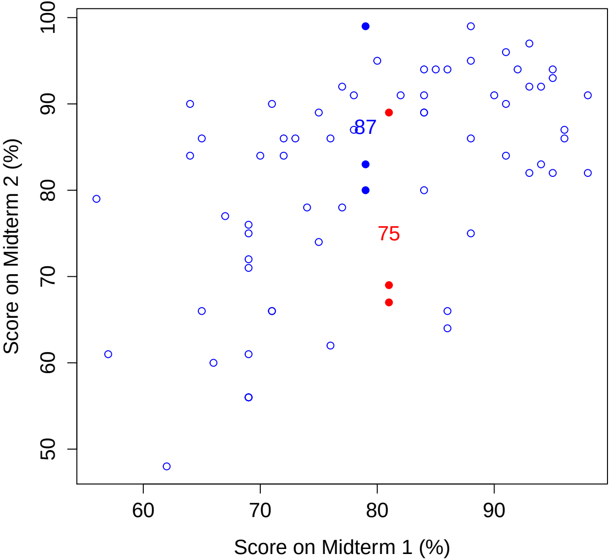
Predict Second Midterm given 81 on First



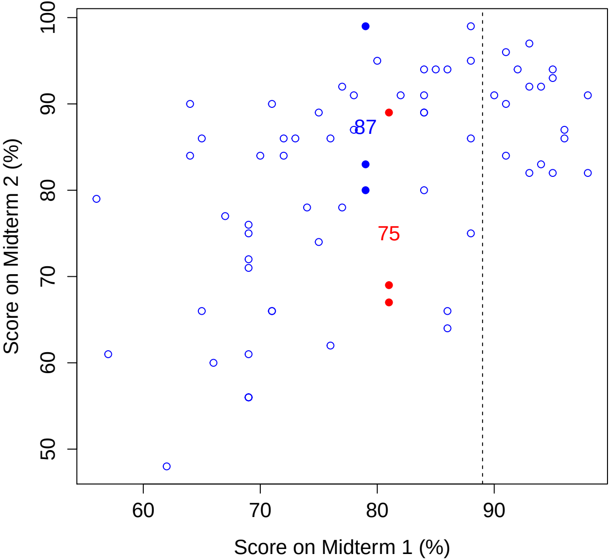
Predict Second Midterm given 81 on First



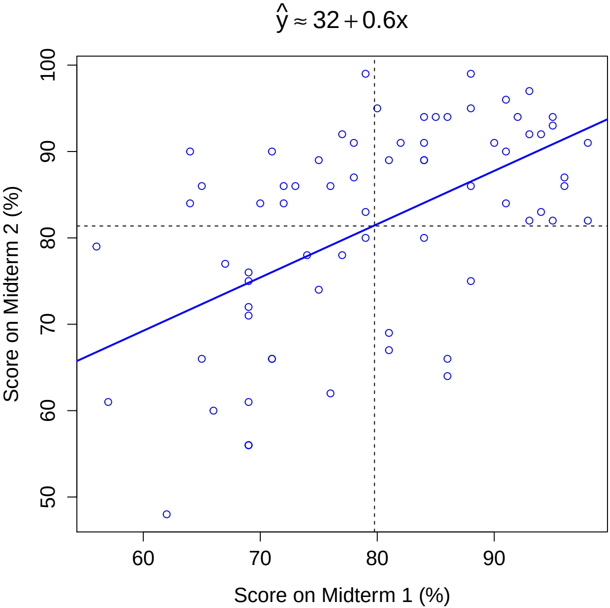
But if they'd only gotten 79 we'd predict higher?!



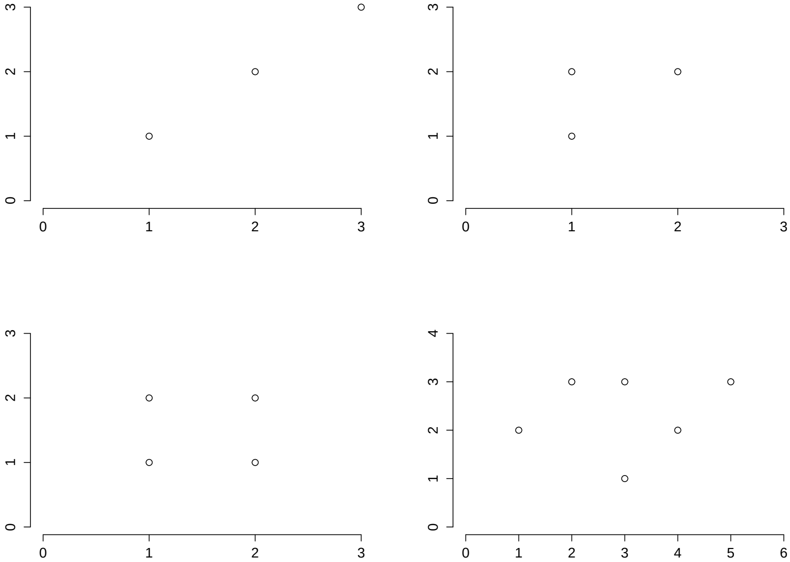
No one who took both exams got 89 on the first!

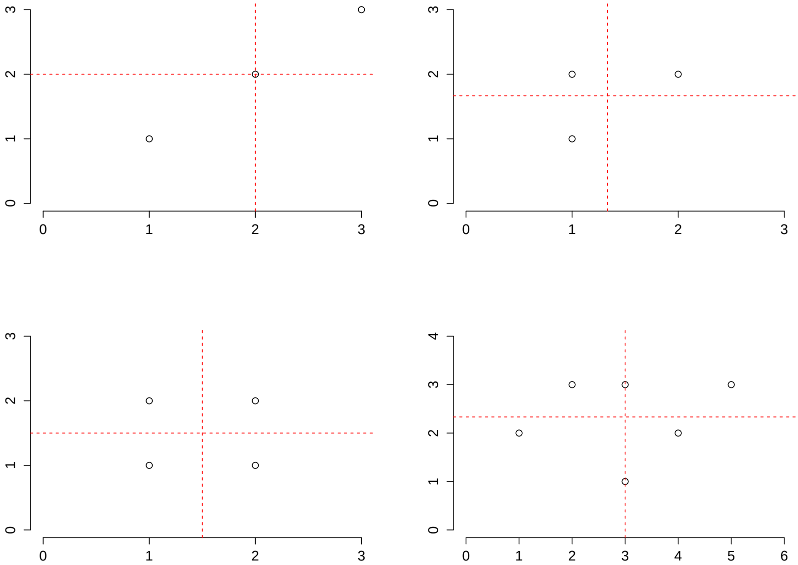


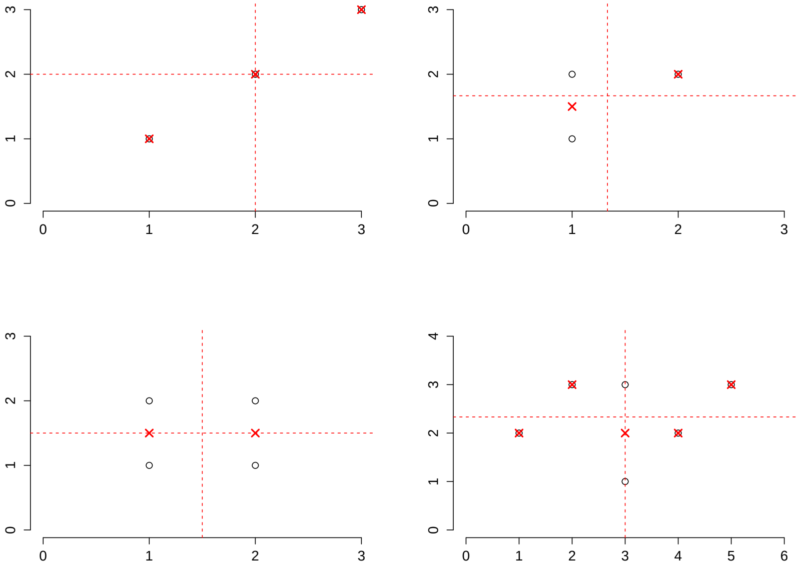
Regression: "Best Fitting" Line Through Cloud of Points



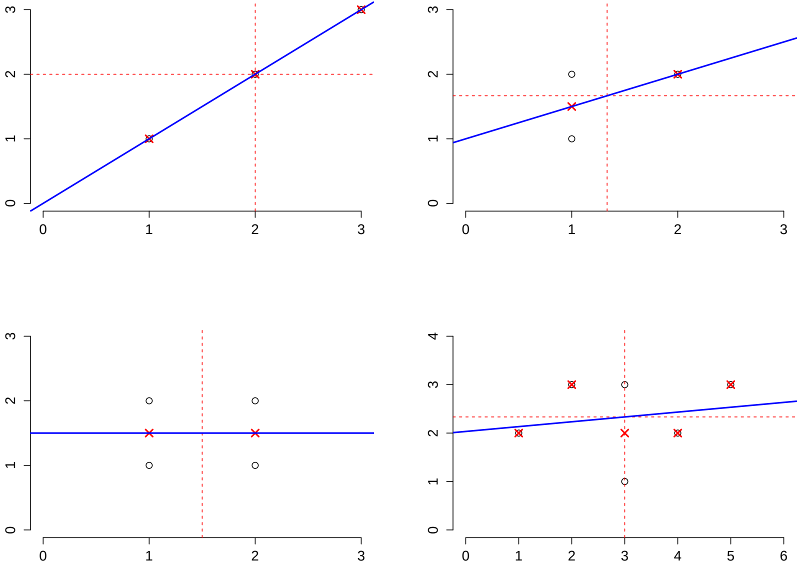
Fitting a Line by Eye




0
1
2
3
0
1
2
3
0
1
2
3
0
1
2
3
0
1
2
3
0
1
2
3
0
1
2
3
4
5
6
0
1
2
3
4



0
1
2
3
0
1
2
3
0
1
2
3
4
5
6


0
1
2
3
0
1
2
3
0
1
2
3
0
1
2
3
0
1
2
3
0
1
2
3
0
1
2
3
4
5
6
0
1
2
3
4

But How to Do this Formally?

Least Squares Regression – Predict Using a Line

The Prediction

Predict score $\hat{y} = a + bx$ on 2nd midterm if you scored x on 1st

How to choose (a, b)?

Linear regression chooses the slope (b) and intercept (a) that

minimize the sum of squared vertical deviations

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

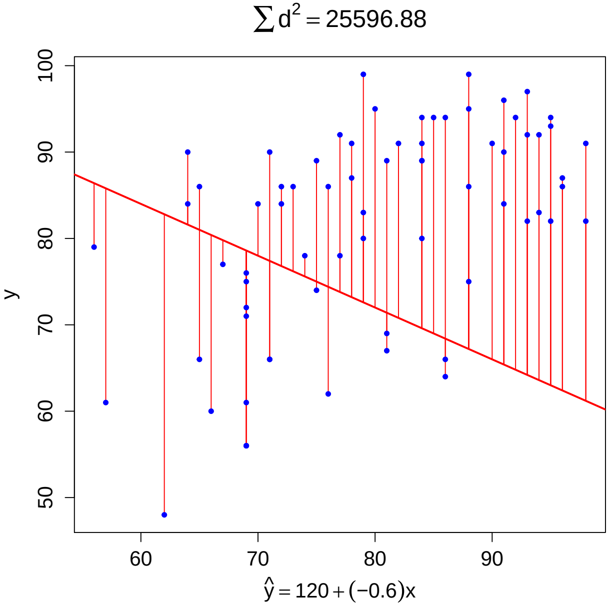
Why Squared Deviations?

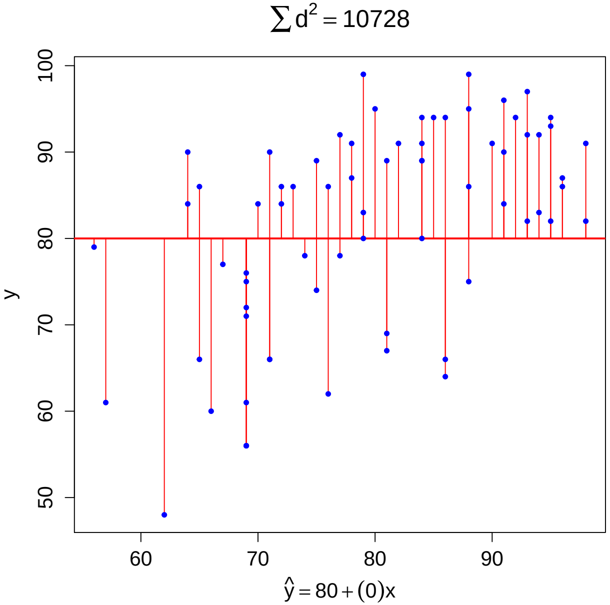
Important Point About Notation

$$\boxed{\underset{a,b}{\text{minimize}} \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2}$$

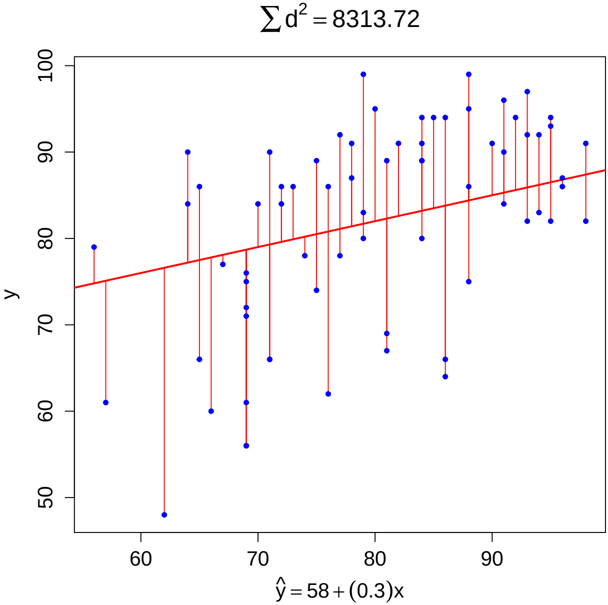
$$\boxed{\hat{y} = a + bx}$$

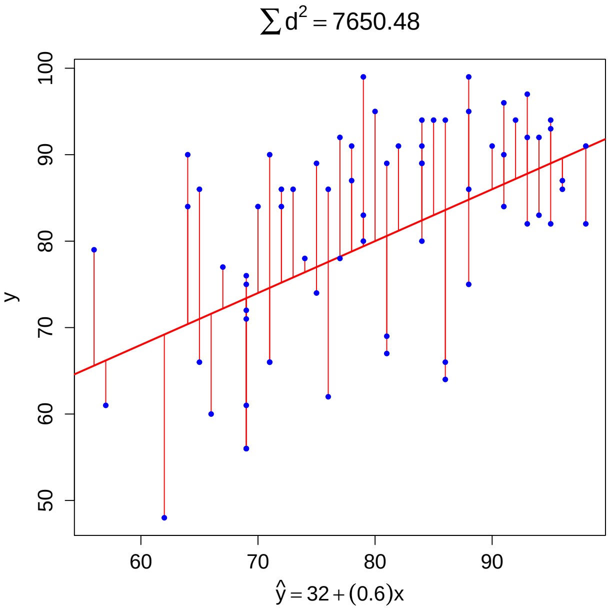
- $(x_i, y_i)_{i=1}^n$ are the observed data
- $\widehat{y}$ is our prediction for a given value of x
- Neither x nor $\widehat{y}$ needs to be in out dataset!


$\sum d^2 = 25596.88$
y
100
90
80
70
60
50
60
70
80
90
$\hat{y} = 120 + (-0.6)x$

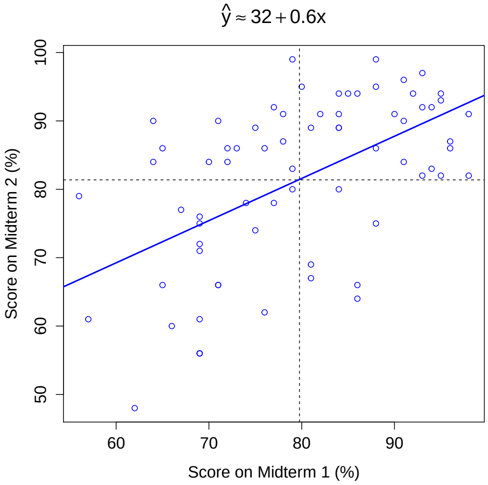


$\sum d^2 = 10728$
y
100
90
80
70
60
50
60
70
80
90
$\hat{y} = 80 + (0)x$


$\sum d^2 = 8313.72$
y
100
90
80
70
60
50
60
70
80
90
$\hat{y} = 58 + (0.3)x$


$\sum d^2 = 7650.48$
y
100
90
80
70
60
50
60
70
80
90
$\hat{y} = 32 + (0.6)x$

Prediction given 89 on Midterm 1?



$$32 + 0.6 \times 89 = 32 + 53.4 = 85.4$$

You Need to Know How To Derive This

Minimize the sum of squared vertical deviations from the line:

$$\min_{a,b} \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

How should we proceed?

(a) Differentiate with respect to x

(b) Differentiate with respect to y

(c) Differentiate with respect to x, y

(d) Differentiate with respect to a, b

(e) Can't solve this with calculus.

Objective Function

$$\min_{a,b} \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

FOC with respect to a

Objective Function

$$\min_{a,b} \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

FOC with respect to a

$$-2 \sum_{i=1}^{n} (y_i - a - bx_i) \;=\; 0$$

Objective Function

$$\min_{a,b} \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

FOC with respect to *a*

$$\begin{aligned} -2\sum_{i=1}^{n}\left(y_i - a - bx_i\right) &= 0 \\ \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a - b\sum_{i=1}^{n} x_i &= 0 \end{aligned}$$

Objective Function

$$\min_{a,b} \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

FOC with respect to a

$$\begin{aligned}
-2\sum_{i=1}^{n}(y_i - a - bx_i) &= 0 \\
\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a - b\sum_{i=1}^{n} x_i &= 0 \\
\frac{1}{n}\sum_{i=1}^{n} y_i - \frac{na}{n} - \frac{b}{n}\sum_{i=1}^{n} x_i &= 0
\end{aligned}$$

Objective Function

$$\min_{a,b} \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

FOC with respect to a

$$\begin{aligned}
-2\sum_{i=1}^{n}(y_i - a - bx_i) &= 0 \\
\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a - b\sum_{i=1}^{n} x_i &= 0 \\
\frac{1}{n}\sum_{i=1}^{n} y_i - \frac{na}{n} - \frac{b}{n}\sum_{i=1}^{n} x_i &= 0 \\
\bar{y} - a - b\bar{x} &= 0
\end{aligned}$$

Regression Line Goes Through the Means!

$$\boxed{\bar{y} = a + b\bar{x}}$$

Substitute $a = \bar{y} - b\bar{x}$

$$\sum_{i=1}^{n}(y_i - a - bx_i)^2 \quad =$$

Substitute $a = \bar{y} - b\bar{x}$

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - a - bx_i)^2 &= \sum_{i=1}^{n}(y_i - \bar{y} + b\bar{x} - bx_i)^2 \\
&=
\end{aligned}$$

Substitute $a = \bar{y} - b\bar{x}$

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - a - bx_i)^2 &= \sum_{i=1}^{n}(y_i - \bar{y} + b\bar{x} - bx_i)^2 \\
&= \sum_{i=1}^{n}\left[(y_i - \bar{y}) - b\left(x_i - \bar{x}\right)\right]^2
\end{aligned}$$

FOC wrt b

Substitute $a = \bar{y} - b\bar{x}$

$$\begin{aligned}\sum_{i=1}^{n}(y_i - a - bx_i)^2 &= \sum_{i=1}^{n}(y_i - \bar{y} + b\bar{x} - bx_i)^2 \\ &= \sum_{i=1}^{n}\left[(y_i - \bar{y}) - b\left(x_i - \bar{x}\right)\right]^2\end{aligned}$$

FOC wrt b

$$-2\sum_{i=1}^{n}\left[(y_i - \bar{y}) - b\left(x_i - \bar{x}\right)\right]\left(x_i - \bar{x}\right) = 0$$

Substitute $a = \bar{y} - b\bar{x}$

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - a - bx_i)^2 &= \sum_{i=1}^{n}(y_i - \bar{y} + b\bar{x} - bx_i)^2 \\
&= \sum_{i=1}^{n}\left[(y_i - \bar{y}) - b\left(x_i - \bar{x}\right)\right]^2
\end{aligned}$$

FOC wrt b

$$\begin{aligned}
-2\sum_{i=1}^{n}\left[(y_i - \bar{y}) - b\left(x_i - \bar{x}\right)\right]\left(x_i - \bar{x}\right) &= 0 \\
\sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right) - b\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 &= 0
\end{aligned}$$

Substitute $a = \bar{y} - b\bar{x}$

$$\begin{aligned} \sum_{i=1}^{n}(y_i - a - bx_i)^2 &= \sum_{i=1}^{n}(y_i - \bar{y} + b\bar{x} - bx_i)^2 \\ &= \sum_{i=1}^{n}\left[(y_i - \bar{y}) - b\left(x_i - \bar{x}\right)\right]^2 \end{aligned}$$

FOC wrt b

$$\begin{aligned} -2\sum_{i=1}^{n}\left[(y_i - \bar{y}) - b\left(x_i - \bar{x}\right)\right]\left(x_i - \bar{x}\right) &= 0 \\ \sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right) - b\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 &= 0 \end{aligned}$$

$$b = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

Simple Linear Regression

Problem

$$\min_{a,b} \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

Solution

$$b = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$a = \bar{y} - b\bar{x}$$

Relating Regression to Covariance and Correlation

$$b = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right)}{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \frac{s_{xy}}{s_x^2}$$

$$r = \frac{s_{xy}}{s_x s_y} = b\frac{s_x}{s_y}$$

Comparing Regression, Correlation and Covariance

Units

Correlation is unitless, covariance and regression coefficients (a, b) are not. (What are the units of these?)

Symmetry

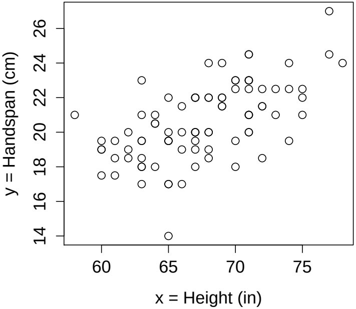
Correlation and covariance are symmetric, regression isn't. (Switching x and y axes changes the slope and intercept.)

On the Homework

Regression with z-scores rather than raw data gives $a = 0, b = r_{xy}$

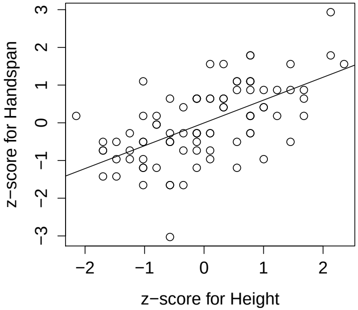
$$s_{xy} = 6, \quad s_x = 5, \quad s_y = 2, \quad \bar{x} = 68, \quad \bar{y} = 21$$

What is the sample correlation between height (x) and handspan (y)?



$$s_{xy} = 6, \quad s_x = 5, \quad s_y = 2, \quad \bar{x} = 68, \quad \bar{y} = 21$$

What is the sample correlation between height (x) and handspan (y)?



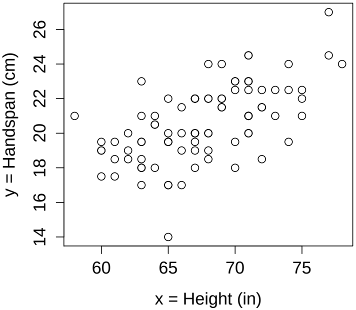
$$r = \frac{s_{xy}}{s_x s_y} = \frac{6}{5 \times 2} = 0.6$$

$$s_{xy} = 6, \quad s_x = 5, \quad s_y = 2, \quad \bar{x} = 68, \quad \bar{y} = 21$$

What is the value of b for the regression:

$$\hat{y} = a + bx$$

where x is height and y is handspan?

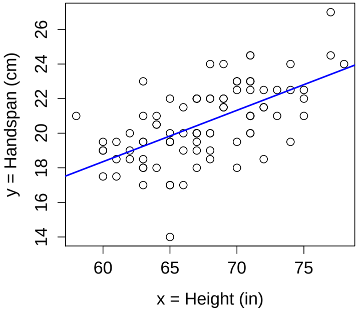


$$s_{xy} = 6, \quad s_x = 5, \quad s_y = 2, \quad \bar{x} = 68, \quad \bar{y} = 21$$

What is the value of b for the regression:

$$\hat{y} = a + bx$$

where x is height and y is handspan?



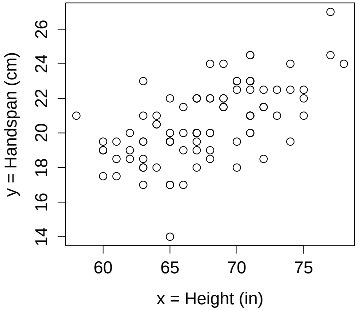
$$b = \frac{s_{xy}}{s_x^2} = \frac{6}{5^2} = 6/25 = 0.24$$

$$s_{xy} = 6, \quad s_x = 5, \quad s_y = 2, \quad \bar{x} = 68, \quad \bar{y} = 21$$

What is the value of a for the regression:

$$\hat{y} = a + bx$$

where x is height and y is handspan? (prev. slide $b = 0.24$)

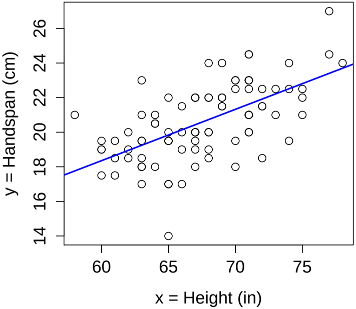


$$s_{xy} = 6, \quad s_x = 5, \quad s_y = 2, \quad \bar{x} = 68, \quad \bar{y} = 21$$

What is the value of a for the regression:

$$\hat{y} = a + bx$$

where x is height and y is handspan?
(prev. slide $b = 0.24$)



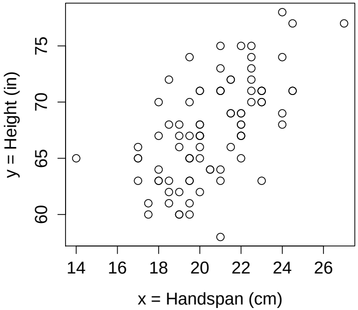
$$a = \bar{y} - b\bar{x} = 21 - 0.24 \times 68 = 4.68$$

$$s_{xy} = 6, \quad s_y = 5, \quad s_x = 2, \quad \bar{y} = 68, \quad \bar{x} = 21$$

What is the value of b for the regression:

$$\hat{y} = a + bx$$

where x is handspan and y is height?

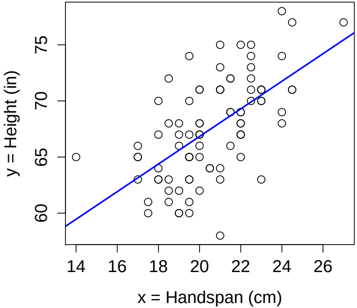


$s_{xy} = 6, \quad s_y = 5, \quad s_x = 2, \quad \bar{y} = 68, \quad \bar{x} = 21$

What is the value of b for the regression:

$$\hat{y} = a + bx$$

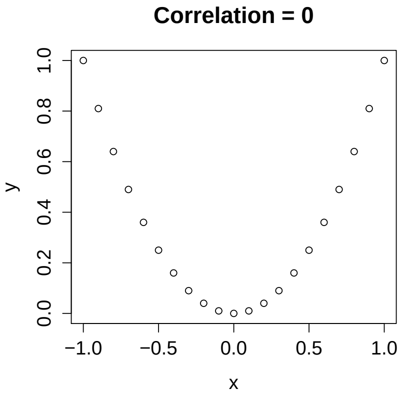
where x is handspan and y is height?



$$b = \frac{s_{xy}}{s_x^2} = 6/2^2 = 1.5$$

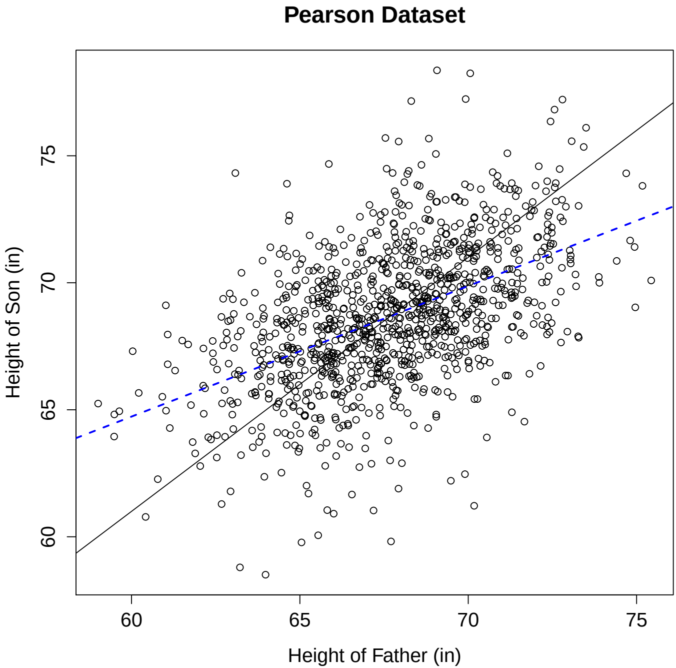
EXTREMELY IMPORTANT

- Regression, Covariance and Correlation: linear association.
- Linear association $\neq$ causation.
- Linear is not the only kind of association!



Regression to the Mean and the Regression Fallacy

Please read Chapter 17 of “Thinking Fast and Slow” by Daniel Kahnemann which I have posted on Piazza. This reading is fair game on an exam or quiz.


Pearson Dataset
Height of Son (in)
60
65
70
75
Height of Father (in)
60
65
70
75

Regression to the Mean

Skill and Luck / Genes and Random Environmental Factors

Unless $r_{xy} = 1$, There Is Regression to the Mean

$$\frac{\hat{y} - \bar{y}}{s_y} = r_{xy}\frac{x - \bar{x}}{s_x}$$

Least-squares Prediction $\hat{y}$ closer to $\bar{y}$ than x is to $\bar{x}$

You will derive the above formula in this week's homework.

Lecture #5 – Basic Probability I

Probability as Long-run Relative Frequency

Sets, Events and Axioms of Probability

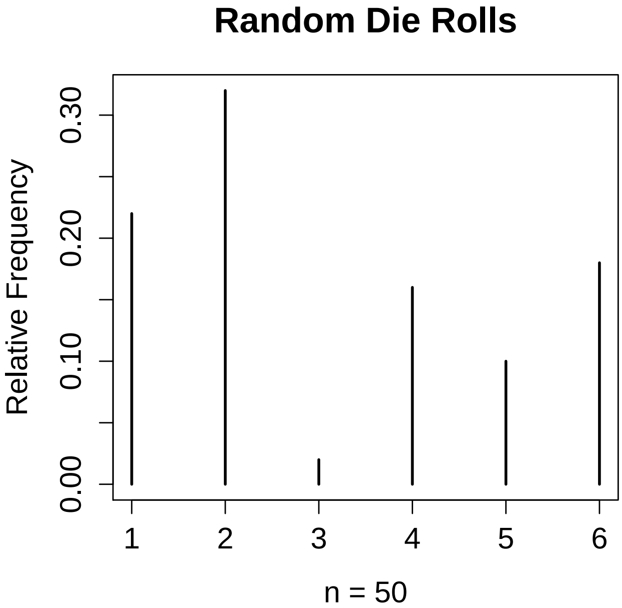
"Classical" Probability

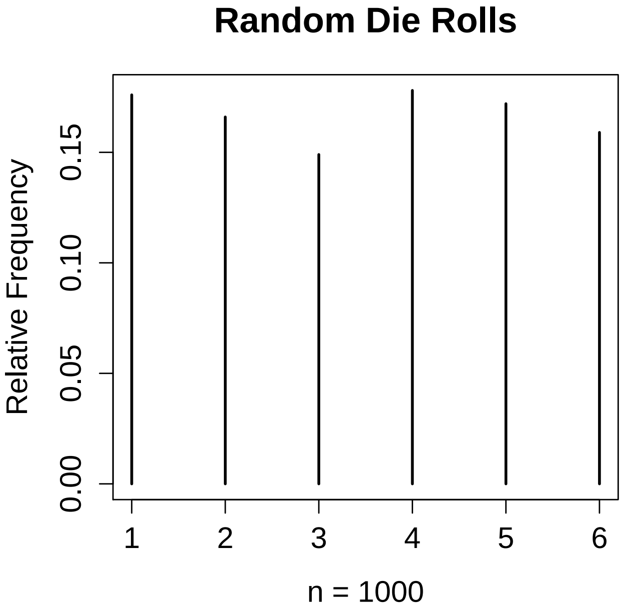
Our Definition of Probability for this Course

Probability = Long-run Relative Frequency

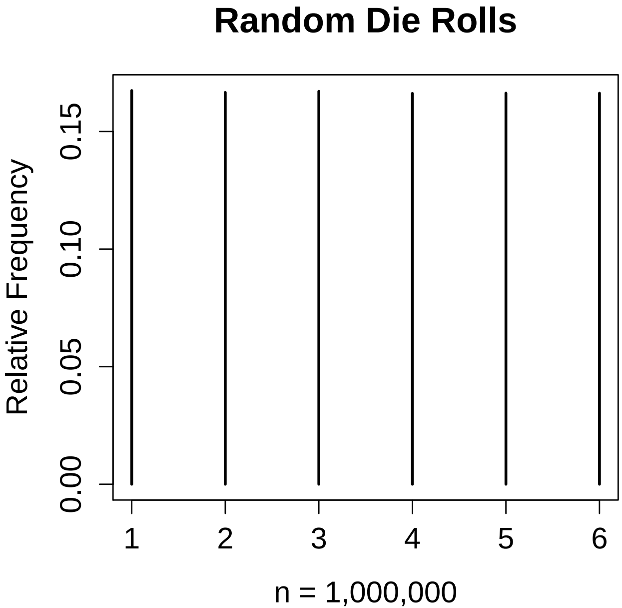
That is, relative frequencies settle down to probabilities if we carry out an experiment over, and over, and over...


Random Die Rolls
Relative Frequency
0.00
0.10
0.20
0.30
1
2
4
5
6
n = 10


Random Die Rolls
Relative Frequency
0.00
0.10
0.20
0.30
1
2
3
4
5
6
n = 50



Random Die Rolls
Relative Frequency
0.00
0.05
0.10
0.15
1
2
3
4
5
6
n = 1000


Random Die Rolls
Relative Frequency
0.00
0.05
0.10
0.15
1
2
3
4
5
6
n = 1,000,000

What do you think of this argument?

- The probability of flipping heads is $1/2$: if we flip a coin many times, about half of the time it will come up heads.
- The last ten throws in a row the coin has come up heads.
- The coin is bound to come up tails next time – it would be very rare to get 11 heads in a row.

(a) Agree

(b) Disagree

The Gambler's Fallacy

Relative frequencies settle down to probabilities, but this does not mean that the trials are dependent.

Dependent = "Memory" of Prev. Trials

Independent = No "Memory" of Prev. Trials

Terminology

Random Experiment

An experiment whose outcomes are random.

Basic Outcomes

Possible outcomes (mutually exclusive) of random experiment.

Sample Space: S

Set of all basic outcomes of a random experiment.

Event: E

A subset of the Sample Space (i.e. a collection of basic outcomes).

In set notation we write $E \subseteq S$.

Example

Random Experiment

Tossing a pair of dice.

Basic Outcome

An ordered pair (a, b) where $a, b \in \{1, 2, 3, 4, 5, 6\}$, e.g. $(2, 5)$

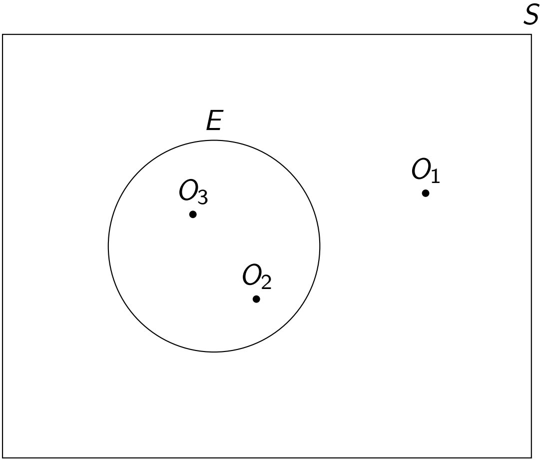
Sample Space: S

All ordered pairs (a, b) where $a, b \in \{1, 2, 3, 4, 5, 6\}$

Event: $E = \{\text{Sum of two dice is less than 4}\}$

$\{(1, 1), (1, 2), (2, 1)\}$

Visual Representation



Probability is Defined on *Sets*, and Events are Sets

Complement of an Event: $A^c =$ not A

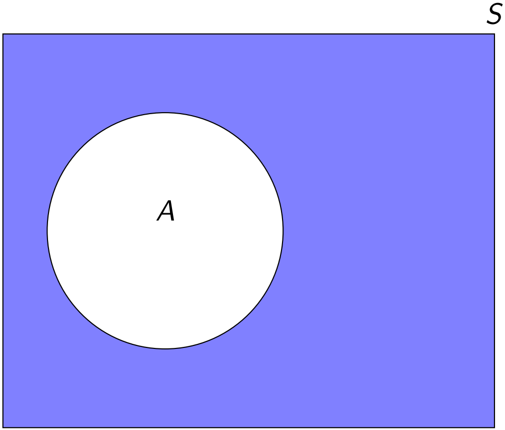


Figure: The complement A^c of an event $A \subseteq S$ is the collection of all basic outcomes from S not contained in A.

Intersection of Events: $A \cap B = A$ and B

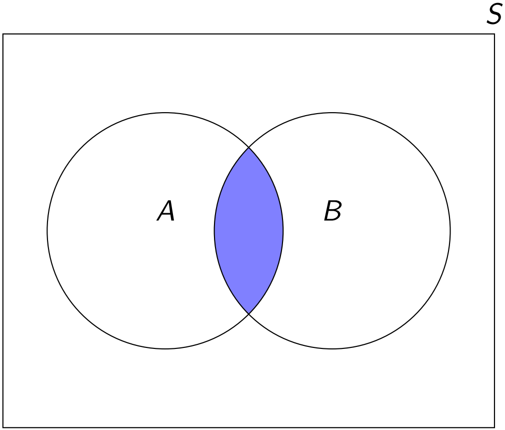


Figure: The intersection $A \cap B$ of two events $A, B \subseteq S$ is the collection of all basic outcomes from S contained in both A and B

Union of Events: $A \cup B = A$ or B

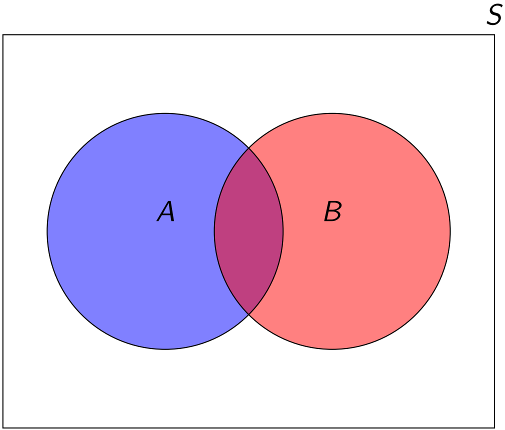


Figure: The union $A \cup B$ of two events $A, B \subseteq S$ is the collection of all basic outcomes from S contained in A, B or both.

Mutually Exclusive and Collectively Exhaustive

Mutually Exclusive Events

A collection of events $E_1, E_2, E_3, \ldots$ is *mutually exclusive* if the intersection $E_i \cap E_j$ of *any two different events* is empty.

Collectively Exhaustive Events

A collection of events $E_1, E_2, E_3, \ldots$ is *collectively exhaustive* if, taken together, they contain *all of the basic outcomes in S*.

Another way of saying this is that the union $E_1 \cup E_2 \cup E_3 \cup \cdots$ is S.

Implications

Mutually Exclusive Events

If one of the events occurs, then none of the others did.

Collectively Exhaustive Events

One of these events *must* occur.

Mutually Exclusive but *not Collectively Exhaustive*

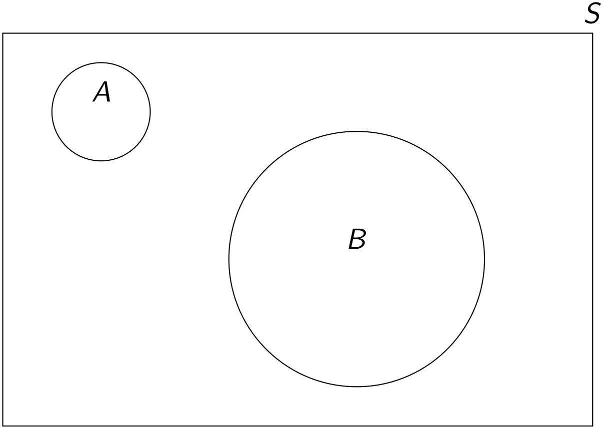


Figure: Although A and B don't overlap, they also don't cover S.

Collectively Exhaustive but *not Mutually Exclusive*

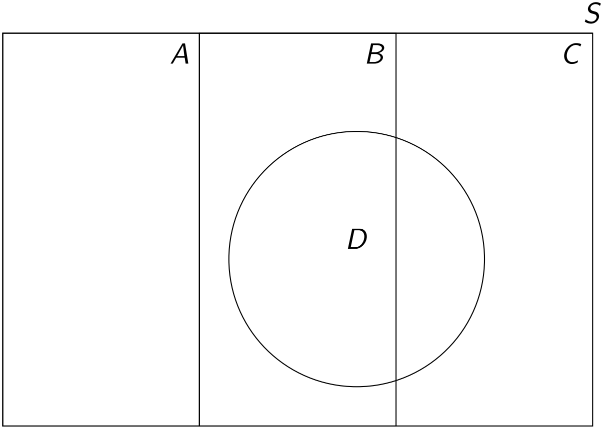


Figure: Together A, B, C and D cover S, but D overlaps with B and C.

Collectively Exhaustive *and* Mutually Exclusive

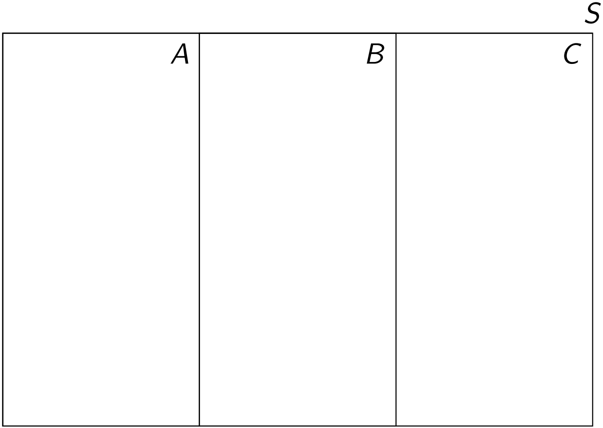


Figure: A, B, and C cover S and don't overlap.

Axioms of Probability

We assign every event A in the sample space S a real number $P(A)$ called the probability of A such that:

Axiom 1 $0 \leq P(A) \leq 1$

Axiom 2 $P(S) = 1$

Axiom 3 If $A_1, A_2, A_3, \dots$ are mutually exclusive events, then $P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \dots$

“Classical” Probability

When all of the basic outcomes are equally likely, calculating the probability of an event is simply a matter of counting – count up all the basic outcomes that make up the event, and divide by the total number of basic outcomes.

Recall from High School Math:

Multiplication Rule for Counting

n_1 ways to make first decision, n_2 ways to make second, $\ldots$, n_k ways to make kth $\Rightarrow n_1 \times n_2 \times \cdots \times n_k$ total ways to decide.

Corollary – Number of Possible Orderings

$k \times (k-1) \times (k-2) \times \cdots \times 2 \times 1 = k!$

Permutations – Order n people in k slots

$P_k^n = \frac{n!}{(n-k)!}$ (Order Matters)

Combinations – Choose committee of k from group of n

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$, where $0! = 1$ (Order Doesn't Matter)

Poker – Deal 5 Cards, Order Doesn't Matter

Basic Outcomes

$\binom{52}{5}$ possible hands

How Many Hands have Four Aces?

Poker – Deal 5 Cards, Order Doesn't Matter

Basic Outcomes

$\binom{52}{5}$ possible hands

How Many Hands have Four Aces?

48 (# of ways to choose the single card that is not an ace)

Probability of Getting Four Aces

$48/\binom{52}{5} \approx 0.00002$

Poker – Deal 5 Cards, Order Doesn't Matter

What is the probability of getting 4 of a kind?

- 13 ways to choose *which* card we have four of
- 48 ways to choose the last card in the hand
- $13 \times 48 = 624$

$$624/\binom{52}{5} \approx 0.00024$$

A Fairly Ridiculous Example

Roger Federer and Novak Djokovic have agreed to play in a tennis tournament against six Penn professors. Each player in the tournament is randomly allocated to one of the eight rungs in the ladder (next slide). Federer always beats Djokovic and, naturally, either of the two pros always beats any of the professors. What is the probability that Djokovic gets second place in the tournament?

Rung 1

Quarterfinalist 1

Rung 2

Semifinalist 1

Rung 3

Quarterfinalist 2

Rung 4

Winner

Rung 5

Quarterfinalist 3

Rung 6

Semifinalist 2

Rung 7

Quarterfinalist 4

Rung 8

Solution: Order Matters!

Denominator

8! basic outcomes – ways to arrange players on tournament ladder.

Numerator

Sequence of three decisions:

1. Which rung to put Federer on? (8 possibilities)
2. Which rung to put Djokovic on?
   - For any given rung that Federer is on, only 4 rungs prevent Djokovic from meeting him until the final.
3. How to arrange the professors? (6! ways)

$$\frac{8 \times 4 \times 6!}{8!} = \frac{8 \times 4}{7 \times 8} = 4/7 \approx 0.57$$

Lecture #6 – Basic Probability II

Complement Rule, Logical Consequence Rule, Addition Rule

Conditional Probability

Independence, Multiplication Rule

Law of Total Probability

Prediction Markets, "Dutch Book"

Recall: Axioms of Probability

Let S be the sample space. With each event $A \subseteq S$ we associate a real number $P(A)$ called the probability of A, satisfying the following conditions:

- Axiom 1 $0 \leq P(A) \leq 1$
- Axiom 2 $P(S) = 1$
- Axiom 3 If $A_1, A_2, A_3, \ldots$ are mutually exclusive events, then $P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \ldots$

The Complement Rule: $P(A^c) = 1 - P(A)$

Since A, A^c are mutually exclusive and collectively exhaustive:

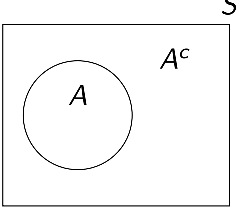


Figure: $A \cap A^c = \emptyset$, $A \cup A^c = S$

The Complement Rule: $P(A^c) = 1 - P(A)$

Since A, A^c are mutually exclusive and collectively exhaustive:

$$P(A \cup A^c) = P(A) + P(A^c) =$$

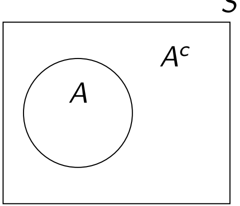


Figure: $A \cap A^c = \emptyset$, $A \cup A^c = S$

The Complement Rule: $P(A^c) = 1 - P(A)$

Since A, A^c are mutually exclusive and collectively exhaustive:

$$P(A \cup A^c) = P(A) + P(A^c) = P(S) = 1$$

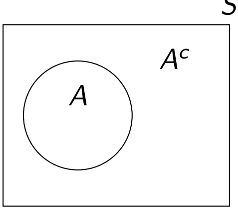


Figure: $A \cap A^c = \emptyset$, $A \cup A^c = S$

The Complement Rule: $P(A^c) = 1 - P(A)$

Since A, A^c are mutually exclusive and collectively exhaustive:

$$P(A \cup A^c) = P(A) + P(A^c) = P(S) = 1$$

Rearranging:

$$P(A^c) = 1 - P(A)$$

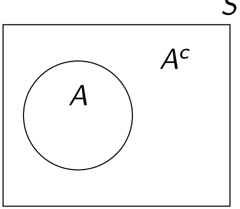


Figure: $A \cap A^c = \emptyset$, $A \cup A^c = S$

Another Important Rule – Equivalent Events

If A and B are Logically Equivalent, then $P(A) = P(B)$.

In other words, if A and B contain exactly the same basic outcomes, then $P(A) = P(B)$.

Although this seems obvious it's important to keep in mind...

The Logical Consequence Rule

If B Logically Entails A, then $P(B) \leq P(A)$

For example, the probability that someone comes from Texas cannot exceed the probability that she comes from the USA.

In Set Notation

$B \subseteq A \Rightarrow P(B) \leq P(A)$

Why is this so?

If $B \subseteq A$, then all the basic outcomes in B are also in A.

Proof of Logical Consequence Rule (Optional)

Proof won't be on a quiz or exam but is good practice with probability axioms.

Since $B \subseteq A$, we have $B = A \cap B$ and $A = B \cup (A \cap B^c)$. Combining these,

$$A = (A \cap B) \cup (A \cap B^c)$$

Now since $(A \cap B) \cap (A \cap B^c) = \emptyset$,

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) \\ &= P(B) + P(A \cap B^c) \\ &\geq P(B) \end{aligned}$$

because $0 \leq P(A \cap B^c) \leq 1$.

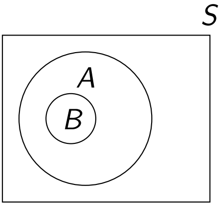


Figure:

$B = A \cap B$, and

$A = B \cup (A \cap B^c)$

“Odd Question” # 2

Pia is thirty-one years old, single, outspoken, and smart. She was a philosophy major. When a student, she was an ardent supporter of Native American rights, and she picketed a department store that had no facilities for nursing mothers. Rank the following statements in order from most probable to least probable.

(A) Pia is an active feminist.

(B) Pia is a bank teller.

(C) Pia works in a small bookstore.

(D) Pia is a bank teller and an active feminist.

(E) Pia is a bank teller and an active feminist who takes yoga classes.

(F) Pia works in a small bookstore and is an active feminist who takes yoga classes.

Using the Logical Consequence Rule...

(A) Pia is an active feminist.

(B) Pia is a bank teller.

(C) Pia works in a small bookstore.

(D) Pia is a bank teller and an active feminist.

(E) Pia is a bank teller and an active feminist who takes yoga classes.

(F) Pia works in a small bookstore and is an active feminist who takes yoga classes.

Any Correct Ranking Must Satisfy:

$$P(\mathrm{A}) \geq P(\mathrm{D}) \geq P(\mathrm{E})$$
$$P(\mathrm{B}) \geq P(\mathrm{D}) \geq P(\mathrm{E})$$
$$P(\mathrm{A}) \geq P(\mathrm{F})$$
$$P(\mathrm{C}) \geq P(\mathrm{F})$$

Throw a Fair Die Once

$E =$ roll an even number

What are the basic outcomes?

$\{1, 2, 3, 4, 5, 6\}$

Throw a Fair Die Once

$E =$ roll an even number

What are the basic outcomes?

$\{1, 2, 3, 4, 5, 6\}$

What is $P(E)$?

Throw a Fair Die Once

$E =$ roll an even number

What are the basic outcomes?

$\{1, 2, 3, 4, 5, 6\}$

What is $P(E)$?

$E = \{2, 4, 6\}$ and the basic outcomes are equally likely (and mutually exclusive), so

$$P(E) = 1/6 + 1/6 + 1/6 = 3/6 = 1/2$$

Throw a Fair Die Once

$E =$ roll an even number $\qquad M =$ roll a 1 or a prime number

What is $P(E \cup M)$?

Throw a Fair Die Once

$E =$ roll an even number $\qquad M =$ roll a 1 or a prime number

What is $P(E \cup M)$?

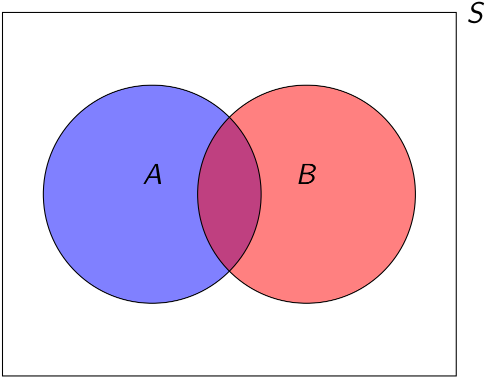
Key point: E and M are not mutually exclusive!

$$\begin{aligned} P(E \cup M) &= P(\{1,2,3,4,5,6\}) = 1 \\ P(E) &= P(\{2,4,6\}) = 1/2 \\ P(M) &= P(\{1,2,3,5\}) = 4/6 = 2/3 \end{aligned}$$

$$P(E) + P(M) = 1/2 + 2/3 = 7/6 \neq P(E \cup M) = 1$$

The Addition Rule – Don't Double-Count!

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Construct a formal proof as an optional homework problem.

Who's on the other side?

Three Cards, Each with a Face on the Front and Back

1. Gaga/Gaga
2. Obama/Gaga
3. Obama/Obama

I draw a card at random and look at one side: it's Obama.

What is the probability that the other side is also Obama?

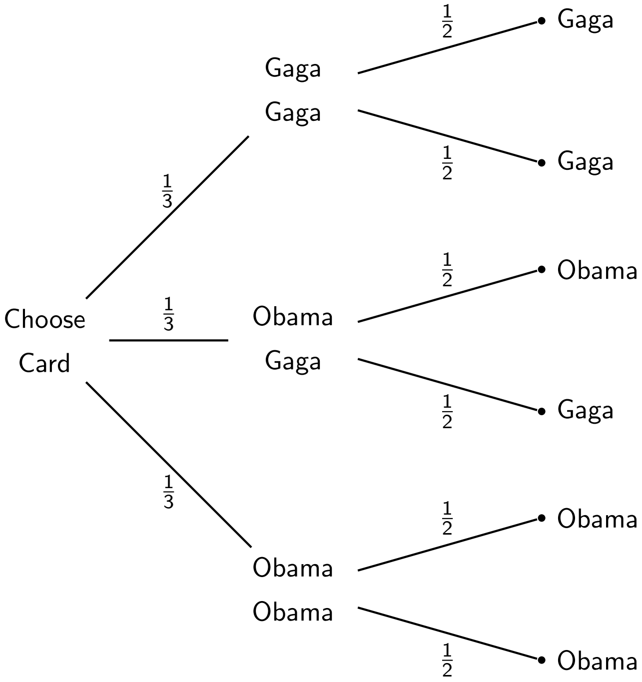
Let's Try The Method of Monte Carlo...

When you don't know how to calculate, simulate.

Procedure

1. Close your eyes and thoroughly shuffle your cards.
2. Keeping eyes closed, draw a card and place it on your desk.
3. Stand if Obama is face-up on your chosen card.
4. We'll count those standing and call the total N
5. Of those standing, sit down if Obama is *not* on the back of your chosen card.
6. We'll count those *still* standing and call the total m.

Monte Carlo Approximation of Desired Probability $= \frac{m}{N}$


Choose
Card
1/3
1/3
1/3
Gaga
Gaga
Obama
Gaga
Obama
Obama
1/2
1/2
1/2
1/2
1/2
1/2
Gaga
Gaga
Obama
Gaga
Obama
Obama

Conditional Probability – Reduced Sample Space

Set of relevant outcomes restricted by condition

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ provided } P(B) > 0$$

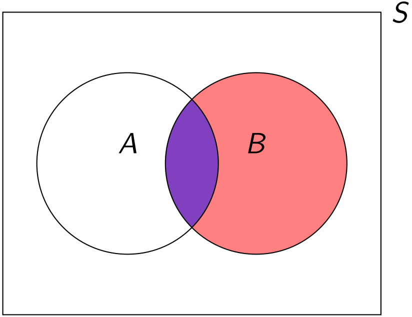


Figure: B becomes the "new sample space" so we need to re-scale by $P(B)$ to keep probabilities between zero and one.

Who's on the other side?

Let F be the event that Obama is on the front of the card of the card we draw and B be the event that he is on the back.

$$P(B|F) = \frac{P(B \cap F)}{P(F)} =$$

Who's on the other side?

Let F be the event that Obama is on the front of the card of the card we draw and B be the event that he is on the back.

$$P(B|F) = \frac{P(B \cap F)}{P(F)} = \frac{1/3}{1/2} =$$

Who's on the other side?

Let F be the event that Obama is on the front of the card of the card we draw and B be the event that he is on the back.

$$P(B|F) = \frac{P(B \cap F)}{P(F)} = \frac{1/3}{1/2} = 2/3$$

Conditional Versions of Probability Axioms

1. $0 \leq P(A|B) \leq 1$
2. $P(B|B) = 1$
3. If $A_1, A_2, A_3, \ldots$ are mutually exclusive given B, then $P(A_1 \cup A_2 \cup A_3 \cup \cdots |B) = P(A_1|B) + P(A_2|B) + P(A_3|B) \ldots$

Conditional Versions of Other Probability Rules

- $P(A|B) = 1 - P(A^c|B)$
- A_1 logically equivalent to $A_2 \iff P(A_1|B) = P(A_2|B)$
- $A_1 \subseteq A_2 \implies P(A_1|B) \leq P(A_2|B)$
- $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$

However: $P(A|B) \neq P(B|A)$ and $P(A|B^c) \neq 1 - P(A|B)$!

Independence and The Multiplication Rule

The Multiplication Rule

Rearrange the definition of conditional probability:

$P(A \cap B) = P(A|B)P(B)$

Independence and The Multiplication Rule

The Multiplication Rule

Rearrange the definition of conditional probability:

$P(A \cap B) = P(A|B)P(B)$

Statistical Independence

$P(A \cap B) = P(A)P(B)$

Independence and The Multiplication Rule

The Multiplication Rule

Rearrange the definition of conditional probability:

$$P(A \cap B) = P(A|B)P(B)$$

Statistical Independence

$$P(A \cap B) = P(A)P(B)$$

By the Multiplication Rule

Independence $\iff P(A|B) = P(A)$

Independence and The Multiplication Rule

The Multiplication Rule

Rearrange the definition of conditional probability:

$P(A \cap B) = P(A|B)P(B)$

Statistical Independence

$P(A \cap B) = P(A)P(B)$

By the Multiplication Rule

Independence $\iff P(A|B) = P(A)$

Interpreting Independence

Knowledge that B has occurred tells nothing about whether A will.

Will Having 5 Children Guarantee a Boy?

A couple plans to have five children. Assuming that each birth is independent and male and female children are equally likely, what is the probability that they have at least one boy?

Will Having 5 Children Guarantee a Boy?

A couple plans to have five children. Assuming that each birth is independent and male and female children are equally likely, what is the probability that they have at least one boy?

By Independence and the Complement Rule,

$$\begin{aligned} P(\text{no boys}) &= P(\text{5 girls}) \\ &= 1/2 \times 1/2 \times 1/2 \times 1/2 \times 1/2 \\ &= 1/32 \end{aligned}$$

Will Having 5 Children Guarantee a Boy?

A couple plans to have five children. Assuming that each birth is independent and male and female children are equally likely, what is the probability that they have at least one boy?

By Independence and the Complement Rule,

$$\begin{aligned} P(\text{no boys}) &= P(\text{5 girls}) \\ &= 1/2 \times 1/2 \times 1/2 \times 1/2 \times 1/2 \\ &= 1/32 \end{aligned}$$

$$\begin{aligned} P(\text{at least 1 boy}) &= 1 - P(\text{no boys}) \\ &= 1 - 1/32 = 31/32 = 0.97 \end{aligned}$$

The Law of Total Probability

If $E_1, E_2, \ldots, E_k$ are mutually exclusive, collectively exhaustive events and A is another event, then

$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \ldots + P(A|E_k)P(E_k)$$

Example of Law of Total Probability

Define the following events:

$$\begin{aligned} F &= \text{Obama on front of card} \\ A &= \text{Draw card with two Gagas} \\ B &= \text{Draw card with two Obamas} \\ C &= \text{Draw card with BOTH Obama and Gaga} \end{aligned}$$

$$P(F) = P(F|A)P(A) + P(F|B)P(B) + P(F|C)P(C)$$

Example of Law of Total Probability

Define the following events:

$$\begin{aligned} F &= \text{Obama on front of card} \\ A &= \text{Draw card with two Gagas} \\ B &= \text{Draw card with two Obamas} \\ C &= \text{Draw card with BOTH Obama and Gaga} \end{aligned}$$

$$\begin{aligned} P(F) &= P(F|A)P(A) + P(F|B)P(B) + P(F|C)P(C) \\ &= 0 \times 1/3 + 1 \times 1/3 + 1/2 \times 1/3 \\ &= 1/2 \end{aligned}$$

Deriving the Law of Total Probability For $k = 2$

Optional proof: You are not responsible for this proof on quizzes and exams.

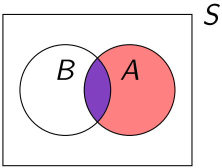


Figure:

$A = (A \cap B) \cup (A \cap B^c)$,

$(A \cap B) \cap (A \cap B^c) = \emptyset$

Deriving the Law of Total Probability For $k = 2$

Optional proof: You are not responsible for this proof on quizzes and exams.

Since $A \cap B$ and $A \cap B^c$ are mutually exclusive and their union equals A,

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

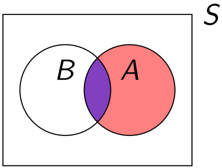


Figure:
$A = (A \cap B) \cup (A \cap B^c)$,
$(A \cap B) \cap (A \cap B^c) = \emptyset$

Deriving the Law of Total Probability For $k = 2$

Optional proof: You are not responsible for this proof on quizzes and exams.

Since $A \cap B$ and $A \cap B^c$ are mutually exclusive and their union equals A,

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

But by the multiplication rule:

$$\begin{aligned} P(A \cap B) &= P(A|B)P(B) \\ P(A \cap B^c) &= P(A|B^c)P(B^c) \end{aligned}$$

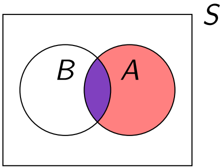


Figure: $A = (A \cap B) \cup (A \cap B^c)$, $(A \cap B) \cap (A \cap B^c) = \emptyset$

Deriving the Law of Total Probability For $k = 2$

Optional proof: You are not responsible for this proof on quizzes and exams.

Since $A \cap B$ and $A \cap B^c$ are mutually exclusive and their union equals A,

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

But by the multiplication rule:

$$\begin{aligned} P(A \cap B) &= P(A|B)P(B) \\ P(A \cap B^c) &= P(A|B^c)P(B^c) \end{aligned}$$

Combining,

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

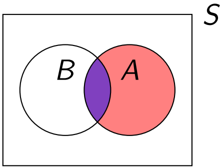


Figure:
$A = (A \cap B) \cup (A \cap B^c)$,
$(A \cap B) \cap (A \cap B^c) = \emptyset$

How do prediction markets work?

To learn more, see Wolfers & Zitzewitz (2004)

> THIS CERTIFICATE ENTITLES THE BEARER TO $1 IF THE PATRIOTS WIN THE 2016-2017 SUPERBOWL.

Buyers – Purchase Right to Collect

Patriots very likely to win $\Rightarrow$ buy for close to \$1.

Patriots very unlikely to win $\Rightarrow$ buy for close to \$0.

Sellers – Sell Obligation to Pay

Patriots very likely to win $\Rightarrow$ sell for close to \$1.

Patriots very unlikely to win $\Rightarrow$ sell for close to \$0.

Probabilities from Beliefs

Market price of contract encodes market participants' beliefs in the form of probability:

$$\text{Price/Payout} \approx \text{Subjective Probability}$$

Probabilities from Beliefs

Market price of contract encodes market participants' beliefs in the form of probability:

$$\text{Price/Payout} \approx \text{Subjective Probability}$$

"Dutch Book"

If the probabilities implied by prediction market prices violate *any* of our probability rules there is a *pure arbitrage opportunity*: a way make to make a guaranteed, risk-free profit.

Probabilities from Beliefs

Market price of contract encodes market participants' beliefs in the form of probability:

$$\text{Price/Payout} \approx \text{Subjective Probability}$$

"Dutch Book"

If the probabilities implied by prediction market prices violate *any* of our probability rules there is a *pure arbitrage opportunity*: a way make to make a guaranteed, risk-free profit.

A Real-world Dutch Book

Courtesy of Eric Crampton

November 5th, 2012

- \$2.30 for contract paying \$10 if Romney wins on BetFair
- \$6.58 for contract paying \$10 if Obama wins on InTrade

Implied Probabilities

- BetFair: $P(Romney) \approx 0.23$
- InTrade: $P(Obama) \approx 0.66$

What's Wrong with This?

A Real-world Dutch Book

Courtesy of $\boxed{\text{Eric Crampton}}$

November 5th, 2012

- \$2.30 for contract paying \$10 if Romney wins on BetFair
- \$6.58 for contract paying \$10 if Obama wins on InTrade

Implied Probabilities

- BetFair: $P(Romney) \approx 0.23$
- InTrade: $P(Obama) \approx 0.66$

What's Wrong with This?

Violates complement rule! $P(Obama) = 1 - P(Romney)$ but the implied probabilities here don't sum up to one!

A Real-world Dutch Book

Courtesy of Eric Crampton

November 5th, 2012

- \$2.30 for contract paying \$10 if Romney wins on BetFair
- \$6.58 for contract paying \$10 if Obama wins on InTrade

Arbitrage Strategy

Buy Equal Numbers of Each

- Cost = \$2.30 + \$6.58 =\$8.88 per pair
- Payout if Romney Wins: \$10
- Payout if Obama Wins: \$10
- Guaranteed Profit: \$10 - \$8.88 = \$1.12 per pair

Lecture #7 – Basic Probability III / Discrete RVs I

Bayes' Rule and the Base Rate Fallacy

Overview of Random Variables

Probability Mass Functions

Four Volunteers Please!

The Lie Detector Problem

From accounting records, we know that 10% of employees in the store are stealing merchandise.

The managers want to fire the thieves, but their only tool in distinguishing is a lie detector test that is 80% accurate:

$$\begin{aligned} \text{Innocent} \quad &\Rightarrow \quad \text{Pass test with 80\% Probability} \\ \text{Thief} \quad &\Rightarrow \quad \text{Fail test with 80\% Probability} \end{aligned}$$

The Lie Detector Problem

From accounting records, we know that 10% of employees in the store are stealing merchandise.

The managers want to fire the thieves, but their only tool in distinguishing is a lie detector test that is 80% accurate:

$$\begin{aligned} \text{Innocent} \quad &\Rightarrow \quad \text{Pass test with 80\% Probability} \\ \text{Thief} \quad &\Rightarrow \quad \text{Fail test with 80\% Probability} \end{aligned}$$

What is the probability that someone is a thief *given* that she has failed the lie detector test?

Monte Carlo Simulation – Roll a 10-sided Die Twice

Managers will split up and visit employees. Employees roll the die twice but keep the results secret!

First Roll – Thief or not?

$0 \Rightarrow$ Thief, $1 - 9 \Rightarrow$ Innocent

Second Roll – Lie Detector Test

$0, 1 \Rightarrow$ Incorrect Test Result, $2 - 9$ Correct Test Result

	0 or 1	2–9
Thief	Pass	Fail
Innocent	Fail	Pass

What percentage of those who failed the test are guilty?

\# Who Failed Lie Detector Test:

\# Of Thieves Among Those Who Failed:

Base Rate Fallacy – Failure to Consider Prior Information

Base Rate – Prior Information

Before the test we know that 10% of Employees are stealing.

People tend to focus on the fact that the test is 80% accurate and ignore the fact that only 10% of the employees are theives.

Thief (Y/N), Lie Detector (P/F)

	0	1	2	3	4	5	6	7	8	9
0	YP	YP	YF	YF	YF	YF	YF	YF	YF	YF
1	NF	NF	NP	NP	NP	NP	NP	NP	NP	NP
2	NF	NF	NP	NP	NP	NP	NP	NP	NP	NP
3	NF	NF	NP	NP	NP	NP	NP	NP	NP	NP
4	NF	NF	NP	NP	NP	NP	NP	NP	NP	NP
5	NF	NF	NP	NP	NP	NP	NP	NP	NP	NP
6	NF	NF	NP	NP	NP	NP	NP	NP	NP	NP
7	NF	NF	NP	NP	NP	NP	NP	NP	NP	NP
8	NF	NF	NP	NP	NP	NP	NP	NP	NP	NP
9	NF	NF	NP	NP	NP	NP	NP	NP	NP	NP

Table: Each outcome in the table is equally likely. The 26 given in red correspond to failing the test, but only 8 of these (YF) correspond to being a thief.

Base Rate of Thievery is 10%

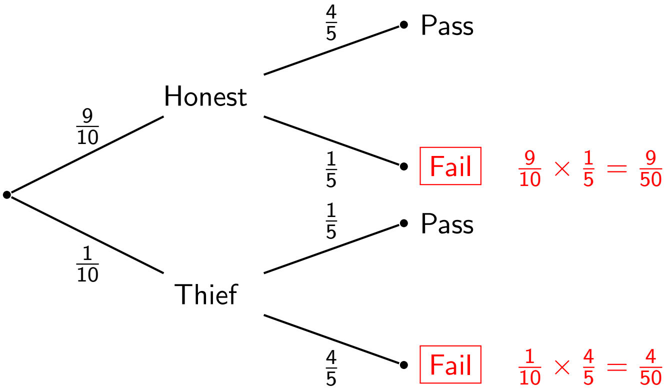


Figure: Although $\frac{9}{50} + \frac{4}{50} = \frac{13}{50}$ fail the test, only $\frac{4/50}{13/50} = \frac{4}{13} \approx 0.31$ are actually theives!

Deriving Bayes' Rule

Intersection is symmetric: $A \cap B = B \cap A$ so $P(A \cap B) = P(B \cap A)$

Deriving Bayes' Rule

Intersection is symmetric: $A \cap B = B \cap A$ so $P(A \cap B) = P(B \cap A)$

By the definition of conditional probability,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Deriving Bayes' Rule

Intersection is symmetric: $A \cap B = B \cap A$ so $P(A \cap B) = P(B \cap A)$

By the definition of conditional probability,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

And by the multiplication rule:

$$P(B \cap A) = P(B|A)P(A)$$

Deriving Bayes' Rule

Intersection is symmetric: $A \cap B = B \cap A$ so $P(A \cap B) = P(B \cap A)$

By the definition of conditional probability,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

And by the multiplication rule:

$$P(B \cap A) = P(B|A)P(A)$$

Finally, combining these

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Understanding Bayes' Rule

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

Reversing the Conditioning

Express $P(A|B)$ in terms of $P(B|A)$. *Relative magnitudes* of the two conditional probabilities determined by the ratio $P(A)/P(B)$.

Base Rate

$P(A)$ is called the "base rate" or the "prior probability."

Denominator

Typically, we calculate $P(B)$ using the law of toal probability

In General $P(A|B) \neq P(B|A)$

Question

Most college students are Democrats. Does it follow that most Democrats are college students? (A = YES, B = NO)

In General $P(A|B) \neq P(B|A)$

Question

Most college students are Democrats. Does it follow that most Democrats are college students? (A = YES, B = NO)

Answer

There are many more Democracts than college students:

$$P(\text{Dem}) > P(\text{Student})$$

so $P(\text{Student}|\text{Dem})$ is small even though $P(\text{Dem}|\text{Student})$ is large.

Solving the Lie Detector Problem with Bayes' Rule

$\boxed{T = \text{Employee is a Thief}, \; F = \text{Employee Fails Lie Detector Test}}$

$$P(T|F) = \frac{P(F|T)P(T)}{P(F)}$$

Solving the Lie Detector Problem with Bayes' Rule

$\boxed{T = \text{Employee is a Thief}, \; F = \text{Employee Fails Lie Detector Test}}$

$$P(T|F) = \frac{P(F|T)P(T)}{P(F)}$$

$$P(F) = P(F|T)P(T) + P(F|T^c)P(T^c)$$

Solving the Lie Detector Problem with Bayes' Rule

$\boxed{T = \text{Employee is a Thief, } F = \text{Employee Fails Lie Detector Test}}$

$$P(T|F) = \frac{P(F|T)P(T)}{P(F)}$$

$$\begin{aligned} P(F) &= P(F|T)P(T) + P(F|T^c)P(T^c) \\ &= 0.8 \times 0.1 + 0.2 \times 0.9 \end{aligned}$$

Solving the Lie Detector Problem with Bayes' Rule

$\boxed{T = \text{Employee is a Thief},\ F = \text{Employee Fails Lie Detector Test}}$

$$P(T|F) = \frac{P(F|T)P(T)}{P(F)}$$

$$\begin{aligned} P(F) &= P(F|T)P(T) + P(F|T^c)P(T^c) \\ &= 0.8 \times 0.1 + 0.2 \times 0.9 \\ &= 0.08 + 0.18 = 0.26 \end{aligned}$$

Solving the Lie Detector Problem with Bayes' Rule

$\boxed{T = \text{Employee is a Thief, } F = \text{Employee Fails Lie Detector Test}}$

$$P(T|F) = \frac{P(F|T)P(T)}{P(F)}$$

$$\begin{aligned} P(F) &= P(F|T)P(T) + P(F|T^c)P(T^c) \\ &= 0.8 \times 0.1 + 0.2 \times 0.9 \\ &= 0.08 + 0.18 = 0.26 \end{aligned}$$

$$P(T|F) = \frac{0.08}{0.26} =$$

Solving the Lie Detector Problem with Bayes' Rule

$\boxed{T = \text{Employee is a Thief, } F = \text{Employee Fails Lie Detector Test}}$

$$P(T|F) = \frac{P(F|T)P(T)}{P(F)}$$

$$\begin{aligned} P(F) &= P(F|T)P(T) + P(F|T^c)P(T^c) \\ &= 0.8 \times 0.1 + 0.2 \times 0.9 \\ &= 0.08 + 0.18 = 0.26 \end{aligned}$$

$$P(T|F) = \frac{0.08}{0.26} = \frac{8}{26} = \frac{4}{13} \approx 0.31$$

Random Variables

Random Variables

A random variable is neither random nor a variable.

Random Variable (RV): X

A *fixed* function that assigns a *number* to each basic outcome of a random experiment.

Realization: x

A particular numeric value that an RV could take on. We write $\{X = x\}$ to refer to the *event* that the RV X took on the value x.

Support Set (aka Support)

The set of all possible realizations of a RV.

Random Variables (continued)

Notation

Capital latin letters for RVs, e.g. X, Y, Z, and the corresponsing lowercase letters for their realizations, e.g. x, y, z.

Intuition

You can think of an RV as a machine that spits out random numbers: although the machine is deterministic, its inputs, the outcomes of a random experiment, are not.

Example: Coin Flip Random Variable

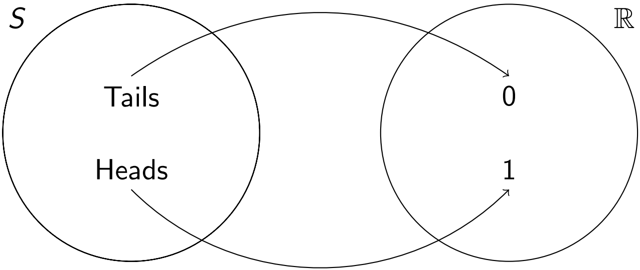


Figure: This random variable assigns numeric values to the random experiment of flipping a fair coin once: Heads is assigned 1 and Tails 0.

Which of these is a realization of the Coin Flip RV?

(a) Tails

(b) 2

(c) 0

(d) Heads

(e) $1/2$

What is the support set of the Coin Flip RV?

(a) $\{\text{Heads, Tails}\}$

(b) $1/2$

(c) 0

(d) $\{0, 1\}$

(e) 1

Let X denote the Coin Flip RV

What is $P(X = 1)$?

(a) 0

(b) 1

(c) 1/2

(d) Not enough information to determine

Two Kinds of RVs: Discrete and Continuous

Discrete: support set is discrete, e.g. $\{0, 1, 2\}$, $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$

Continuous: support set is continuous, e.g. $[-1, 1]$, $\mathbb{R}$.

Start with the discrete case since it's easier, but most of the ideas we learn will carry over to the continuous case.

Discrete Random Variables I

Probability Mass Function (pmf)

A function that gives $P(X = x)$ for any realization x in the support set of a discrete RV X. We use the following notation for the pmf:

$$p(x) = P(X = x)$$

Plug in a realization x, get out a probability $p(x)$.

Probability Mass Function for Coin Flip RV

$$X = \begin{cases} 0, \text{Tails} \\ 1, \text{Heads} \end{cases}$$

$$\begin{aligned} p(0) &= 1/2 \\ p(1) &= 1/2 \end{aligned}$$

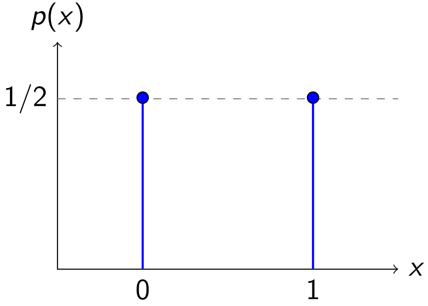


Figure: Plot of pmf for Coin Flip Random Variable

Important Note about Support Sets

Whenever you write down the pmf of a RV, it is crucial to also write down its Support Set. Recall that this is the set of *all possible realizations for a RV*. Outside of the support set, all probabilities are zero. In other words, the pmf is only defined on the support.

Properties of Probability Mass Functions

If $p(x)$ is the pmf of a random variable X, then

(i) $0 \leq p(x) \leq 1$ for all x

(ii) $\sum_{\text{all } x} p(x) = 1$

where "all x" is shorthand for "all x in the support of X."

Lecture #8 – Discrete RVs II

Cumulative Distribution Functions (CDFs)

The Bernoulli Random Variable

Definition of Expected Value

Expected Value of a Function

Linearity of Expectation

Recall: Properties of Probability Mass Functions

If $p(x)$ is the pmf of a random variable X, then

(i) $0 \leq p(x) \leq 1$ for all x

(ii) $\displaystyle\sum_{\text{all } x} p(x) = 1$

where "all x" is shorthand for "all x in the support of X."

Cumulative Distribution Function (CDF)

This Def. is the same for continuous RVs.

The CDF gives the probability that a RV X does not exceed a specified threshold x_0, as a function of x_0

$$F(x_0) = P(X \leq x_0)$$

Important!

The threshold x_0 is allowed to be *any real number*. In particular, it doesn't have to be in the support of X!

Discrete RVs: Sum the pmf to get the CDF

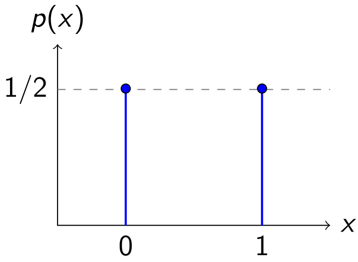
$$\boxed{F(x_0) = \sum_{x \leq x_0} p(x)}$$

Why?

The events $\{X = x\}$ are mutually exclusive, so we sum to get the probability of their union for all $x \leq x_0$:

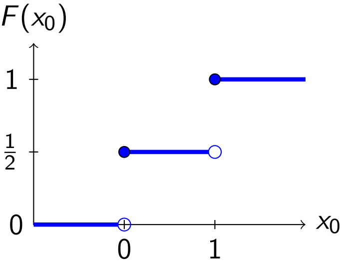
$$F(x_0) = P(X \leq x_0) = P\left(\bigcup_{x \leq x_0} \{X = x\}\right) = \sum_{x \leq x_0} P(X = x) = \sum_{x \leq x_0} p(x)$$

Probability Mass Function



$$\begin{aligned} p(0) &= 1/2 \\ p(1) &= 1/2 \end{aligned}$$

Cumulative Dist. Function



$$F(x_0) = \begin{cases} 0, & x_0 < 0 \\ \frac{1}{2}, & 0 \leq x_0 < 1 \\ 1, & x_0 \geq 1 \end{cases}$$

Properties of CDFs

These are also true for continuous RVs.

1. $\lim_{x_0 \to \infty} F(x_0) = 1$
2. $\lim_{x_0 \to -\infty} F(x_0) = 0$
3. Non-decreasing: $x_0 < x_1 \Rightarrow F(x_0) \leq F(x_1)$
4. Right-continuous ("open" versus "closed" on prev. slide)

Since $F(x_0) = P(X \leq x_0)$, we have $0 \leq F(x_0) \leq 1$ for all x_0

Bernoulli Random Variable – Generalization of Coin Flip

Support Set

$\{0, 1\}$ – 1 traditionally called "success," 0 "failure"

Probability Mass Function

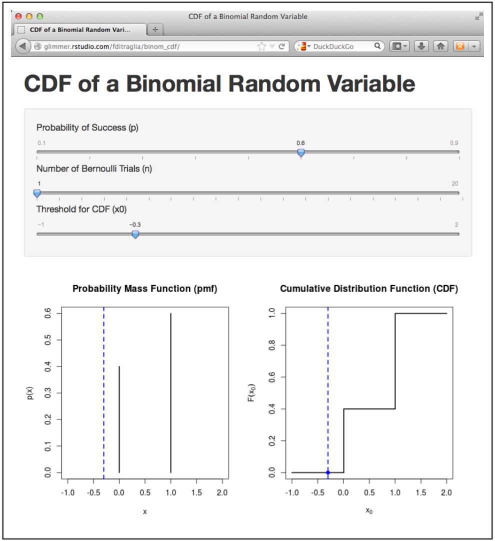
$$
\begin{aligned}
p(0) &= 1 - p \\
p(1) &= p
\end{aligned}
$$

Cumulative Distribution Function

$$
F(x_0) = \begin{cases} 0, & x_0 < 0 \\ 1 - p, & 0 \leq x_0 < 1 \\ 1, & x_0 \geq 1 \end{cases}
$$

http://fditraglia.shinyapps.io/binom_cdf/

Set the second slider to 1 and play around with the others.



Average Winnings Per Trial

If the realizations of the coin-flip RV were payoffs, how much would you expect to win per play *on average* in a long sequence of plays?

$$X = \begin{cases} \$0, \text{Tails} \\ \$1, \text{Heads} \end{cases}$$

Expected Value (aka Expectation)

The expected value of a discrete RV X is given by

$$E[X] = \sum_{\text{all } x} x \cdot p(x)$$

In other words, the expected value of a discrete RV is the *probability-weighted average of its realizations*.

Notation

We sometimes write μ as shorthand for $E[X]$.

Expected Value of Bernoulli RV

$$X = \begin{cases} 0, \text{Failure: } 1-p \\ 1, \text{Success: } p \end{cases}$$

$$\sum_{\text{all } x} x \cdot p(x) = 0 \cdot (1-p) + 1 \cdot p = p$$

Your Turn to Calculate an Expected Value

Let X be a random variable with support set $\{1, 2, 3\}$ where $p(1) = p(2) = 1/3$. Calculate $E[X]$.

Your Turn to Calculate an Expected Value

Let X be a random variable with support set $\{1, 2, 3\}$ where $p(1) = p(2) = 1/3$. Calculate $E[X]$.

$$E[X] = \sum_{\text{all } x} x \cdot p(x) = 1 \times 1/3 + 2 \times 1/3 + 3 \times 1/3 = 2$$

Random Variables and Parameters

Notation: $X \sim \text{Bernoulli}(p)$

Means X is a Bernoulli RV with $P(X = 1) = p$ and $P(X = 0) = 1 - p$. The tilde is read "distributes as."

Parameter

Any constant that appears in the definition of a RV, here p.

Constants Versus Random Variables

This is a crucial distinction that students sometimes miss:

Random Variables

- Suppose X is a RV – the values it takes on are random
- A function $g(X)$ of a RV is itself a RV as we'll learn today.

Constants

- $E[X]$ is a constant (you should convince yourself of this)
- Realizations x are constants. What is random is *which* realization the RV takes on.
- Parameters are constants (e.g. p for Bernoulli RV)
- Sample size n is a constant

The St. Petersburg Game

How Much Would You Pay?

How much would you be willing to pay for the right to play the following game?

> *Imagine a fair coin. The coin is tossed once. If it falls heads, you receive a prize of \$2 and the game stops. If not, it is tossed again. If it falls heads on the second toss, you get \$4 and the game stops. If not, it is tossed again. If it falls heads on the third toss, you get \$8 and the game stops, and so on. The game stops after the first head is thrown. If the first head is thrown on the* x^{th} *toss, the prize is* \$$2^x$

$X =$ Trial Number of First Head

x	2^x	$p(x)$	$2^x \cdot p(x)$

$$E[Y] = \sum_{\text{all } x} 2^x \cdot p(x) =$$

X = Trial Number of First Head

x	2^x	$p(x)$	$2^x \cdot p(x)$
1	2	$1/2$	1
2	4	$1/4$	1
3	8	$1/8$	1

$$E[Y] = \sum_{\text{all } x} 2^x \cdot p(x) =$$

X = Trial Number of First Head

x	2^x	$p(x)$	$2^x \cdot p(x)$
1	2	$1/2$	1
2	4	$1/4$	1
3	8	$1/8$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	2^n	$1/2^n$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$

$$E[Y] = \sum_{\text{all } x} 2^x \cdot p(x) =$$

X = Trial Number of First Head

x	2^x	$p(x)$	$2^x \cdot p(x)$
1	2	$1/2$	1
2	4	$1/4$	1
3	8	$1/8$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	2^n	$1/2^n$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$

$$E[Y] = \sum_{\text{all } x} 2^x \cdot p(x) = 1 + 1 + 1 + \dots$$

X = Trial Number of First Head

x	2^x	$p(x)$	$2^x \cdot p(x)$
1	2	$1/2$	1
2	4	$1/4$	1
3	8	$1/8$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	2^n	$1/2^n$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$

$$E[Y] = \sum_{\text{all } x} 2^x \cdot p(x) = 1 + 1 + 1 + \ldots = \infty$$

Functions of Random Variables are Themselves Random Variables

Example: $X \sim \text{Bernoulli}(p)$, $Y = (X+1)^2$

Support Set for Y

Example: $X \sim \text{Bernoulli}(p)$, $Y = (X+1)^2$

Support Set for Y

$\{(0+1)^2, (1+1)^2\} = \{1, 4\}$

Probability Mass Function for Y

Example: $X \sim \text{Bernoulli}(p)$, $Y = (X+1)^2$

Support Set for Y

$\{(0+1)^2, (1+1)^2\} = \{1, 4\}$

Probability Mass Function for Y

$$p_Y(y) = \begin{cases} 1-p & y = 1 \\ p & y = 4 \\ 0 & \text{otherwise} \end{cases}$$

Example: $X \sim \text{Bernoulli}(p)$, $Y = (X+1)^2$

Support Set for Y

$\{(0+1)^2, (1+1)^2\} = \{1, 4\}$

Probability Mass Function for Y

$$p_Y(y) = \begin{cases} 1-p & y = 1 \\ p & y = 4 \\ 0 & \text{otherwise} \end{cases}$$

Expected Value of Y

$$\sum_{y \in \{1,4\}} y \times p_Y(y)$$

Example: $X \sim \text{Bernoulli}(p)$, $Y = (X+1)^2$

Support Set for Y

$\{(0+1)^2, (1+1)^2\} = \{1, 4\}$

Probability Mass Function for Y

$$p_Y(y) = \begin{cases} 1-p & y = 1 \\ p & y = 4 \\ 0 & \text{otherwise} \end{cases}$$

Expected Value of Y

$$\sum_{y \in \{1,4\}} y \times p_Y(y) = 1 \times (1-p) + 4 \times p$$

Example: $X \sim \text{Bernoulli}(p)$, $Y = (X+1)^2$

Support Set for Y

$\{(0+1)^2, (1+1)^2\} = \{1, 4\}$

Probability Mass Function for Y

$$p_Y(y) = \begin{cases} 1-p & y = 1 \\ p & y = 4 \\ 0 & \text{otherwise} \end{cases}$$

Expected Value of Y

$$\sum_{y \in \{1,4\}} y \times p_Y(y) = 1 \times (1-p) + 4 \times p = 1 + 3p$$

Example: $X \sim \text{Bernoulli}(p)$, $Y = (X+1)^2$

$E[g(X)] = E[(X+1)^2]$

$$\sum_{y \in \{1,4\}} y \times p_Y(y) = 1 \times (1-p) + 4 \times p = 1 + 3p$$

Example: $X \sim \text{Bernoulli}(p)$, $Y = (X+1)^2$

$E[g(X)] = E[(X+1)^2]$

$$\sum_{y \in \{1,4\}} y \times p_Y(y) = 1 \times (1-p) + 4 \times p = 1 + 3p$$

$g(E[X]) = (E[X]+1)^2$

$$(E[X]+1)^2 = (p+1)^2 = 1 + 2p + p^2$$

Example: $X \sim \text{Bernoulli}(p)$, $Y = (X+1)^2$

$E[g(X)] = E[(X+1)^2]$

$$\sum_{y \in \{1,4\}} y \times p_Y(y) = 1 \times (1-p) + 4 \times p = 1 + 3p$$

$g(E[X]) = (E[X]+1)^2$

$$(E[X]+1)^2 = (p+1)^2 = 1 + 2p + p^2$$

In general: $1 + 3p \neq 1 + 2p + p^2$!

$$E[g(X)] \neq g(E[X])$$

(Expected value of Function $\neq$ Function of Expected Value)

Expectation of a Function of a Discrete RV

Let X be a random variable and g be a function. Then:

$$\boxed{E[g(X)] = \sum_{\text{all } x} g(x)p(x)}$$

This is how we proceeded in the St. Petersburg Game Example

Your Turn: Calculate $E[X^2]$

X has support $\{-1, 0, 1\}$, $p(-1) = p(0) = p(1) = 1/3$.

Your Turn: Calculate $E[X^2]$

X has support $\{-1, 0, 1\}$, $p(-1) = p(0) = p(1) = 1/3$.

$$E[X^2] = \sum_{\text{all } x} x^2 p(x) = \sum_{x \in \{-1,0,1\}} x^2 p(x)$$

Your Turn: Calculate $E[X^2]$

X has support $\{-1, 0, 1\}$, $p(-1) = p(0) = p(1) = 1/3$.

$$\begin{aligned} E[X^2] &= \sum_{\text{all } x} x^2 p(x) = \sum_{x \in \{-1,0,1\}} x^2 p(x) \\ &= (-1)^2 \cdot (1/3) + (0)^2 \cdot (1/3) + (1)^2 \cdot (1/3) \end{aligned}$$

Your Turn: Calculate $E[X^2]$

X has support $\{-1, 0, 1\}$, $p(-1) = p(0) = p(1) = 1/3$.

$$\begin{aligned} E[X^2] &= \sum_{\text{all } x} x^2 p(x) = \sum_{x \in \{-1,0,1\}} x^2 p(x) \\ &= (-1)^2 \cdot (1/3) + (0)^2 \cdot (1/3) + (1)^2 \cdot (1/3) \\ &= 1/3 + 1/3 \\ &= 2/3 \approx 0.67 \end{aligned}$$

Linearity of Expectation

Holds for Continuous RVs as well, but proof is different.

Let X be a RV and a, b be constants. Then:

$$E[a + bX] = a + bE[X]$$

This is a Crucial Exception

In general $E[g(X)]$ does not equal $g(E[X])$. But in the special case where g is a linear function, $g(X) = a + bX$, the two are equal.

Example: Linearity of Expectation

Let $X \sim \text{Bernoulli}(1/3)$ and define $Y = 3X + 2$

1. What is $E[X]$?

Example: Linearity of Expectation

Let $X \sim \text{Bernoulli}(1/3)$ and define $Y = 3X + 2$

1. What is $E[X]$? $E[X] = 0 \times 2/3 + 1 \times 1/3 = 1/3$

Example: Linearity of Expectation

Let $X \sim \text{Bernoulli}(1/3)$ and define $Y = 3X + 2$

1. What is $E[X]$? $E[X] = 0 \times 2/3 + 1 \times 1/3 = 1/3$
2. What is $E[Y]$?

Example: Linearity of Expectation

Let $X \sim \text{Bernoulli}(1/3)$ and define $Y = 3X + 2$

1. What is $E[X]$? $E[X] = 0 \times 2/3 + 1 \times 1/3 = 1/3$
2. What is $E[Y]$? $E[Y] = E[3X + 2] = 3E[X] + 2 = 3$

Proof: Linearity of Expectation For Discrete RV

$$\begin{aligned}
E[a + bX] &= \sum_{\text{all } x} (a + bx)p(x) \\
&= \sum_{\text{all } x} p(x) \cdot a + \sum_{\text{all } x} p(x) \cdot bx \\
&= a \sum_{\text{all } x} p(x) + b \sum_{\text{all } x} x \cdot p(x) \\
&= a + bE[X]
\end{aligned}$$

Lecture #9 – Discrete RVs III

Variance and Standard Deviation of a Random Variable

Binomial Random Variable

Joint Probability Mass Function

Joint versus Marginal Probability Mass Function

Variance and Standard Deviation of a RV

The Defs are the same for continuous RVs, but the method of calculating will differ.

Variance (Var)

$$\sigma^2 = Var(X) = E\left[(X-\mu)^2\right] = E\left[(X-E[X])^2\right]$$

Standard Deviation (SD)

$$\sigma = \sqrt{\sigma^2} = SD(X)$$

Key Point

Variance and std. dev. are *expectations of functions of a RV*

It follows that:

1. Variance and SD are constants
2. To derive facts about them you can use the facts you know about expected value

How To Calculate Variance for Discrete RV?

Remember: it's just a function of X!

Recall that $\mu = E[X] = \sum_{\text{all } x} x p(x)$

$$Var(X) = E\left[(X-\mu)^2\right] = \sum_{\text{all } x} (x-\mu)^2 p(x)$$

Shortcut Formula For Variance

This is *not* the definition, it's a shortcut for doing calculations:

$$Var(X) = E\left[(X-\mu)^2\right] = E[X^2] - (E[X])^2$$

We'll prove this in an upcoming lecture.

Example: The Shortcut Formula

Let $X \sim \text{Bernoulli}(1/2)$. Calculate $Var(X)$.

Example: The Shortcut Formula

Let $X \sim \text{Bernoulli}(1/2)$. Calculate $Var(X)$.

$$\begin{aligned} E[X] &= 0 \times 1/2 + 1 \times 1/2 = 1/2 \\ E[X^2] &= 0^2 \times 1/2 + 1^2 \times 1/2 = 1/2 \end{aligned}$$

Example: The Shortcut Formula

Let $X \sim \text{Bernoulli}(1/2)$. Calculate $Var(X)$.

$$
\begin{aligned}
E[X] &= 0 \times 1/2 + 1 \times 1/2 = 1/2 \\
E[X^2] &= 0^2 \times 1/2 + 1^2 \times 1/2 = 1/2
\end{aligned}
$$

$$E[X^2] - (E[X])^2 = 1/2 - (1/2)^2 = 1/4$$

Variance of Bernoulli RV – via the Shortcut Formula

Step 1 – $E[X]$

$$\mu = E[X] = \sum_{x \in \{0,1\}} p(x) \cdot x = (1-p) \cdot 0 + p \cdot 1 = p$$

Step 2 – $E[X^2]$

$$E[X^2] = \sum_{x \in \{0,1\}} x^2 p(x) = 0^2(1-p) + 1^2 p = p$$

Step 3 – Combine with Shortcut Formula

$$\sigma^2 = Var[X] = E[X^2] - (E[X])^2 = p - p^2 = p(1-p)$$

Variance of a Linear Transformation

$$
\begin{aligned}
Var(a+bX) &= E\left[\{(a+bX)-E(a+bX)\}^2\right] \\
&= E\left[\{(a+bX)-(a+bE[X])\}^2\right] \\
&= E\left[(bX-bE[X])^2\right] \\
&= E[b^2(X-E[X])^2] \\
&= b^2E[(X-E[X])^2] \\
&= b^2Var(X) = b^2\sigma^2
\end{aligned}
$$

The key point here is that variance is defined in terms of expectation and expectation is linear.

Variance and SD are *NOT* Linear

$$Var(a + bX) = b^2\sigma^2$$

$$SD(a + bX) = |b|\sigma$$

These should look familiar from the related results for sample variance and std. dev. that you worked out on an earlier problem set.

Binomial Random Variable

Let $X =$ the sum of n independent Bernoulli trials, each with probability of success p. Then we say that: $X \sim \text{Binomial}(n, p)$

Parameters

$p =$ probability of "success," $n = \#$ of trials

Support

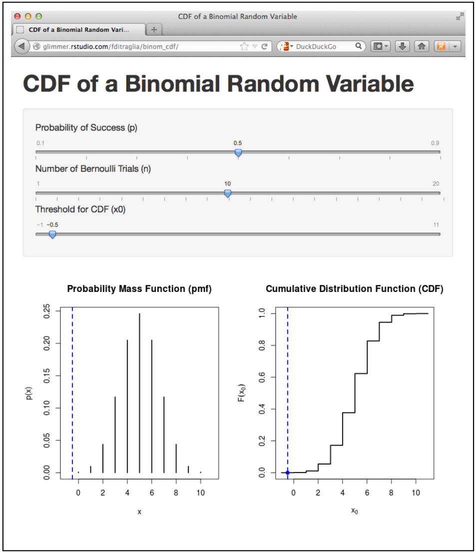
$\{0, 1, 2, \ldots, n\}$

Probability Mass Function (pmf)

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

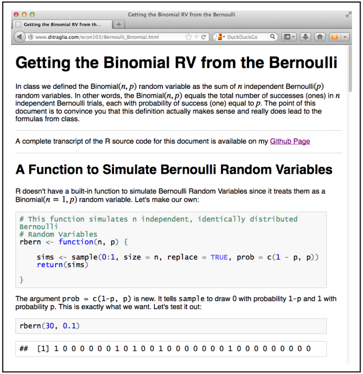
http://fditraglia.shinyapps.io/binom_cdf/

Try playing around with all three sliders. If you set the second to 1 you get a Bernoulli.



http://fditraglia.github.com/Econ103Public/Rtutorials/Bernoulli_Binomial.html

Source Code on my Github Page



Getting the Binomial RV from the Bernoulli

www.ditraglia.com/econ103/Bernoulli_Binomial.html

Getting the Binomial RV from the Bernoulli

In class we defined the Binomial(n, p) random variable as the sum of n independent Bernoulli(p) random variables. In other words, the Binomial(n, p) equals the total number of successes (ones) in n independent Bernoulli trials, each with probability of success (one) equal to p. The point of this document is to convince you that this definition actually makes sense and really does lead to the formulas from class.

A complete transcript of the R source code for this document is available on my Github Page

A Function to Simulate Bernoulli Random Variables

R doesn't have a built-in function to simulate Bernoulli Random Variables since it treats them as a Binomial$(n = 1, p)$ random variable. Let's make our own:

```
# This function simulates n independent, identically distributed
Bernoulli
# Random Variables
rbern <- function(n, p) {

    sims <- sample(0:1, size = n, replace = TRUE, prob = c(1 - p, p))
    return(sims)

}
```

The argument `prob = c(1-p, p)` is new. It tells `sample` to draw 0 with probability `1-p` and 1 with probability `p`. This is exactly what we want. Let's test it out:

```
rbern(30, 0.1)
```

```
##  [1] 1 0 0 0 0 0 0 1 0 1 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
```

Don't forget this!

Binomial RV counts up the *total* number of successes (ones) in n indep. Bernoulli trials, each with prob. of success p.

Where does the Binomial pmf come from?

Question

Suppose we flip a fair coin 3 times. What is the probability that we get exactly 2 heads?

Where does the Binomial pmf come from?

Question

Suppose we flip a fair coin 3 times. What is the probability that we get exactly 2 heads?

Answer

Three basic outcomes make up this event: $\{HHT, HTH, THH\}$, each has probability $1/8 = 1/2 \times 1/2 \times 1/2$. Basic outcomes are mutually exclusive, so sum to get $3/8 = 0.375$

Where does the Binomial pmf come from?

Question

Suppose we flip an *unfair* coin 3 times, where the probability of heads is $1/3$. What is the probability that we get exactly 2 heads?

Answer

No longer true that *all* basic outcomes are equally likely, but those with exactly two heads *still are*

$$
\begin{aligned}
P(HHT) &= (1/3)^2(1-1/3) = 2/27 \\
P(THH) &= 2/27 \\
P(HTH) &= 2/27
\end{aligned}
$$

Summing gives $2/9 \approx 0.22$

Where does the Binomial pmf come from?

Starting to see a pattern?

Suppose we flip an unfair coin *4* times, where the probability of heads is $1/3$. What is the probability that we get exactly 2 heads?

HHTT	TTHH
HTHT	THTH
HTTH	THHT

Six equally likely, mutually exclusive basic outcomes make up this event:

$$\binom{4}{2}(1/3)^2(2/3)^2$$

Multiple RVs *at once* - Definition of Joint PMF

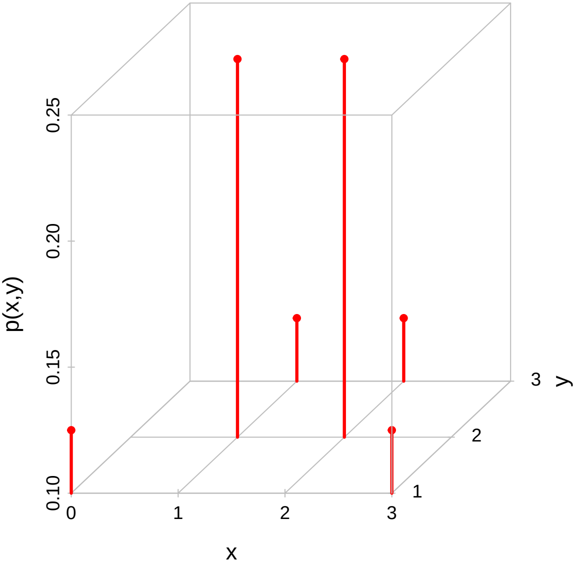
Let X and Y be discrete random variables. The joint probability mass function $p_{XY}(x, y)$ gives the probability of each pair of realizations (x, y) in the support:

$$\boxed{p_{XY}(x, y) = P(X = x \cap Y = y)}$$

Example: Joint PMF in Tabular Form

		Y		
		1	2	3
X	0	1/8	0	0
	1	0	1/4	1/8
	2	0	1/4	1/8
	3	1/8	0	0

Plot of Joint PMF



What is $p_{XY}(1,2)$?

		Y		
		1	2	3
X	0	1/8	0	0
	1	0	1/4	1/8
	2	0	1/4	1/8
	3	1/8	0	0

What is $p_{XY}(1,2)$?

		Y		
		1	2	3
X	0	1/8	0	0
	1	0	1/4	1/8
	2	0	1/4	1/8
	3	1/8	0	0

$$p_{XY}(1,2) \quad = \quad P(X = 1 \cap Y = 2) = 1/4$$

What is $p_{XY}(1,2)$?

		Y		
		1	2	3
X	0	1/8	0	0
	1	0	1/4	1/8
	2	0	1/4	1/8
	3	1/8	0	0

$$p_{XY}(1,2) = P(X = 1 \cap Y = 2) = 1/4$$

$$p_{XY}(2,1) = P(X = 2 \cap Y = 1) = 0$$

Properties of Joint PMF

1. $0 \leq p_{XY}(x, y) \leq 1$ for any pair (x, y)
2. The sum of $p_{XY}(x, y)$ over all pairs (x, y) in the support is 1:

$$\sum_x \sum_y p(x, y) = 1$$

Joint versus Marginal PMFs

Joint PMF

$p_{XY}(x,y) = P(X = x \cap Y = y)$

Marginal PMFs

$p_X(x) = P(X = x)$

$p_Y(y) = P(Y = y)$

You can't calculate a joint pmf from marginals alone but you *can* calculate marginals from the joint!

Marginals from Joint

$$\boxed{p_X(x) = \sum_{\text{all } y} p_{XY}(x, y)}$$

$$\boxed{p_Y(y) = \sum_{\text{all } x} p_{XY}(x, y)}$$

Why?

$$\begin{aligned} p_Y(y) &= P(Y = y) = P\left(\bigcup_{\text{all } x} \{X = x \cap Y = y\}\right) \\ &= \sum_{\text{all } x} P(X = x \cap Y = y) = \sum_{\text{all } x} p_{XY}(x, y) \end{aligned}$$

To get the marginals sum "into the margins" of the table.

		Y			
		1	2	3	
X	0	1/8	0	0	
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	

To get the marginals sum "into the margins" of the table.

		Y			
		1	2	3	
	0	1/8	0	0	1/8
X	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	

$$p_X(0) \;=\; 1/8 + 0 + 0 = 1/8$$

To get the marginals sum "into the margins" of the table.

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	
	3	1/8	0	0	

$$p_X(0) = 1/8 + 0 + 0 = 1/8$$
$$p_X(1) = 0 + 1/4 + 1/8 = 3/8$$

To get the marginals sum "into the margins" of the table.

		Y			
		1	2	3	
	0	1/8	0	0	1/8
X	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	

$$\begin{aligned} p_X(0) &= 1/8 + 0 + 0 = 1/8 \\ p_X(1) &= 0 + 1/4 + 1/8 = 3/8 \\ p_X(2) &= 0 + 1/4 + 1/8 = 3/8 \end{aligned}$$

To get the marginals sum “into the margins” of the table.

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
					1

$$\begin{aligned} p_X(0) &= 1/8 + 0 + 0 = 1/8 \\ p_X(1) &= 0 + 1/4 + 1/8 = 3/8 \\ p_X(2) &= 0 + 1/4 + 1/8 = 3/8 \\ p_X(3) &= 1/8 + 0 + 0 = 1/8 \end{aligned}$$

What is $p_Y(2)$?

		Y			
		1	2	3	
X	0	1/8	0	0	
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	

What is $p_Y(2)$?

		Y			
		1	2	3	
X	0	1/8	0	0	
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	
		1/4			

$$p_Y(1) = 1/8 + 0 + 0 + 1/8 = 1/4$$

What is $p_Y(2)$?

			Y		
		1	2	3	
X	0	1/8	0	0	
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	
		1/4	1/2		

$$
\begin{aligned}
p_Y(1) &= 1/8 + 0 + 0 + 1/8 = 1/4 \\
p_Y(2) &= 0 + 1/4 + 1/4 + 0 = 1/2
\end{aligned}
$$

What is $p_Y(2)$?

		Y = 1	2	3	
X	0	1/8	0	0	
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	
		1/4	1/2	1/4	1

$$
\begin{aligned}
p_Y(1) &= 1/8 + 0 + 0 + 1/8 = 1/4 \\
p_Y(2) &= 0 + 1/4 + 1/4 + 0 = 1/2 \\
p_Y(3) &= 0 + 1/8 + 1/8 + 0 = 1/4
\end{aligned}
$$

Lecture #10 – Discrete RVs IV

Conditional Probability Mass Function & Independence

Conditional Expectation & The Law of Iterated Expectations

Expectation of a Function of Two Discrete RVs, Covariance

Linearity of Expectation Reprise, Properties of Binomial RV

Definition of Conditional PMF

How does the distribution of y change with x?

$$p_{Y|X}(y|x) = P(Y = y|X = x) = \frac{P(Y = y \cap X = x)}{P(X = x)} = \frac{p_{XY}(x, y)}{p_X(x)}$$

Conditional PMF of Y given $X = 2$

		Y 1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8

$$p_{Y|X}(1|2) = \frac{p_{XY}(2,1)}{p_X(2)} = \frac{0}{3/8} = 0$$

$$p_{Y|X}(2|2) = \frac{p_{XY}(2,2)}{p_X(2)} = \frac{1/4}{3/8} = 2/3$$

$$p_{Y|X}(3|2) = \frac{p_{XY}(2,3)}{p_X(2)} = \frac{1/8}{3/8} = 1/3$$

What is $p_{X|Y}(1|2)$?

		Y			
		1	2	3	
X	0	1/8	0	0	
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	
		1/4	1/2	1/4	

What is $p_{X|Y}(1|2)$?

		Y			
		1	2	3	
X	0	1/8	0	0	
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	
		1/4	1/2	1/4	

$$p_{X|Y}(1|2) = \frac{p_{XY}(1,2)}{p_Y(2)} = \frac{1/4}{1/2} = 1/2$$

What is $p_{X|Y}(1|2)$?

		Y			
		1	2	3	
X	0	1/8	0	0	
	1	0	1/4	1/8	
	2	0	1/4	1/8	
	3	1/8	0	0	
		1/4	1/2	1/4	

$$p_{X|Y}(1|2) = \frac{p_{XY}(1,2)}{p_Y(2)} = \frac{1/4}{1/2} = 1/2$$

Similarly:

$$p_{X|Y}(0|2) = 0, \quad p_{X|Y}(2|2) = 1/2, \quad p_{X|Y}(3|2) = 0$$

Independent RVs: Joint Equals Product of Marginals

Definition

Two discrete RVs are independent if and only if

$$p_{XY}(x,y) = p_X(x)p_Y(y)$$

for all pairs (x,y) in the support.

Equivalent Definition

$$p_{Y|X}(y|x) = p_Y(y) \text{ and } p_{X|Y}(x|y) = p_X(x)$$

for all pairs (x,y) in the support.

Are X and Y Independent?

(A = YES, B = NO)

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

Are X and Y Independent?

(A = YES, B = NO)

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

$$
\begin{aligned}
p_{XY}(2,1) &= 0 \\
p_X(2) \times p_Y(1) &= (3/8) \times (1/4) \neq 0
\end{aligned}
$$

Therefore X and Y are *not* independent.

Conditional Expectation

Intuition

$E[Y|X] =$ "best guess" of realization that Y after observing realization of X.

$E[Y|X]$ is a Random Variable

While $E[Y]$ is a constant, $E[Y|X]$ is a function of X, hence a Random Variable.

$E[Y|X = x]$ is a Constant

The constant $E[Y|X = x]$ is the "guess" of Y if we see $X = x$.

Calculating $E[Y|X = x]$

Take the mean of the conditional pmf of Y given $X = x$.

Conditional Expectation: $E[Y|X=2]$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

We showed above that the conditional pmf of $Y|X=2$ is:

$$\boxed{p_{Y|X}(1|2) = 0 \qquad p_{Y|X}(2|2) = 2/3 \qquad p_{Y|X}(3|2) = 1/3}$$

Hence

$$E[Y|X=2] = 2 \times 2/3 + 3 \times 1/3 = 7/3$$

Conditional Expectation: $E[Y|X=0]$

		Y 1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

The conditional pmf of $Y|X=0$ is

$$\boxed{p_{Y|X}(1|0) = 1 \qquad p_{Y|X}(2|0) = 0 \qquad p_{Y|X}(3|0) = 0}$$

Hence $E[Y|X=0] = 1$

Calculate $E[Y|X = 3]$

		Y 1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

The conditional pmf of $Y|X = 3$ is

$$\boxed{p_{Y|X}(1|3) = 1 \qquad p_{Y|X}(2|3) = 0 \qquad p_{Y|X}(3|3) = 0}$$

Hence $E[Y|X = 3] = 1$

Calculate $E[Y|X = 1]$

		Y			
		1	2	3	
	0	1/8	0	0	1/8
X	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

Calculate $E[Y|X=1]$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

The conditional pmf of $Y|X=1$ is

$$\boxed{p_{Y|X}(1|1)=0 \qquad p_{Y|X}(2|1)=2/3 \qquad p_{Y|X}(3|1)=1/3}$$

Hence

$$E[Y|X=1] = 2\times 2/3 + 3\times 1/3 = 7/3$$

$E[Y|X]$ is a Random Variable

For this example:

$$E[Y|X] = \begin{cases} 1 & X = 0 \\ 7/3 & X = 1 \\ 7/3 & X = 2 \\ 1 & X = 3 \end{cases}$$

From above the marginal distribution of X is:

$$P(X = 0) = 1/8 \qquad P(X = 1) = 3/8$$
$$P(X = 2) = 3/8 \qquad P(X = 3) = 1/8$$

$E[Y|X]$ takes the value 1 with prob. 1/4 and 7/3 with prob. 3/4.

The Law of Iterated Expectations

$E[Y|X]$ is an RV so what is its expectation?

For any RVs X and Y

$$E\left[E\left[Y|X\right]\right] = E[Y]$$

Option proof $\boxed{\text{HERE}}$. (Helpful for Econ 104...)

Law of Iterated Expectations for Our Example

Marginal pmf of Y

$$
\begin{aligned}
P(Y = 1) &= 1/4 \\
P(Y = 2) &= 1/2 \\
P(Y = 3) &= 1/4
\end{aligned}
$$

$$
\begin{aligned}
E[Y] &= 1 \times 1/4 + 2 \times 1/2 + 3 \times 1/4 \\
&= 2
\end{aligned}
$$

$E[Y|X]$

$$
E[Y|X] = \begin{cases} 1 & \text{w/ prob. } 1/4 \\ 7/3 & \text{w/ prob. } 3/4 \end{cases}
$$

$$
\begin{aligned}
E\left[E\left[Y|X\right]\right] &= 1 \times 1/4 + 7/3 \times 3/4 \\
&= 2
\end{aligned}
$$

Expectation of Function of Two Discrete RVs

$$\boxed{E[g(X, Y)] = \sum_x \sum_y g(x, y) p_{XY}(x, y)}$$

Some Extremely Important Examples

Same For Continuous Random Variables

Let $\mu_X = E[X], \mu_Y = E[Y]$

Covariance

$$\sigma_{XY} = Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

Correlation

$$\rho_{XY} = Corr(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Shortcut Formula for Covariance

Much easier for calculating:

$$\boxed{Cov(X, Y) = E[XY] - E[X]E[Y]}$$

I'll mention this again in a few slides...

Calculating $Cov(X, Y)$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

$$E[X] = 3/8 + 2 \times 3/8 + 3 \times 1/8 = 3/2$$

$$E[Y] = 1/4 + 2 \times 1/2 + 3 \times 1/4 = 2$$

Calculating $Cov(X, Y)$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

$$E[X] = 3/8 + 2 \times 3/8 + 3 \times 1/8 = 3/2$$

$$E[Y] = 1/4 + 2 \times 1/2 + 3 \times 1/4 = 2$$

$$\begin{aligned} E[XY] &= 1/4 \times (2+4) + 1/8 \times (3+6+3) \\ &= 3 \end{aligned}$$

Calculating $Cov(X, Y)$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

$$E[X] = 3/8 + 2 \times 3/8 + 3 \times 1/8 = 3/2$$

$$E[Y] = 1/4 + 2 \times 1/2 + 3 \times 1/4 = 2$$

$$\begin{aligned} E[XY] &= 1/4 \times (2+4) + 1/8 \times (3+6+3) \\ &= 3 \end{aligned}$$

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$

Calculating $Cov(X, Y)$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

$$E[X] = 3/8 + 2 \times 3/8 + 3 \times 1/8 = 3/2$$

$$E[Y] = 1/4 + 2 \times 1/2 + 3 \times 1/4 = 2$$

$$\begin{aligned} E[XY] &= 1/4 \times (2+4) + 1/8 \times (3+6+3) \\ &= 3 \end{aligned}$$

$$\begin{aligned} Cov(X, Y) &= E[XY] - E[X]E[Y] \\ &= 3 - 3/2 \times 2 = 0 \end{aligned}$$

Calculating $Cov(X, Y)$

		Y			
		1	2	3	
X	0	1/8	0	0	1/8
	1	0	1/4	1/8	3/8
	2	0	1/4	1/8	3/8
	3	1/8	0	0	1/8
		1/4	1/2	1/4	

$$E[X] = 3/8 + 2 \times 3/8 + 3 \times 1/8 = 3/2$$

$$E[Y] = 1/4 + 2 \times 1/2 + 3 \times 1/4 = 2$$

$$\begin{aligned} E[XY] &= 1/4 \times (2+4) + 1/8 \times (3+6+3) \\ &= 3 \end{aligned}$$

$$\begin{aligned} Cov(X, Y) &= E[XY] - E[X]E[Y] \\ &= 3 - 3/2 \times 2 = 0 \end{aligned}$$

$$Corr(X, Y) = Cov(X, Y)/\left[SD(X)SD(Y)\right] = 0$$

Zero Covariance versus Independence

- From this example we learn that zero covariance (correlation) *does not* imply independence.
- However, it turns out that independence *does* imply zero covariance (correlation).

Optional proof that independence implies zero covariance HERE.

Linearity of Expectation, Again

Holds for Continuous RVs as well, but different proof.

In general, $E[g(X, Y)] \neq g(E[X], E[Y])$. The key exception is when g is a linear function:

$$\boxed{E[aX + bY + c] = aE[X] + bE[Y] + c}$$

where X, Y are random variables and a, b, c are constants.

Optional proof HERE.

Application: Shortcut Formula for Variance

By the Linearity of Expectation,

$$Var(X) \quad = \quad E[(X-\mu)^2] =$$

Application: Shortcut Formula for Variance

By the Linearity of Expectation,

$$
\begin{aligned}
Var(X) &= E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] \\
&=
\end{aligned}
$$

Application: Shortcut Formula for Variance

By the Linearity of Expectation,

$$\begin{aligned} Var(X) &= E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= \end{aligned}$$

Application: Shortcut Formula for Variance

By the Linearity of Expectation,

$$\begin{aligned} Var(X) &= E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - 2\mu^2 + \mu^2 \\ &= \end{aligned}$$

Application: Shortcut Formula for Variance

By the Linearity of Expectation,

$$\begin{aligned} Var(X) &= E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - 2\mu^2 + \mu^2 \\ &= E[X^2] - \mu^2 \end{aligned}$$

We saw in a previous lecture that it's typically much easier to calculate variances using the shortcut formula.

Another Application: Shortcut Formula for Covariance

Similar to Shortcut for Variance: in fact $Var(X) = Cov(X, X)$

$$
\begin{aligned}
Cov(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\
&= E[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y] \\
&\vdots \\
&= E[XY] - E[X]E[Y]
\end{aligned}
$$

You'll fill in the details for homework...

Expected Value of Sum = Sum of Expected Values

Repeatedly applying the linearity of expectation,

$$E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n]$$

regardless of how the RVs $X_1, \ldots, X_n$ are related to each other. In particular it doesn't matter if they're dependent or independent.

Independent and Identically Distributed (iid) RVs

Example

$X_1, X_2, \ldots X_n \sim$ iid Bernoulli(p)

Independent

Realization of one of the RVs gives no information about the others.

Identically Distributed

Each X_i is the same kind of RV, with the same values for any parameters. (Hence same pmf, cdf, mean, variance, etc.)

Recall: Binomial(n, p) Random Variable

Definition

Sum of n independent Bernoulli RVs, each with probability of "success," i.e. 1, equal to p

Using Our New Notation

Let $X_1, X_2, \ldots, X_n \sim$ iid Bernoulli(p), $Y = X_1 + X_2 + \ldots + X_n$.
Then $Y \sim$ Binomial(n, p).

Expected Value of Binomial RV

Use the fact that a Binomial(n, p) RV is defined as the sum of n iid Bernoulli(p) Random Variables and the Linearity of Expectation:

$$\begin{aligned} E[Y] &= E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n] \\ &= p + p + \ldots + p \\ &= np \end{aligned}$$

Variance of a Sum $\neq$ Sum of Variances!

$$Var(aX + bY) = E\left[\{(aX + bY) - E[aX + bY]\}^2\right]$$

Variance of a Sum $\neq$ Sum of Variances!

$$
\begin{aligned}
Var(aX + bY) &= E\left[\{(aX + bY) - E[aX + bY]\}^2\right] \\
&= E\left[\{a(X - \mu_X) + b(Y - \mu_Y)\}^2\right]
\end{aligned}
$$

Variance of a Sum $\neq$ Sum of Variances!

$$
\begin{aligned}
Var(aX + bY) &= E\left[\{(aX + bY) - E[aX + bY]\}^2\right] \\
&= E\left[\{a(X - \mu_X) + b(Y - \mu_Y)\}^2\right] \\
&= E\left[a^2(X - \mu_X)^2 + b^2(Y - \mu_Y)^2 + 2ab(X - \mu_X)(Y - \mu_Y)\right]
\end{aligned}
$$

Variance of a Sum $\neq$ Sum of Variances!

$$
\begin{aligned}
Var(aX + bY) &= E\left[\{(aX + bY) - E[aX + bY]\}^2\right] \\
&= E\left[\{a(X - \mu_X) + b(Y - \mu_Y)\}^2\right] \\
&= E\left[a^2(X - \mu_X)^2 + b^2(Y - \mu_Y)^2 + 2ab(X - \mu_X)(Y - \mu_Y)\right] \\
&= a^2E[(X - \mu_X)^2] + b^2E[(Y - \mu_Y)^2] + 2abE[(X - \mu_X)(Y - \mu_Y)]
\end{aligned}
$$

Variance of a Sum $\neq$ Sum of Variances!

$$\begin{aligned}
Var(aX + bY) &= E\left[\{(aX + bY) - E[aX + bY]\}^2\right] \\
&= E\left[\{a(X - \mu_X) + b(Y - \mu_Y)\}^2\right] \\
&= E\left[a^2(X - \mu_X)^2 + b^2(Y - \mu_Y)^2 + 2ab(X - \mu_X)(Y - \mu_Y)\right] \\
&= a^2 E[(X - \mu_X)^2] + b^2 E[(Y - \mu_Y)^2] + 2abE[(X - \mu_X)(Y - \mu_Y)] \\
&= a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)
\end{aligned}$$

Since $\sigma_{XY} = \rho\sigma_X\sigma_Y$, this is sometimes written as:

$$\boxed{Var(aX + bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y}$$

Independence $\Rightarrow Var(X+Y) = Var(X) + Var(Y)$

X and Y independent $\Longrightarrow Cov(X,Y) = 0$. Hence, independence implies

$$\begin{aligned} Var(X+Y) &= Var(X) + Var(Y) + 2Cov(X,Y) \\ &= Var(X) + Var(Y) \end{aligned}$$

Also true for three or more RVs

If $X_1, X_2, \ldots, X_n$ are independent, then

$$Var(X_1 + X_2 + \ldots X_n) = Var(X_1) + Var(X_2) + \ldots + Var(X_n)$$

Crucial Distinction

Expected Value

Always true that

$$E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n]$$

Variance

Not true in general that

$$Var[X_1 + X_2 + \ldots + X_n] = Var[X_1] + Var[X_2] + \ldots + Var[X_n]$$

except in the special case where $X_1, \ldots X_n$ are independent (or at least uncorrelated).

Variance of Binomial Random Variable

Definition from Sequence of Bernoulli Trials

If $X_1, X_2, \ldots, X_n \sim$ iid Bernoulli(p) then

$$Y = X_1 + X_2 + \ldots + X_n \sim \text{Binomial}(n, p)$$

Using Independence

$$\begin{aligned}
Var[Y] &= Var[X_1 + X_2 + \ldots + X_n] \\
&= Var[X_1] + Var[X_2] + \ldots + Var[X_n] \\
&=
\end{aligned}$$

Variance of Binomial Random Variable

Definition from Sequence of Bernoulli Trials

If $X_1, X_2, \ldots, X_n \sim$ iid Bernoulli(p) then

$$Y = X_1 + X_2 + \ldots + X_n \sim \text{Binomial}(n, p)$$

Using Independence

$$\begin{aligned} Var[Y] &= Var[X_1 + X_2 + \ldots + X_n] \\ &= Var[X_1] + Var[X_2] + \ldots + Var[X_n] \\ &= p(1-p) + p(1-p) + \ldots + p(1-p) \\ &= \end{aligned}$$

Variance of Binomial Random Variable

Definition from Sequence of Bernoulli Trials

If $X_1, X_2, \ldots, X_n \sim$ iid Bernoulli(p) then

$$Y = X_1 + X_2 + \ldots + X_n \sim \text{Binomial}(n, p)$$

Using Independence

$$\begin{aligned} Var[Y] &= Var[X_1 + X_2 + \ldots + X_n] \\ &= Var[X_1] + Var[X_2] + \ldots + Var[X_n] \\ &= p(1-p) + p(1-p) + \ldots + p(1-p) \\ &= np(1-p) \end{aligned}$$

Lecture #11 – Continuous RVs I

Introduction: Probability as Area

Probability Density Function (PDF)

Relating the PDF to the CDF

Calculating the Probability of an Interval

Calculating Expected Value for Continuous RVs

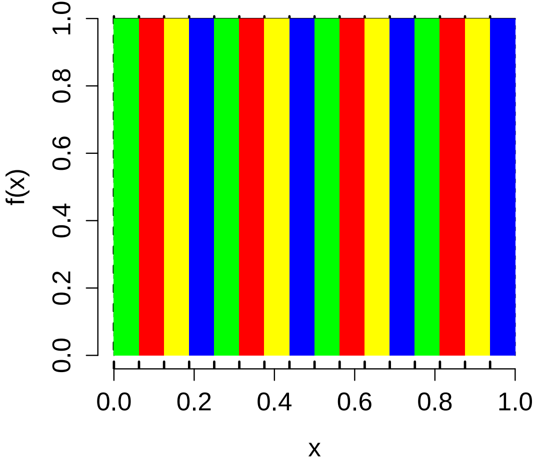
Continuous RVs – What Changes?

1. Probability Density Functions replace Probability Mass Functions (aka Probability Distributions)
2. Integrals Replace Sums

Everything Else is Essentially Unchanged!

What is the probability of "Yellow?"

From Twister to Density – Probability as *Area*



Continuous Random Variables

For continuous RVs, probability is a matter of finding the area of *intervals*. Individual *points* have *zero* probability.

Probability Density Function (PDF)

For a continuous random variable X,

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

where $f(x)$ is the *probability density function* for X.

Extremely Important

For any realization x, $P(X = x) = 0 \neq f(x)$!

Properties of PDFs

1. $\int_{-\infty}^{\infty} f(x)\, dx = 1$
2. $f(x) \geq 0$ for all x
3. $f(x)$ is *not* a probability and can be greater than one!
4. $P(X \leq x_0) = F(x_0) = \int_{-\infty}^{x_0} f(x)\, dx$

Simplest Possible Continuous RV: Uniform$(0, 1)$

You'll look at a generalization, Uniform(a, b) for homework.

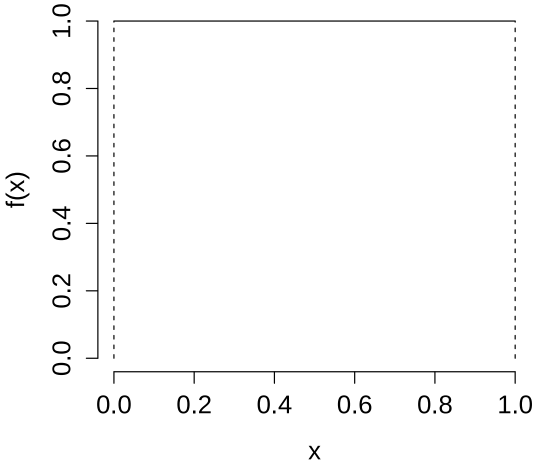
$X \sim$ Uniform$(0, 1)$

A Uniform(0,1) RV is equally likely to take on *any value* in the range $[0, 1]$ and never takes on a value outside this range.

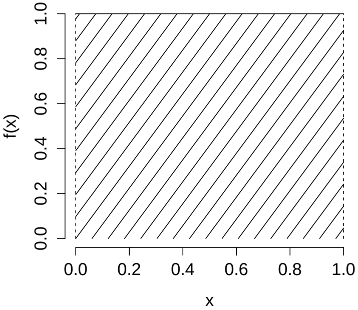
Uniform PDF

$f(x) = 1$ for $0 \leq x \leq 1$, zero elsewhere.

Uniform$(0, 1)$ PDF

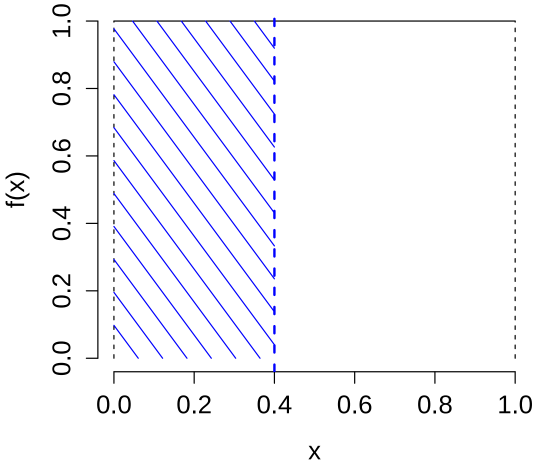


Check that Uniform pdf Integrates to 1

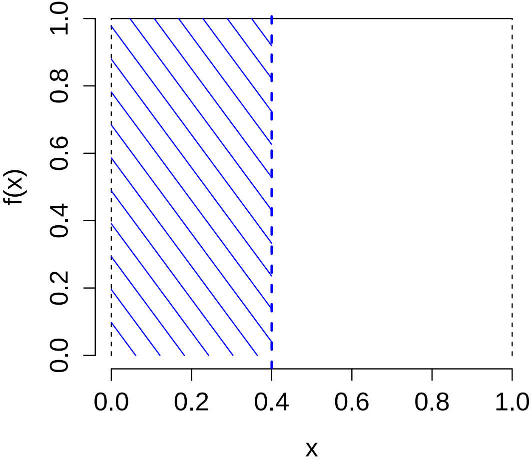


$$\int_{-\infty}^{\infty} f(x)\, dx = \int_0^1 1\, dx = x\big|_0^1 = 1 - 0 = 1$$

What is the area of the shaded region?



$$F(0.4) = P(X \leq 0.4) = 0.4$$



Relationship between PDF and CDF

Integrate pdf $\rightarrow$ CDF

$$F(x_0) = P(X \leq x_0) = \int_{-\infty}^{x_0} f(x)\, dx$$

Differentiate CDF $\rightarrow$ pdf

$$f(x) = \frac{d}{dx}F(x)$$

This is just the First Fundamental Theorem of Calculus.

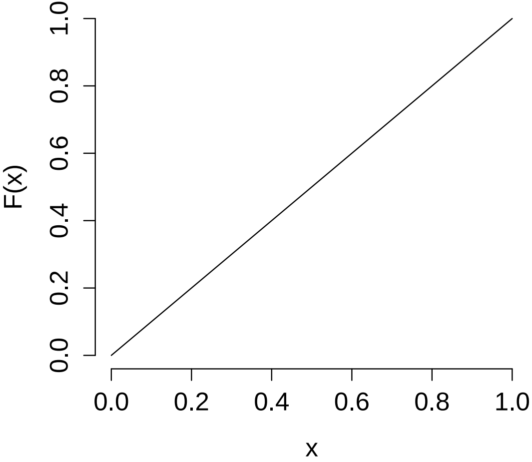
Example: Uniform$(0, 1)$ RV

Integrate the pdf, $f(x) = 1$, to get the CDF

$$F(x_0) = \int_{-\infty}^{x_0} f(x)\,dx = \int_0^{x_0} 1\,dx = x\big|_0^{x_0} = x_0 - 0 = x_0$$

$$F(x_0) = \begin{cases} 0, x_0 < 0 \\ x_0, 0 \leq x_0 \leq 1 \\ 1, x_0 > 1 \end{cases}$$

Uniform$(0, 1)$ CDF

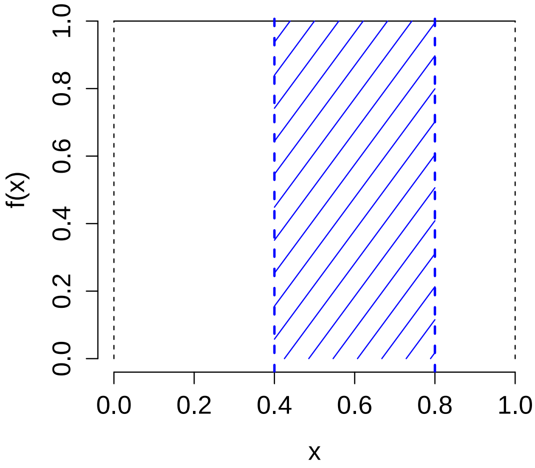


Example: Uniform$(0, 1)$ RV

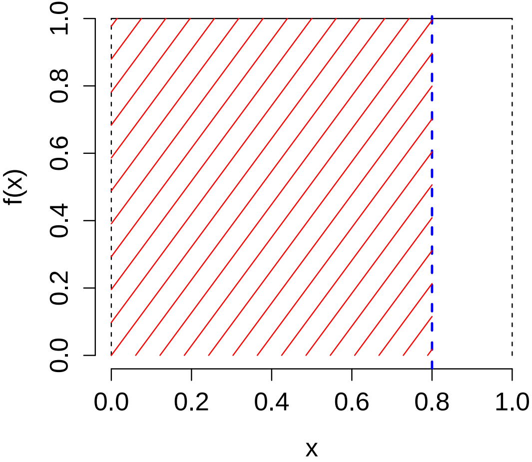
Differentiate the CDF, $F(x_0) = x_0$, to get the pdf

$$\frac{d}{dx}F(x) = 1 = f(x)$$

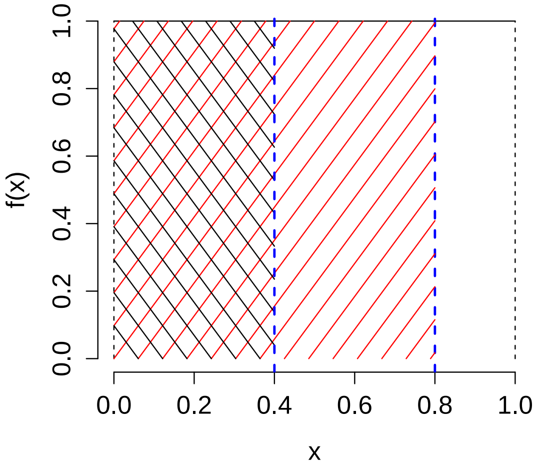
What is $P(0.4 \leq X \leq 0.8)$ if $X \sim \text{Uniform}(0, 1)$?



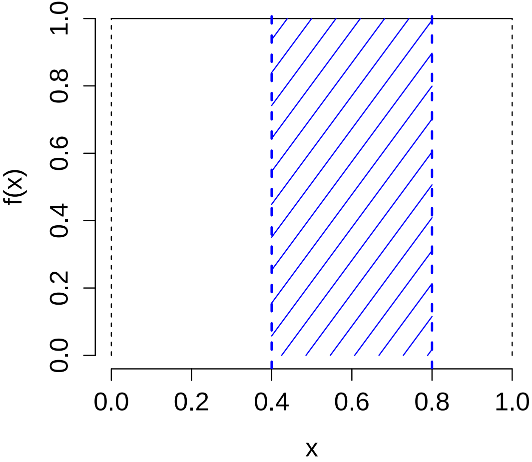
$F(0.8) = P(X \leq 0.8)$



$F(0.8) - F(0.4) = ?$



$$F(0.8) - F(0.4) = P(0.4 \le X \le 0.8) = 0.4$$



Key Idea: Probability of Interval for Continuous RV

$$\boxed{P(a \leq X \leq b) = \int_a^b f(x)\, dx = F(b) - F(a)}$$

This is just the Second Fundamental Theorem of Calculus.

Expected Value for Continuous RVs

$$\boxed{\int_{-\infty}^{\infty} xf(x)\,dx}$$

Remember: Integrals Replace Sums!

Example: Uniform(0,1) Random Variable

$$E[X] \; = \; \int_{-\infty}^{\infty} x f(x)\, dx =$$

Example: Uniform(0,1) Random Variable

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^1 x \cdot 1\, dx \\
&= \left.\frac{x^2}{2}\right|_0^1 = 1/2 - 0 = 1/2
\end{aligned}
$$

Expected Value of a Function of a Continuous RV

$$\boxed{E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\;dx}$$

Example: Uniform(0,1) Random Variable

$$E[X^2] \;=\; \int_{-\infty}^{\infty} x^2 f(x)\, dx$$

Example: Uniform(0,1) Random Variable

$$
\begin{aligned}
E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_0^1 x^2 \cdot 1\, dx \\
&= \left.\frac{x^3}{3}\right|_0^1 = 1/3
\end{aligned}
$$

What about all those rules for expected value?

- The only difference between expectation for continuous versus discrete is how we do the *calculation*.
- Sum for discrete; integral for continuous.
- All *properties* of expected value continue to hold!
- Includes linearity, shortcut for variance, etc.

Variance of Continuous RV

$$\boxed{Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx}$$

where

$$\mu = E[X] = \int_{-\infty}^{\infty} x f(x)\, dx$$

Shortcut formula still holds for continuous RVs!

$$Var(X) = E[X^2] - (E[X])^2$$

Example: Uniform$(0, 1)$ RV

$$
\begin{aligned}
Var(X) &= E\left[(X - E[X])^2\right] = E[X^2] - (E[X])^2 \\
&= 1/3 - (1/2)^2 \\
&= 1/12 \\
&\approx 0.083
\end{aligned}
$$

So where does that leave us?

What We've Accomplished

We've covered all the basic properties of RVs on this Handout.

Where are we headed next?

Next up is the most important RV of all: the normal RV. After that it's time to do some statistics!

How should you be studying?

If you *master* the material on RVs (both continuous and discrete) and in particular the normal RV the rest of the semester will seem easy. If you don't, you're in for a rough time...

Lecture #12 – Continuous RVs II: The Normal RV

The Standard Normal RV

Linear Combinations and the $N(\mu, \sigma^2)$ RV

Where does the Empirical Rule come from?

From $N(0, 1)$ to $N(\mu, \sigma^2)$ and Back Again

Percentiles/Quantiles for Continuous RVs

Available on Etsy, Made using R!

Figure: Standard Normal RV (PDF)

Standard Normal Random Variable: $N(0, 1)$

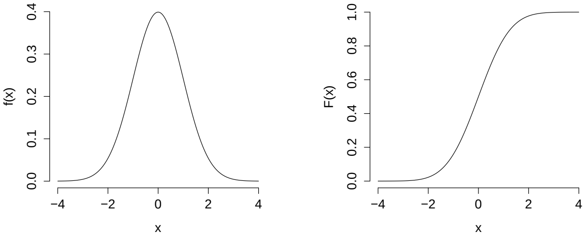
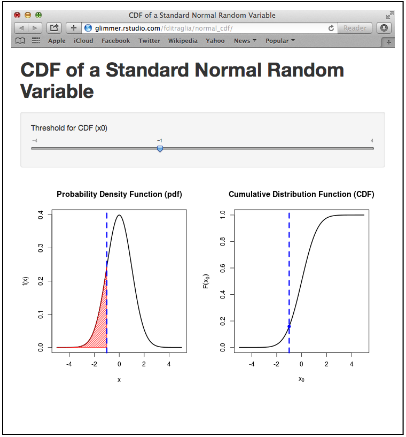


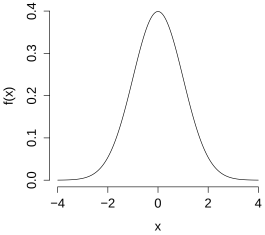
Figure: Standard Normal PDF (left) and CDF (Right)

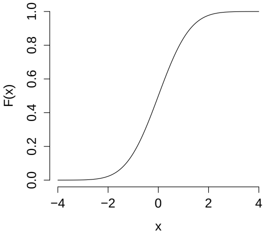
- Notation: $X \sim N(0, 1)$
- Symmetric, Bell-shaped, $E[X] = 0$, $Var[X] = 1$
- Support Set $= (-\infty, \infty)$

https://fditraglia.shinyapps.io/normal_cdf/



Standard Normal Random Variable: $N(0,1)$





- There is no closed-form expression for the $N(0,1)$ CDF.
- For Econ 103, don't need to know formula for $N(0,1)$ PDF.
- You *do need* to know the R commands. . .

R Commands for the Standard Normal RV

`dnorm` – Standard Normal PDF

- Mnemonic: `d` = density, `norm` = normal
- Example: `dnorm(0)` gives height of $N(0,1)$ PDF at zero.

R Commands for the Standard Normal RV

`dnorm` – Standard Normal PDF

- Mnemonic: `d` = density, `norm` = normal
- Example: `dnorm(0)` gives height of $N(0,1)$ PDF at zero.

`pnorm` – Standard Normal CDF

- Mnemonic: `p` = probability, `norm` = normal
- Example: `pnorm(1)` $= P(X \leq 1)$ if $X \sim N(0,1)$.

R Commands for the Standard Normal RV

`dnorm` – Standard Normal PDF

- Mnemonic: `d` = density, `norm` = normal
- Example: `dnorm(0)` gives height of $N(0,1)$ PDF at zero.

`pnorm` – Standard Normal CDF

- Mnemonic: `p` = probability, `norm` = normal
- Example: `pnorm(1)` $= P(X \leq 1)$ if $X \sim N(0,1)$.

`rnorm` – Simulate Standard Normal Draws

- Mnemonic: `r` = random, `norm` = normal.
- Example: `rnorm(10)` makes ten iid $N(0,1)$ draws.

$\Phi(x_0)$ Denotes the $N(0,1)$ CDF

You will sometimes encounter the notation $\Phi(x_0)$. It means the same thing as `pnorm`(x_0) but it's not an R command.

The $N(\mu, \sigma^2)$ Random Variable

Idea

Take a linear function of the $N(0, 1)$ RV.

Formal Definition

$N(\mu, \sigma^2) \equiv \mu + \sigma X$ where $X \sim N(0, 1)$ and μ, σ are constants.

Properties of $N(\mu, \sigma^2)$ RV

- Parameters: Expected Value $= \mu$, Variance $= \sigma^2$
- Symmetric and bell-shaped.
- Support Set $= (-\infty, \infty)$
- $N(0, 1)$ is the special case where $\mu = 0$ and $\sigma^2 = 1$.

Expected Value: μ *shifts* PDF

all of these have $\sigma = 1$

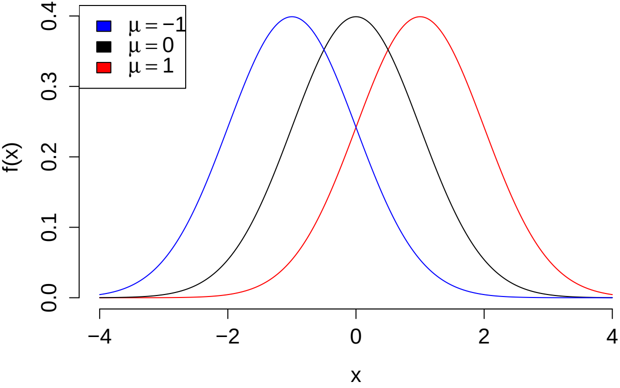


Figure: Blue $\mu = -1$, Black $\mu = 0$, Red $\mu = 1$

Standard Deviation: σ *scales* PDF

all of these have $\mu = 0$

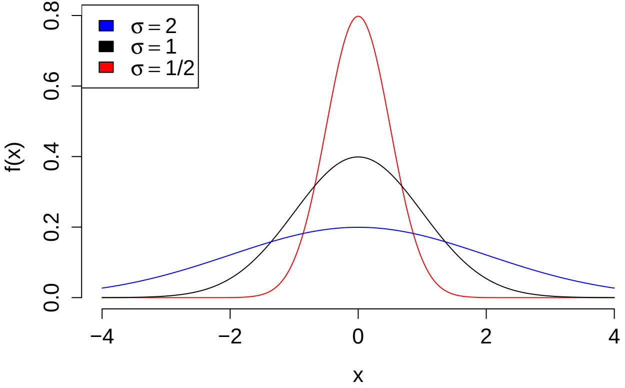


Figure: Blue $\sigma^2 = 4$, Black $\sigma^2 = 1$, Red $\sigma^2 = 1/4$

Linear Function of Normal RV is a Normal RV

Suppose that $X \sim N(\mu, \sigma^2)$. Then if a and b constants,

$$\boxed{a + bX \sim N(a + b\mu, b^2\sigma^2)}$$

Important

- For *any* RV X, $E[a + bX] = a + bE[X]$ and $Var(a + bX) = b^2 Var(X)$.
- Key point: linear transformation of normal is still normal!
- Linear transformation of Binomial is *not* Binomial!

Example

Suppose $X \sim N(\mu, \sigma^2)$ and let $Z = (X - \mu)/\sigma$. What is the distribution of Z?

(a) $N(\mu, \sigma^2)$

(b) $N(\mu, \sigma)$

(c) $N(0, \sigma^2)$

(d) $N(0, \sigma)$

(e) $N(0, 1)$

Linear Combinations of *Multiple Independent* Normals

Let $X \sim N(\mu_x, \sigma_x^2)$ independent of $Y \sim N(\mu_y, \sigma_y^2)$. Then if a, b, c are constants:

$$\boxed{aX + bY + c \sim N(a\mu_x + b\mu_y + c, a^2\sigma_x^2 + b^2\sigma_y^2)}$$

Important

- Result assumes independence
- Particular to Normal RV
- Extends to more than two Normal RVs

Suppose $X_1, X_2, \sim$ iid $N(\mu, \sigma^2)$

Let $\bar{X} = (X_1 + X_2)/2$. What is the distribution of $\bar{X}$?

(a) $N(\mu, \sigma^2/2)$

(b) $N(0, 1)$

(c) $N(\mu, \sigma^2)$

(d) $N(\mu, 2\sigma^2)$

(e) $N(2\mu, 2\sigma^2)$

Where does the Empirical Rule come from?

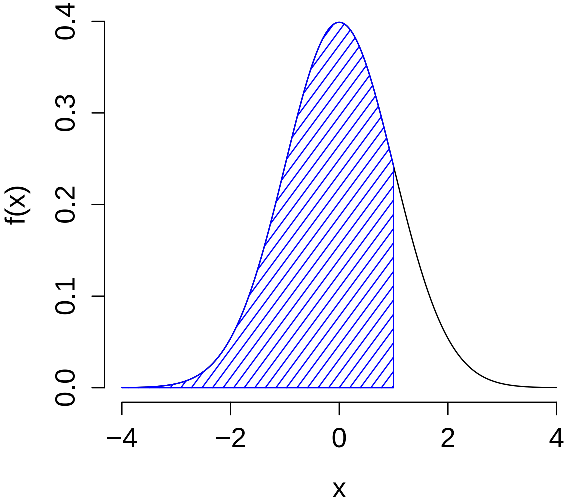
Empirical Rule

Approximately 68% of observations within $\mu \pm \sigma$

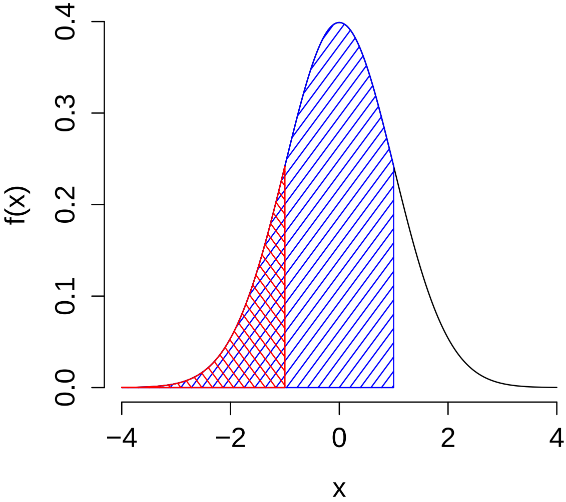
Approximately 95% of observations within $\mu \pm 2\sigma$

Nearly all observations within $\mu \pm 3\sigma$

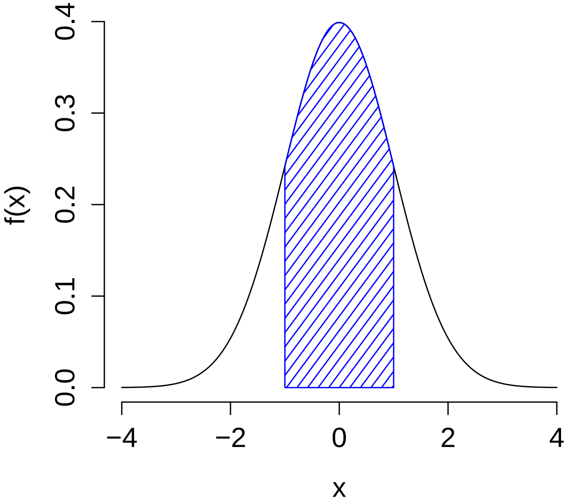
pnorm(1) ≈ 0.84



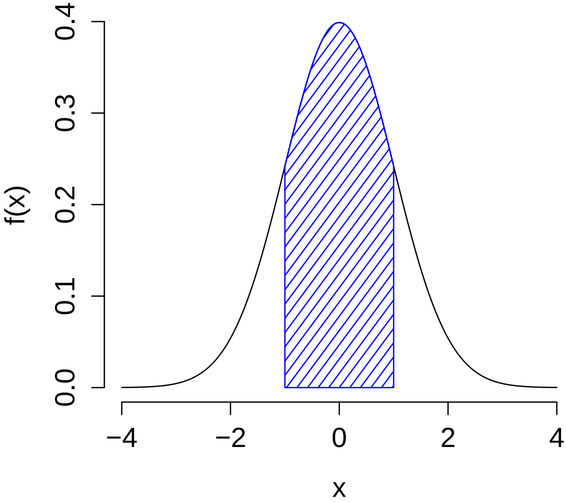
`pnorm(1) - pnorm(-1)` $\approx 0.84 - 0.16$



pnorm(1) - pnorm(-1) ≈ 0.68



Middle 68% of $N(0,1) \Rightarrow$ approx. $(-1,1)$



Suppose $X \sim N(0, 1)$

$$\begin{aligned} P(-1 \leq X \leq 1) &= \texttt{pnorm(1) - pnorm(-1)} \\ &\approx 0.683 \end{aligned}$$

$$\begin{aligned} P(-2 \leq X \leq 2) &= \texttt{pnorm(2) - pnorm(-2)} \\ &\approx 0.954 \end{aligned}$$

$$\begin{aligned} P(-3 \leq X \leq 3) &= \texttt{pnorm(3) - pnorm(-3)} \\ &\approx 0.997 \end{aligned}$$

What if $X \sim N(\mu, \sigma^2)$?

$$P(X \leq a) \quad =$$

What if $X \sim N(\mu, \sigma^2)$?

$$
\begin{aligned}
P(X \leq a) &= P(X - \mu \leq a - \mu) \\
&=
\end{aligned}
$$

What if $X \sim N(\mu, \sigma^2)$?

$$
\begin{aligned}
P(X \leq a) &= P(X - \mu \leq a - \mu) \\
&= P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) \\
&=
\end{aligned}
$$

What if $X \sim N(\mu, \sigma^2)$?

$$
\begin{aligned}
P(X \leq a) &= P(X - \mu \leq a - \mu) \\
&= P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) \\
&= P\left(Z \leq \frac{a - \mu}{\sigma}\right)
\end{aligned}
$$

Where Z is a standard normal random variable, i.e. $N(0, 1)$.

Which of these equals $P\left(Z \leq (a-\mu)/\sigma\right)$ if $Z \sim N(0,1)$?

(a) `pnorm(a)`

(b) $1 -$ `pnorm(a)`

(c) `pnorm(a)`$/\sigma - \mu$

(d) `pnorm`$\left(\frac{a-\mu}{\sigma}\right)$

(e) None of the above.

Probability *Above* a Threshold: $X \sim N(\mu, \sigma^2)$

$$
\begin{aligned}
P(X \geq b) &= 1 - P(X \leq b) = 1 - P\left(\frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\
&= 1 - P\left(Z \leq \frac{b-\mu}{\sigma}\right) \\
&= 1 - \texttt{pnorm}((b-\mu)/\sigma)
\end{aligned}
$$

Where Z is a standard normal random variable.

Probability of an Interval: $X \sim N(\mu, \sigma^2)$

$$
\begin{aligned}
P(a \leq X \leq b) &= P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\
&= P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) \\
&= \texttt{pnorm}((b-\mu)/\sigma) - \texttt{pnorm}((a-\mu)/\sigma)
\end{aligned}
$$

Where Z is a standard normal random variable.

Suppose $X \sim N(\mu, \sigma^2)$

What is $P(\mu - \sigma \leq X \leq \mu + \sigma)$?

Suppose $X \sim N(\mu, \sigma^2)$

What is $P(\mu - \sigma \leq X \leq \mu + \sigma)$?

$$\begin{aligned}
P(\mu - \sigma \leq X \leq \mu + \sigma) &= P\left(-1 \leq \frac{X-\mu}{\sigma} \leq 1\right) \\
&= P\left(-1 \leq Z \leq 1\right) \\
&= \texttt{pnorm(1)} - \texttt{pnorm(-1)} \\
&\approx 0.68
\end{aligned}$$

Percentiles/Quantiles for Continuous RVs

Quantile Function $Q(p)$ is the inverse of CDF $F(x_0)$

Plug in a probability p, get out the value of x_0 such that $F(x_0) = p$

$$Q(p) = F^{-1}(p)$$

In other words:

$$Q(p) = \text{the value of } x_0 \text{ such that } \int_{-\infty}^{x_0} f(x)\,dx = p$$

Inverse exists as long as $F(x_0)$ is *strictly increasing*.

Example: Median

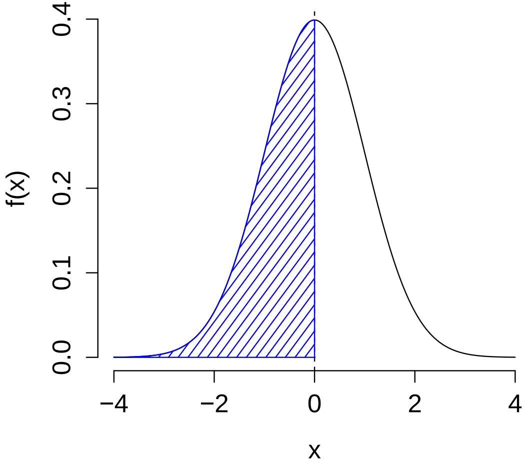
The median of a continuous random variable is $Q(0.5)$, i.e. the value of x_0 such that

$$\int_{-\infty}^{x_0} f(x)\, dx = 1/2$$

What is the median of a standard normal RV?

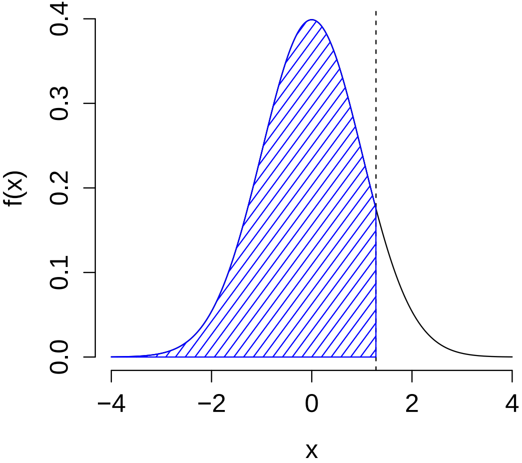
What is the median of a standard normal RV?

By symmetry, $Q(0.5) = 0$. R command: `qnorm()`



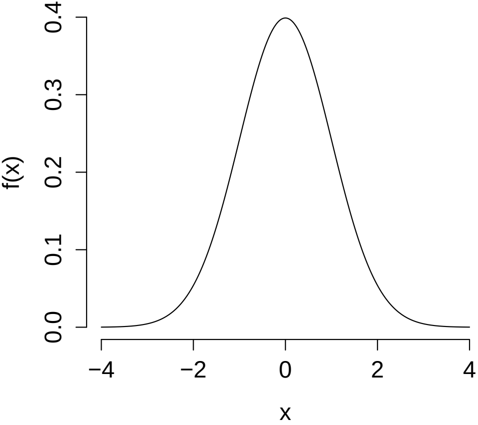
90th Percentile of a Standard Normal

qnorm(0.9) ≈ 1.28



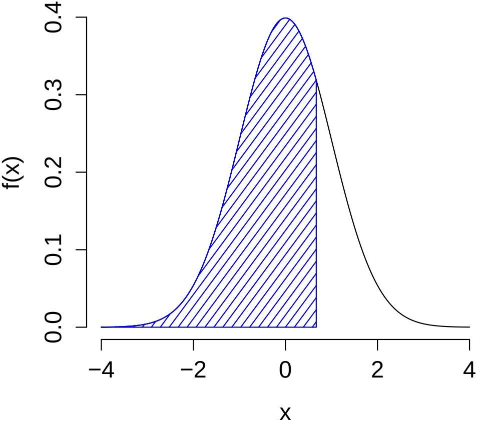
Using Quantile Function to find Symmetric Intervals

Suppose X is a standard normal RV. What is the value of c such that $P(-c \leq X \leq c) = 0.5$?



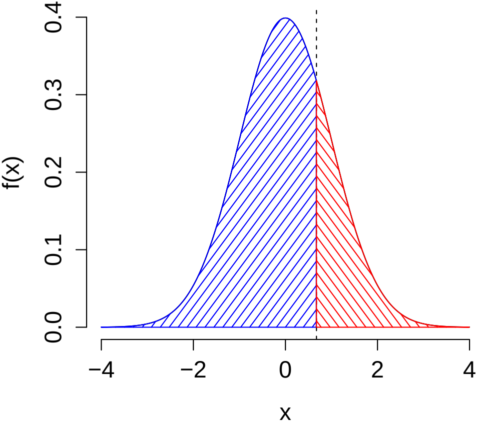
`qnorm(0.75)` ≈ 0.67

Suppose X is a standard normal RV. What is the value of c such that $P(-c \leq X \leq c) = 0.5$?



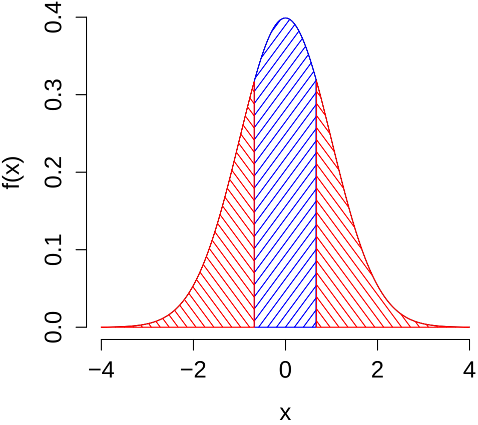
`qnorm(0.75)` ≈ 0.67

Suppose X is a standard normal RV. What is the value of c such that $P(-c \leq X \leq c) = 0.5$?



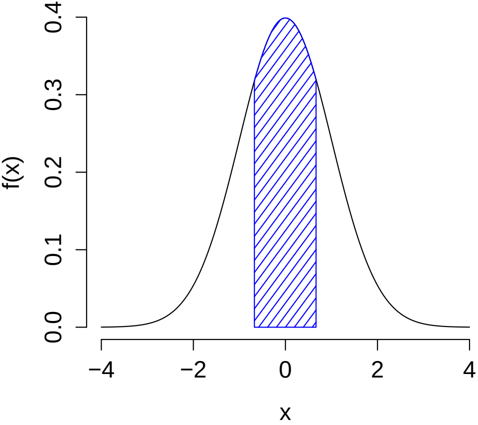
`pnorm(0.67)-pnorm(-0.67)`≈?

Suppose X is a standard normal RV. What is the value of c such that $P(-c \leq X \leq c) = 0.5$?



pnorm(0.67)-pnorm(-0.67)$\approx$ 0.5

Suppose X is a standard normal RV. What is the value of c such that $P(-c \leq X \leq c) = 0.5$?



95% Central Interval for Standard Normal

Suppose X is a standard normal random variable. What value of c ensures that $P(-c \leq X \leq c) \approx 0.95$?

R Commands for *Arbitrary* Normal RVs

Let $X \sim N(\mu, \sigma^2)$. Then we can use R to evaluate the CDF and Quantile function of X as follows:

CDF $F(x)$	`pnorm(x, mean =` μ`, sd =` σ`)`
Quantile Function $Q(p)$	`qnorm(p, mean =` μ`, sd =` σ`)`

Notice that this means you don't have to transform X to a standard normal in order to find areas under its pdf using R.

Example from Homework: $X \sim N(0, 16)$

One Way:

$$\begin{aligned} P(X \geq 10) &= 1 - P(X \leq 10) = 1 - P(X/4 \leq 10/4) \\ &= 1 - P(Z \leq 2.5) = 1 - \Phi(2.5) = 1 - \texttt{pnorm(2.5)} \\ &\approx 0.006 \end{aligned}$$

Example from Homework: $X \sim N(0, 16)$

One Way:

$$\begin{aligned} P(X \geq 10) &= 1 - P(X \leq 10) = 1 - P(X/4 \leq 10/4) \\ &= 1 - P(Z \leq 2.5) = 1 - \Phi(2.5) = 1 - \texttt{pnorm(2.5)} \\ &\approx 0.006 \end{aligned}$$

An Easier Way:

$$\begin{aligned} P(X \geq 10) &= 1 - P(X \leq 10) \\ &= 1 - \texttt{pnorm(10, mean = 0, sd = 4)} \\ &\approx 0.006 \end{aligned}$$

Lecture #13 – Sampling Distributions and Estimation I

Candy Weighing Experiment

Random Sampling Redux

Sampling Distributions

Estimator versus Estimate

Unbiasedness of Sample Mean

Standard Error of the Mean

Weighing a Random Sample

Bag Contains 100 Candies

Estimate total weight of candies by weighing a random sample of size 5 and multiplying the result by 20.

Your Chance to Win

The bag of candies and a digital scale will make their way around the room during the lecture. Each student gets a chance to draw 5 candies and weigh them.

Student with closest estimate wins the bag of candy!

Weighing a Random Sample

Procedure

When the bag and scale reach you, do the following:

1. Fold the top of the bag over and shake to randomize.
2. Randomly draw 5 candies without replacement.
3. Weigh your sample and record the result in grams along with your name on the sign-up sheet.
4. Replace your sample and shake again to re-randomize.
5. Pass bag and scale to next person.

Sampling and Estimation

Questions to Answer

1. How accurately do sample statistics estimate population parameters?
2. How can we quantify the uncertainty in our estimates?
3. What's so good about random sampling?

Random Sample

In Words

Select sample of n objects from population so that:

1. Each member of the population has the same probability of being selected
2. The fact that one individual is selected does not affect the chance that any other individual is selected
3. Each sample of size n is equally likely to be selected

In Math

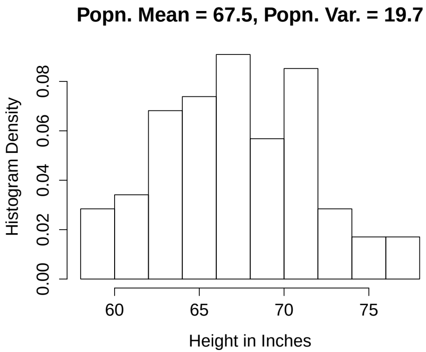
$X_1, X_2, \ldots, X_n \sim$ iid $f(x)$ if continuous

$X_1, X_2, \ldots, X_n \sim$ iid $p(x)$ if discrete

Random Sample Means *Sample With Replacement*

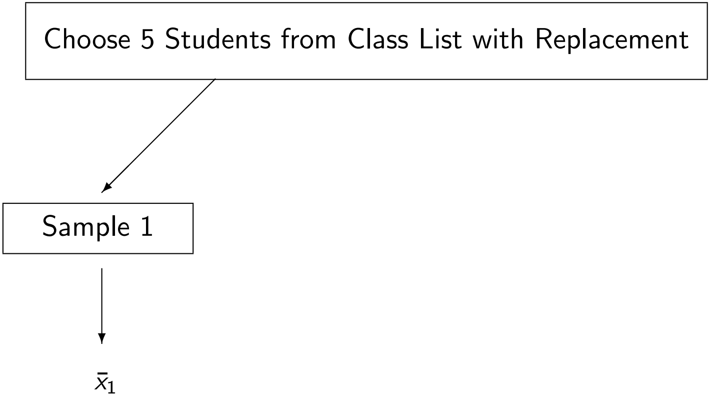
- Without replacement $\Rightarrow$ dependence between samples
- Sample small relative to popn. $\Rightarrow$ dependence negligible.
- This means our candy experiment (in progress) isn't bogus.

Example: Sampling from Econ 103 Class List

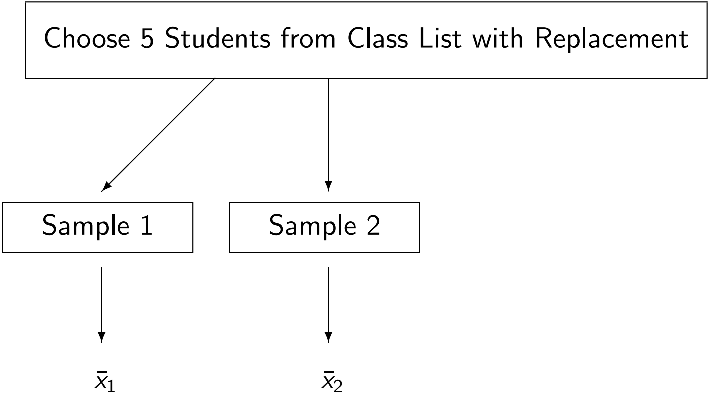


Use R to illustrate the in an example where we *know* the population. Can't do this in the real applications, but simulate it on the computer...

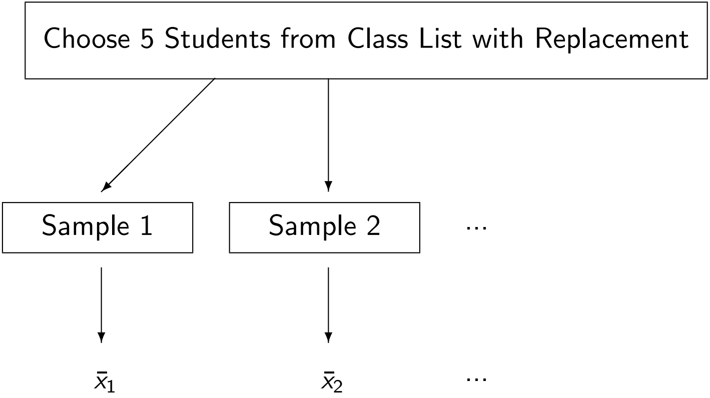
Sampling Distribution of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$



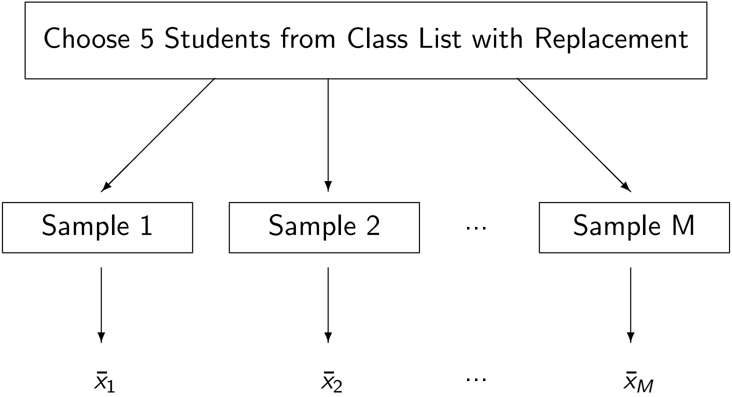
Sampling Distribution of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$



Sampling Distribution of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

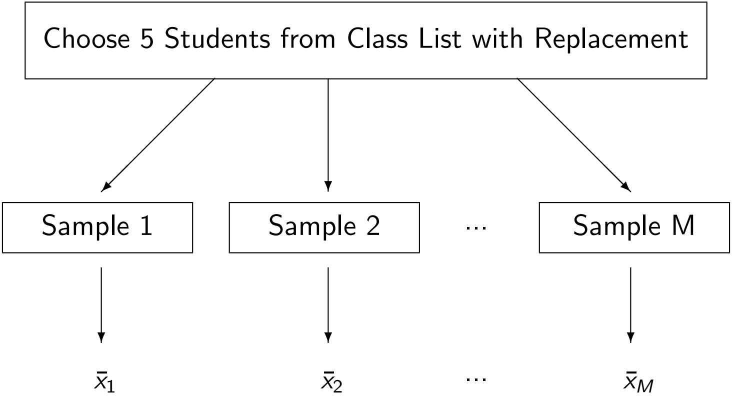


Sampling Distribution of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$



Repeat M times $\rightarrow$ get M different sample means

Sampling Distribution of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$



Repeat M times $\rightarrow$ get M different sample means

Sampling Dist: relative frequencies of the $\bar{x}_i$ when $M = \infty$

Height of Econ 103 Students

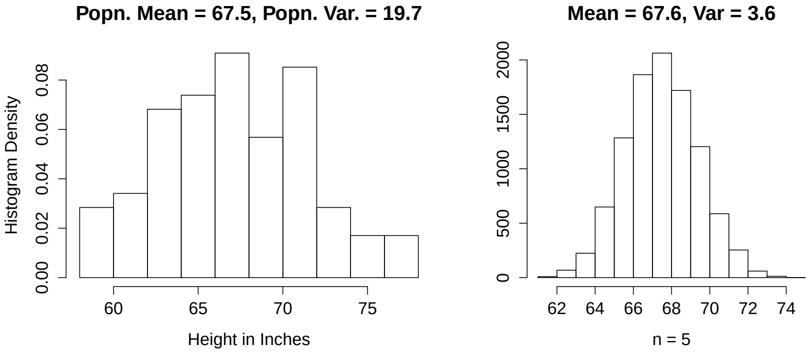
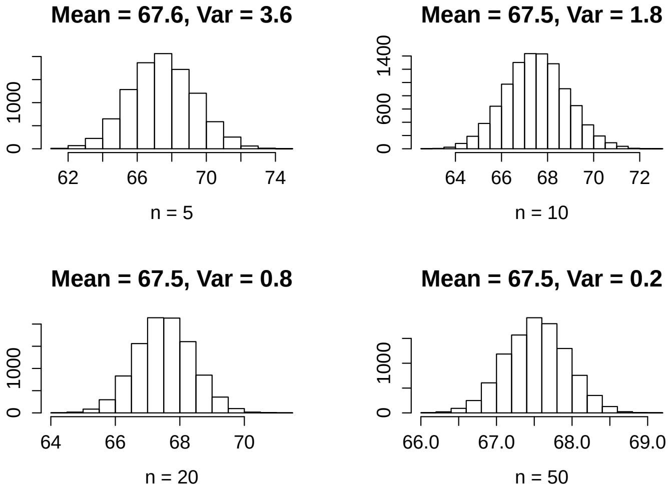


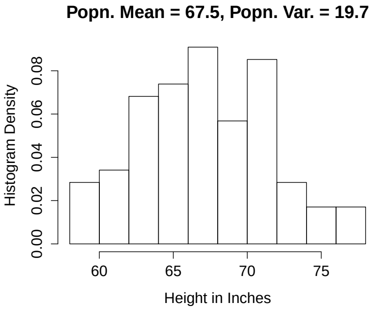
Figure: Left: Population, Right: Sampling distribution of $\bar{X}_5$

Histograms of sampling distribution of sample mean $\bar{X}_n$

Random Sampling With Replacement, 10000 Reps. Each



Population Distribution vs. Sampling Distribution of $\bar{X}_n$



	Sampling Dist. of $\bar{X}_n$	
n	Mean	Variance
5	67.6	3.6
10	67.5	1.8
20	67.5	0.8
50	67.5	0.2

Things to Notice:

1. Sampling dist. "correct on average"
2. Sampling variability decreases with n
3. Sampling dist. bell-shaped even though population isn't!

Step 1: Population as RV rather than List of Objects

Old Way	New Way
In the 2016 election, 65,853,625 out of 137,100,229 voters voted for Hillary Clinton	Bernoulli($p = 0.48$) RV

Old Way	New Way
List of heights for 97 million US adult males with mean 69 in and std. dev. 6 in	$N(\mu = 69, \sigma^2 = 36)$ RV

Second example assumes distribution of height is bell-shaped.

Step 2: iid RVs Represent Random Sampling from Popn.

Hillary Voters Example

Poll random sample of 1000 people who voted in 2016:

$$X_1, \ldots, X_{1000} \sim \text{ iid Bernoulli}(p = 0.48)$$

Height Example

Measure the heights of random sample of 50 US males:

$$Y_1, \ldots, Y_{50} \sim \text{ iid } N(\mu = 69, \sigma^2 = 36)$$

Key Question

What do the properties of the population imply about the properties of the sample?

What does the population imply about the sample?

Suppose that exactly half of US voters chose Hillary Clinton in the 2016 election. If you poll a random sample of 4 voters, what is the probability that *exactly half* were Hillary supporters?

What does the population imply about the sample?

Suppose that exactly half of US voters chose Hillary Clinton in the 2016 election. If you poll a random sample of 4 voters, what is the probability that *exactly half* were Hillary supporters?

$$\binom{4}{2}(1/2)^2(1/2)^2 = 3/8 = 0.375$$

The rest of the probabilities...

Suppose that exactly half of US voters plan to vote for Hillary Clinton and we poll a random sample of 4 voters.

$$
\begin{aligned}
P\,(\text{Exactly 0 Hillary Voters in the Sample}) &= 0.0625 \\
P\,(\text{Exactly 1 Hillary Voters in the Sample}) &= 0.25 \\
P\,(\text{Exactly 2 Hillary Voters in the Sample}) &= 0.375 \\
P\,(\text{Exactly 3 Hillary Voters in the Sample}) &= 0.25 \\
P\,(\text{Exactly 4 Hillary Voters in the Sample}) &= 0.0625
\end{aligned}
$$

You should be able to work these out yourself. If not, review the lecture slides on the Binomial RV.

Population Size is Irrelevant Under Random Sampling

Crucial Point

None of the preceding calculations involved the population size: I didn't even tell you what it was! We'll never talk about population size again in this course.

Why?

Draw with replacement $\implies$ only the sample size and the *proportion* of Hillary supporters in the population matter.

(Sample) Statistic

Any function of the data *alone*, e.g. sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Typically used to estimate an unknown population parameter: e.g. $\bar{x}$ is an estimate of μ.

Step 3: Random Sampling $\Rightarrow$ *Sample Statistics* are RVs

This is *the crucial point of the course*: if we draw a random sample, the dataset we get is random. Since a statistic is a function of the data, it is a random variable!

A Sample Statistic in the Polling Example

Suppose that exactly half of voters in the population support Hillary Clinton and we poll a random sample of 4 voters. If we code Hillary supporters as "1" and everyone else as "0" then what are the possible values of the sample mean in our dataset?

(a) $(0, 1)$

(b) $\{0, 0.25, 0.5, 0.75, 1\}$

(c) $\{0, 1, 2, 3, 4\}$

(d) $(-\infty, \infty)$

(e) Not enough information to determine.

Sampling Distribution

Under random sampling, a statistic is a RV so it has a PDF if continuous or PMF if discrete: this is its sampling distribution.

Sampling Dist. of Sample Mean in Polling Example

$$
\begin{aligned}
p(0) &= 0.0625 \\
p(0.25) &= 0.25 \\
p(0.5) &= 0.375 \\
p(0.75) &= 0.25 \\
p(1) &= 0.0625
\end{aligned}
$$

Contradiction? No, but we need better terminology...

- Under random sampling, a statistic is a RV
- Given dataset is *fixed* so statistic is a *constant number*
- Distinguish between: Estimator vs. Estimate

Estimator

Description of a general procedure.

Estimate

Particular result obtained from applying the procedure.

$\bar{X}_n$ is an Estimator = Procedure = Random Variable

1. Take a random sample: $X_1, \ldots, X_n$
2. Average what you get: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$

$\bar{X}_n$ is an Estimator = Procedure = Random Variable

1. Take a random sample: $X_1, \ldots, X_n$
2. Average what you get: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$

$\bar{x}$ is an Estimate = Result of Procedure = Constant

- Result of taking a random sample was the dataset: $x_1, \ldots, x_n$
- Result of averaging the observed data was $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$

$\bar{X}_n$ is an Estimator = Procedure = Random Variable

1. Take a random sample: $X_1, \ldots, X_n$
2. Average what you get: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\bar{x}$ is an Estimate = Result of Procedure = Constant

- Result of taking a random sample was the dataset: $x_1, \ldots, x_n$
- Result of averaging the observed data was $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Sampling Distribution of $\bar{X}_n$

Thought experiment: suppose I were to repeat the procedure of taking the mean of a random sample over and over forever. What relative frequencies would I get for the sample means?

Mean of Sampling Distribution of $\bar{X}_n$

$X_1, \ldots, X_n \sim$ iid with mean μ

$$E[\bar{X}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$$

Mean of Sampling Distribution of $\bar{X}_n$

$X_1, \ldots, X_n \sim$ iid with mean μ

$$E[\bar{X}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] =$$

Mean of Sampling Distribution of $\bar{X}_n$

$X_1, \ldots, X_n \sim$ iid with mean μ

$$E[\bar{X}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} \mu =$$

Mean of Sampling Distribution of $\bar{X}_n$

$X_1, \ldots, X_n \sim$ iid with mean μ

$$E[\bar{X}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{n\mu}{n} = \mu$$

Hence, sample mean is "correct on average." The formal term for this is *unbiased*.

Variance of Sampling Distribution of $\bar{X}_n$

$X_1, \ldots, X_n \sim$ iid with mean μ and variance σ^2

$$Var[\bar{X}_n] \quad = \quad Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$$

Variance of Sampling Distribution of $\bar{X}_n$

$X_1, \ldots, X_n \sim$ iid with mean μ and variance σ^2

$$
\begin{aligned}
Var[\bar{X}_n] &= Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \\
&=
\end{aligned}
$$

Variance of Sampling Distribution of $\bar{X}_n$

$X_1, \ldots, X_n \sim$ iid with mean μ and variance σ^2

$$\begin{aligned} Var[\bar{X}_n] &= Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \end{aligned}$$

Variance of Sampling Distribution of $\bar{X}_n$

$X_1, \ldots, X_n \sim$ iid with mean μ and variance σ^2

$$\begin{aligned} Var[\bar{X}_n] &= Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

Hence the variance of the sample mean *decreases linearly with sample size*.

Standard Error

Std. Dev. of estimator's sampling dist. is called standard error.

Standard Error of the Sample Mean

$$SE(\bar{X}_n) = \sqrt{Var\left(\bar{X}_n\right)} = \sqrt{\sigma^2/n} = \sigma/\sqrt{n}$$

Lecture #14 – Sampling Distributions and Estimation II

Bias of an Estimator

Why divide by $n-1$ in sample variance?

Biased Sampling and the Candy-Weighing Experiment

Efficiency: Choosing between Unbiased Estimators

Mean-Squared Error: Choosing Between Biased Estimators

Consistency and the Law of Large Numbers

Unbiased means "Right on Average"

Bias of an Estimator

Let $\widehat{\theta}_n$ be a sample estimator of a population parameter θ_0. The *bias* of $\widehat{\theta}_n$ is $E[\widehat{\theta}_n] - \theta_0$.

Unbiased Estimator

A sample estimator $\widehat{\theta}_n$ of a population parameter θ_0 is called *unbiased* if $E[\widehat{\theta}_n] = \theta_0$

Why $(n-1)$ for sample variance?

We will show that having $n-1$ in the denominator ensures:

$$E[S^2] = E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] = \sigma^2$$

under random sampling.

Why $(n-1)$ for sample variance?

Step # 1 – Tedious but straightforward algebra gives:

$$\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \left[\sum_{i=1}^{n}(X_i - \mu)^2\right] - n(\bar{X} - \mu)^2$$

You are not responsible for proving Step #1 on an exam.

$$
\begin{aligned}
\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 &= \sum_{i=1}^{n}\left(X_i-\mu+\mu-\bar{X}\right)^2=\sum_{i=1}^{n}\left[(X_i-\mu)-(\bar{X}-\mu)\right]^2 \\
&= \sum_{i=1}^{n}\left[(X_i-\mu)^2-2(X_i-\mu)(\bar{X}-\mu)+(\bar{X}-\mu)^2\right] \\
&= \sum_{i=1}^{n}(X_i-\mu)^2-\sum_{i=1}^{n}2(X_i-\mu)(\bar{X}-\mu)+\sum_{i=1}^{n}(\bar{X}-\mu)^2 \\
&= \left[\sum_{i=1}^{n}(X_i-\mu)^2\right]-2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\mu)+n(\bar{X}-\mu)^2 \\
&= \left[\sum_{i=1}^{n}(X_i-\mu)^2\right]-2(\bar{X}-\mu)\left(\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}\mu\right)+n(\bar{X}-\mu)^2 \\
&= \left[\sum_{i=1}^{n}(X_i-\mu)^2\right]-2(\bar{X}-\mu)(n\bar{X}-n\mu)+n(\bar{X}-\mu)^2 \\
&= \left[\sum_{i=1}^{n}(X_i-\mu)^2\right]-2n(\bar{X}-\mu)^2+n(\bar{X}-\mu)^2 \\
&= \left[\sum_{i=1}^{n}(X_i-\mu)^2\right]-n(\bar{X}-\mu)^2
\end{aligned}
$$

Why $(n-1)$ for sample variance?

Step # 2 – Take Expectations of Step # 1:

$$
\begin{aligned}
E\left[\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right] &= E\left[\left\{\sum_{i=1}^{n}(X_i-\mu)^2\right\}-n(\bar{X}-\mu)^2\right] \\
&=
\end{aligned}
$$

Why $(n-1)$ for sample variance?

Step # 2 – Take Expectations of Step # 1:

$$
\begin{aligned}
E\left[\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right] &= E\left[\left\{\sum_{i=1}^{n}(X_i-\mu)^2\right\}-n(\bar{X}-\mu)^2\right] \\
&= E\left[\sum_{i=1}^{n}(X_i-\mu)^2\right]-E\left[n(\bar{X}-\mu)^2\right] \\
&=
\end{aligned}
$$

Why $(n-1)$ for sample variance?

Step # 2 – Take Expectations of Step # 1:

$$\begin{aligned}
E\left[\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right] &= E\left[\left\{\sum_{i=1}^{n}(X_i-\mu)^2\right\}-n(\bar{X}-\mu)^2\right] \\
&= E\left[\sum_{i=1}^{n}(X_i-\mu)^2\right]-E\left[n(\bar{X}-\mu)^2\right] \\
&= \sum_{i=1}^{n}E\left[(X_i-\mu)^2\right]-n\,E\left[(\bar{X}-\mu)^2\right]
\end{aligned}$$

Where we have used the linearity of expectation.

Why $(n-1)$ for sample variance?

Step # 3 – Use assumption of random sampling:

$X_1, \ldots, X_n \sim$ iid with mean μ and variance σ^2

$$E\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] = \sum_{i=1}^{n} E\left[(X_i - \mu)^2\right] - n\,E\left[(\bar{X} - \mu)^2\right]$$

$$=$$

Why $(n-1)$ for sample variance?

Step # 3 – Use assumption of random sampling:

$X_1, \ldots, X_n \sim$ iid with mean μ and variance σ^2

$$
\begin{aligned}
E\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] &= \sum_{i=1}^{n} E\left[\left(X_i - \mu\right)^2\right] - n\, E\left[(\bar{X} - \mu)^2\right] \\
&= \sum_{i=1}^{n} Var(X_i) - n\, E\left[(\bar{X} - E[\bar{X}])^2\right] \\
&=
\end{aligned}
$$

Why $(n-1)$ for sample variance?

Step # 3 – Use assumption of random sampling:

$X_1, \ldots, X_n \sim$ iid with mean μ and variance σ^2

$$
\begin{aligned}
E\left[\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2\right] &= \sum_{i=1}^{n} E\left[(X_i - \mu)^2\right] - n\, E\left[(\bar{X} - \mu)^2\right] \\
&= \sum_{i=1}^{n} Var(X_i) - n\, E\left[(\bar{X} - E[\bar{X}])^2\right] \\
&= \sum_{i=1}^{n} Var(X_i) - n\, Var(\bar{X}) = n\sigma^2 - \sigma^2 \\
&=
\end{aligned}
$$

Why $(n-1)$ for sample variance?

Step # 3 – Use assumption of random sampling:

$X_1, \ldots, X_n \sim$ iid with mean μ and variance σ^2

$$
\begin{aligned}
E\left[\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right] &= \sum_{i=1}^{n} E\left[(X_i-\mu)^2\right] - n\,E\left[(\bar{X}-\mu)^2\right] \\
&= \sum_{i=1}^{n} Var(X_i) - n\,E\left[(\bar{X}-E[\bar{X}])^2\right] \\
&= \sum_{i=1}^{n} Var(X_i) - n\,Var(\bar{X}) = n\sigma^2 - \sigma^2 \\
&= (n-1)\sigma^2
\end{aligned}
$$

Since we showed earlier today that $E[\bar{X}] = \mu$ and $Var(\bar{X}) = \sigma^2/n$ under this random sampling assumption.

Why $(n-1)$ for sample variance?

Finally – Divide Step # 3 by $(n-1)$:

$$E[S^2] \;=\; E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right]=\frac{(n-1)\sigma^2}{n-1}=\sigma^2$$

Hence, having $(n-1)$ in the denominator ensures that the sample variance is "correct on average," that is *unbiased*.

A Different Estimator of the Population Variance

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

A Different Estimator of the Population Variance

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

$$E[\widehat{\sigma}^2] = E\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] =$$

A Different Estimator of the Population Variance

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

$$E[\widehat{\sigma}^2] = E\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] = \frac{(n-1)\sigma^2}{n}$$

Bias of $\widehat{\sigma}^2$

$E[\widehat{\sigma}^2] - \sigma^2 =$

A Different Estimator of the Population Variance

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

$$E[\widehat{\sigma}^2] = E\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] = \frac{(n-1)\sigma^2}{n}$$

Bias of $\widehat{\sigma}^2$

$$E[\widehat{\sigma}^2] - \sigma^2 = \frac{(n-1)\sigma^2}{n} - \sigma^2 = \frac{(n-1)\sigma^2}{n} - \frac{n\sigma^2}{n} = -\sigma^2/n$$

How Large is the Average Family?

How many brothers and sisters are in your family, including yourself?

The average number of children per family was about 2.0 twenty years ago.

What's Going On Here?

What's Going On Here?

Biased Sample!

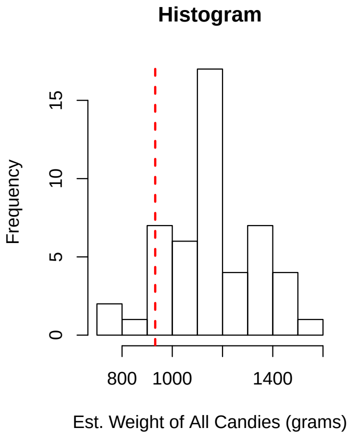
- Zero children $\Rightarrow$ didn't send any to college
- Sampling by *children* so large families oversampled

Candy Weighing: 49 Estimates, Each With $n = 5$

$$\widehat{\theta} = 20 \times (X_1 + \ldots + X_5)$$

Summary of Sampling Dist.	
Overestimates	45
Exactly Correct	0
Underestimates	4
$E[\hat{\theta}]$	1164 grams
$\mathrm{SD}(\widehat{\theta})$	189 grams

Actual Mass: θ_0 =932 grams



What was in the bag?

100 Candies Total:

- 20 Fun Size Snickers Bars (large)
- 30 Reese's Miniatures (medium)
- 50 Tootsie Roll "Midgees" (small)

So What Happened?

What was in the bag?

100 Candies Total:

- 20 Fun Size Snickers Bars (large)
- 30 Reese's Miniatures (medium)
- 50 Tootsie Roll "Midgees" (small)

So What Happened?

Not a random sample! The Snickers bars were *oversampled*.

Could we have avoided this? How?

Let $X_1, X_2, \ldots X_n \sim iid$ mean μ, variance σ^2 and define $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. True or False:

$\bar{X}_n$ *is an unbiased estimator of* μ

(a) True

(b) False

Let $X_1, X_2, \ldots X_n \sim iid$ mean μ, variance σ^2 and define $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. True or False:

$\bar{X}_n$ *is an unbiased estimator of* μ

(a) True

(b) False

TRUE!

Let $X_1, X_2, \dots X_n \sim iid$ mean μ, variance σ^2. True or False:

X_1 *is an unbiased estimator of* μ

(a) True

(b) False

Let $X_1, X_2, \ldots X_n \sim iid$ mean μ, variance σ^2. True or False:

X_1 is an unbiased estimator of μ

(a) True

(b) False

TRUE!

How to choose between two unbiased estimators?

Suppose $X_1, X_2, \ldots X_n \sim iid$ with mean μ and variance σ^2

$$\begin{aligned} E[\bar{X}_n] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu \\ E[X_1] &= \mu \end{aligned}$$

How to choose between two unbiased estimators?

Suppose $X_1, X_2, \ldots X_n \sim iid$ with mean μ and variance σ^2

$$\begin{aligned} E[\bar{X}_n] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu \\ E[X_1] &= \mu \\ Var(\bar{X}_n) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \sigma^2/n \end{aligned}$$

How to choose between two unbiased estimators?

Suppose $X_1, X_2, \ldots X_n \sim iid$ with mean μ and variance σ^2

$$\begin{aligned}
E[\bar{X}_n] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu \\
E[X_1] &= \mu \\
Var(\bar{X}_n) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \sigma^2/n \\
Var(X_1) &= \sigma^2
\end{aligned}$$

Efficiency - Compare Unbiased Estimators by Variance

Let $\widehat{\theta}_1$ and $\widehat{\theta}_2$ be unbiased estimators of θ_0. We say that $\widehat{\theta}_1$ is *more efficient* than $\widehat{\theta}_2$ if $Var(\widehat{\theta}_1) < Var(\widehat{\theta}_2)$.

Mean-Squared Error

Except in very simple situations, unbiased estimators are hard to come by. In fact, in many interesting applications there is a *tradeoff* between bias and variance:

- Low bias estimators often have a high variance
- Low variance estimators often have high bias

Mean-Squared Error

Except in very simple situations, unbiased estimators are hard to come by. In fact, in many interesting applications there is a *tradeoff* between bias and variance:

- Low bias estimators often have a high variance
- Low variance estimators often have high bias

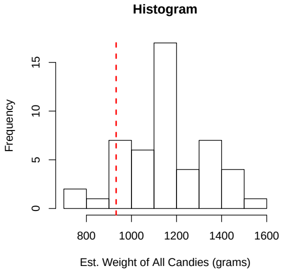
Mean-Squared Error (MSE): Squared Bias plus Variance

$$MSE(\widehat{\theta}) = \text{Bias}(\widehat{\theta})^2 + Var(\widehat{\theta})$$

Root Mean-Squared Error (RMSE): $\sqrt{\text{MSE}}$

Calculate MSE for Candy Experiment

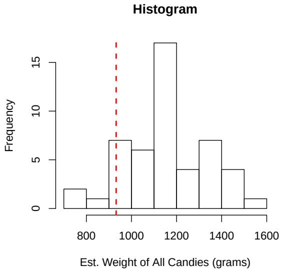
$E[\hat{\theta}]$	1164 grams
θ_0	932 grams
$SD(\widehat{\theta})$	189 grams



$$\text{Bias} =$$

Calculate MSE for Candy Experiment

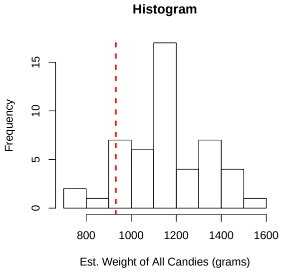
$E[\hat{\theta}]$	1164 grams
θ_0	932 grams
$SD(\widehat{\theta})$	189 grams



$$\begin{aligned} \text{Bias} &= 1164 \text{ grams} - 932 \text{ grams} \\ &= 232 \text{ grams} \\ \text{MSE} &= \end{aligned}$$

Calculate MSE for Candy Experiment

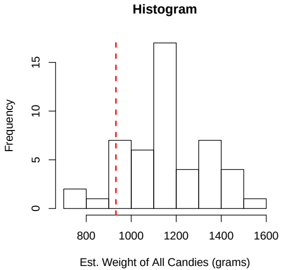
$E[\hat{\theta}]$	1164 grams
θ_0	932 grams
$SD(\widehat{\theta})$	189 grams



$$
\begin{aligned}
\text{Bias} &= 1164 \text{ grams} - 932 \text{ grams} \\
&= \textcolor{red}{232 \text{ grams}} \\
\text{MSE} &= \text{Bias}^2 + \text{Variance} \\
&= (232^2 + 189^2) \text{ grams}^2 \\
&=
\end{aligned}
$$

Calculate MSE for Candy Experiment

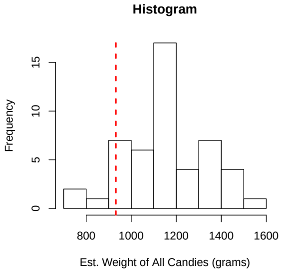
$E[\hat{\theta}]$	1164 grams
θ_0	932 grams
$SD(\widehat{\theta})$	189 grams



$$
\begin{aligned}
\text{Bias} &= 1164 \text{ grams} - 932 \text{ grams} \\
&= {\color{red}232 \text{ grams}} \\
\text{MSE} &= \text{Bias}^2 + \text{Variance} \\
&= (232^2 + 189^2) \text{ grams}^2 \\
&= 8.9545 \times 10^4 \text{ grams}^2
\end{aligned}
$$

Calculate MSE for Candy Experiment

$E[\hat{\theta}]$	1164 grams
θ_0	932 grams
$SD(\widehat{\theta})$	189 grams



$$
\begin{aligned}
\text{Bias} &= 1164 \text{ grams} - 932 \text{ grams} \\
&= {\color{red} 232 \text{ grams}} \\
\text{MSE} &= \text{Bias}^2 + \text{Variance} \\
&= (232^2 + 189^2) \text{ grams}^2 \\
&= 8.9545 \times 10^4 \text{ grams}^2 \\
\text{RMSE} &= \sqrt{\text{MSE}} = 299 \text{ grams}
\end{aligned}
$$

Finite Sample versus Asymptotic Properties of Estimators

Finite Sample Properties

For *fixed sample size n* what are the properties of the sampling distribution of $\widehat{\theta}_n$? (E.g. bias and variance.)

Asymptotic Properties

What happens to the sampling distribution of $\widehat{\theta}_n$ *as the sample size n gets larger and larger?* (That is, $n \to \infty$).

Why Asymptotics?

Law of Large Numbers

Make precise what we mean by "bigger samples are better."

Central Limit Theorem

As $n \to \infty$ *pretty much any* sampling distribution is well-approximated by a normal random variable!

Consistency

Consistency

If an estimator $\widehat{\theta}_n$ (which is a RV) *converges* to θ_0 (a constant) as $n \to \infty$, we say that *$\widehat{\theta}_n$ is consistent for θ_0*.

What does it mean for a *RV* to converge to a *constant?*

For this course we'll use *MSE Consistency*:

$$\lim_{n \to \infty} \text{MSE}(\widehat{\theta}_n) = 0$$

This makes sense since $\text{MSE}(\widehat{\theta}_n)$ is a *constant*, so this is just an ordinary limit from calculus.

Law of Large Numbers (aka Law of Averages)

Let $X_1, X_2, \ldots X_n \sim iid$ mean μ, variance σ^2. Then the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is consistent for the population mean μ.

Law of Large Numbers (aka Law of Averages)

Let $X_1, X_2, \ldots X_n \sim iid$ mean μ, variance σ^2.

$$\begin{aligned} E[\bar{X}_n] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu \\ Var(\bar{X}_n) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \sigma^2/n \end{aligned}$$

Law of Large Numbers (aka Law of Averages)

Let $X_1, X_2, \ldots X_n \sim iid$ mean μ, variance σ^2.

$$
\begin{aligned}
E[\bar{X}_n] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu \\
Var(\bar{X}_n) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \sigma^2/n \\
\text{MSE}(\bar{X}_n) &= \text{Bias}(\bar{X}_n)^2 + Var(\bar{X}_n)
\end{aligned}
$$

Law of Large Numbers (aka Law of Averages)

Let $X_1, X_2, \ldots X_n \sim iid$ mean μ, variance σ^2.

$$
\begin{aligned}
E[\bar{X}_n] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu \\
Var(\bar{X}_n) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \sigma^2/n \\
\text{MSE}(\bar{X}_n) &= \text{Bias}(\bar{X}_n)^2 + Var(\bar{X}_n) \\
&= \left(E[\bar{X}_n] - \mu\right)^2 + Var(\bar{X}_n)
\end{aligned}
$$

Law of Large Numbers (aka Law of Averages)

Let $X_1, X_2, \ldots X_n \sim iid$ mean μ, variance σ^2.

$$
\begin{aligned}
E[\bar{X}_n] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu \\
Var(\bar{X}_n) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \sigma^2/n \\
\text{MSE}(\bar{X}_n) &= \text{Bias}(\bar{X}_n)^2 + Var(\bar{X}_n) \\
&= \big(E[\bar{X}_n] - \mu\big)^2 + Var(\bar{X}_n) \\
&= 0 + \sigma^2/n \\
&\to 0
\end{aligned}
$$

Hence $\bar{X}_n$ is consistent for μ

Important!

An estimator *can* be biased but still consistent, as long as the bias disappears as $n \to \infty$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

Bias of $\widehat{\sigma}^2$

$E[\widehat{\sigma}^2] - \sigma^2 =$

Important!

An estimator *can* be biased but still consistent, as long as the bias disappears as $n \to \infty$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

Bias of $\widehat{\sigma}^2$

$$E[\widehat{\sigma}^2] - \sigma^2 = \frac{(n-1)\sigma^2}{n} - \sigma^2 =$$

Important!

An estimator *can* be biased but still consistent, as long as the bias disappears as $n \to \infty$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

Bias of $\widehat{\sigma}^2$

$$E[\widehat{\sigma}^2] - \sigma^2 = \frac{(n-1)\sigma^2}{n} - \sigma^2 = -\sigma^2/n$$

Important!

An estimator *can* be biased but still consistent, as long as the bias disappears as $n \to \infty$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

Bias of $\widehat{\sigma}^2$

$$E[\widehat{\sigma}^2] - \sigma^2 = \frac{(n-1)\sigma^2}{n} - \sigma^2 = -\sigma^2/n \to 0$$

Suppose $X_1, X_2, \ldots, X_n \sim$ iid $N(\mu, \sigma^2)$. What is the sampling distribution of $\bar{X}_n$?

(a) $N(0, 1)$

(b) $N(\mu, \sigma^2/n)$

(c) $N(\mu, \sigma^2)$

(d) $N(\mu/n, \sigma^2/n)$

(e) $N(n\mu, n\sigma^2)$

But still, how can something random converge to something constant?

Sampling Distribution of $\bar{X}_n$ Collapses to μ

Look at an example where we can directly calculate not only the mean and variance of the sampling distribution of $\bar{X}_n$, but the *sampling distribution itself*:

$$X_1, X_2, \ldots, X_n \sim \text{iid } N(\mu, \sigma^2) \Rightarrow \bar{X}_n \sim N(\mu, \sigma^2/n)$$

Sampling Distribution of $\bar{X}_n$ Collapses to μ

$X_1, X_2, \ldots, X_n \sim$ iid $N(\mu, \sigma^2 \Rightarrow \bar{X}_n \sim N(\mu, \sigma^2/n)$.

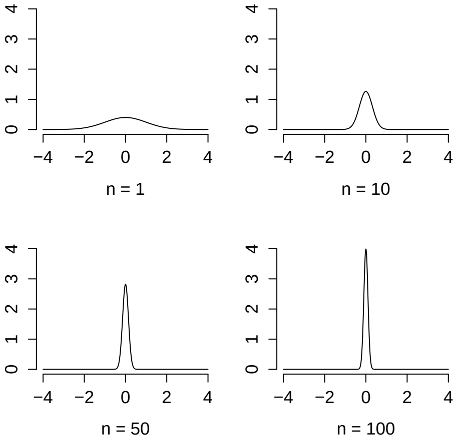


Figure: Sampling Distributions for $\bar{X}_n$ where $X_i \sim$ iid $N(0, 1)$

Another Visualization: Keep Adding Observations

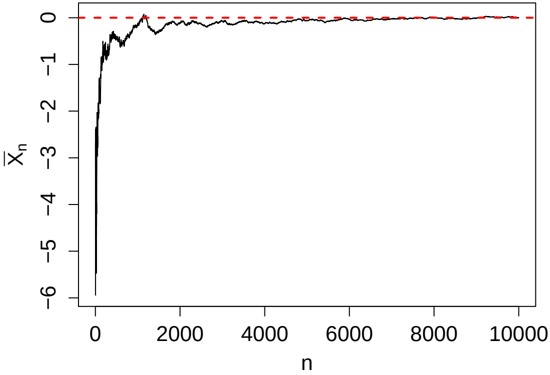


Figure: Running sample means: $X_i \sim$ iid $N(0, 100)$

n	$\bar{X}_n$
1	-2.69
2	-3.18
3	-5.94
4	-4.27
5	-2.62
10	-2.89
20	-5.33
50	-2.94
100	-1.58
500	-0.45
1000	-0.13
5000	-0.05
10000	0.00

Lecture #15 – Confidence Intervals I

Confidence Interval for Mean of Normal Population (σ^2 Known)

Interpreting a Confidence Interval

Margin of Error and Width

Today – Simplest Example of a Confidence Interval

- Suppose the population is $N(\mu, \sigma^2)$
- We know σ^2 but not μ
- Draw random sample $X_1, X_2, \ldots, X_n \sim$ iid $N(\mu, \sigma^2)$

Today – Simplest Example of a Confidence Interval

- Suppose the population is $N(\mu, \sigma^2)$
- We know σ^2 but not μ
- Draw random sample $X_1, X_2, \ldots, X_n \sim$ iid $N(\mu, \sigma^2)$
- Observe value of sample mean $\bar{x}_n$ (e.g. 69 inches)

Today – Simplest Example of a Confidence Interval

- Suppose the population is $N(\mu, \sigma^2)$
- We know σ^2 but not μ
- Draw random sample $X_1, X_2, \ldots, X_n \sim$ iid $N(\mu, \sigma^2)$
- Observe value of sample mean $\bar{x}_n$ (e.g. 69 inches)
- What is a plausible range for μ?

Today – Simplest Example of a Confidence Interval

- Suppose the population is $N(\mu, \sigma^2)$
- We know σ^2 but not μ
- Draw random sample $X_1, X_2, \ldots, X_n \sim$ iid $N(\mu, \sigma^2)$
- Observe value of sample mean $\bar{x}_n$ (e.g. 69 inches)
- What is a plausible range for μ?
- How confident are we? Can we make this precise?

Next time we'll look at more realistic and interesting examples...

Suppose $X_1, X_2, \ldots, X_n \sim$ iid $N(\mu, \sigma^2)$. What is the sampling distribution of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$?

(a) $N(\mu, \sigma^2)$

(b) $N(0, 1)$

(c) $N(0, \sigma)$

(d) $N(\mu, 1)$

(e) Not enough information to determine.

$X_1, X_2, \ldots, X_n \sim$ iid $N(\mu, \sigma^2)$

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X}_n - E[\bar{X}_n]}{SD(\bar{X}_n)} \sim N(0,1)$$

Remember that we call the standard deviation of a sampling distribution the standard error, written SE, so

$$\frac{\bar{X}_n - \mu}{SE(\bar{X}_n)} \sim N(0,1)$$

What happens if I rearrange?

$$P\left(-2 \leq \frac{\bar{X}_n - \mu}{SE(\bar{X}_n)} \leq 2\right) \quad = \quad 0.95$$

What happens if I rearrange?

$$P\left(-2 \leq \frac{\bar{X}_n - \mu}{SE(\bar{X}_n)} \leq 2\right) \quad = \quad 0.95$$

$$P\left(-2 \cdot SE \leq \bar{X}_n - \mu \leq 2 \cdot SE\right) \quad = \quad 0.95$$

What happens if I rearrange?

$$
\begin{aligned}
P\left(-2 \leq \frac{\bar{X}_n - \mu}{SE(\bar{X}_n)} \leq 2\right) &= 0.95 \\
P\left(-2 \cdot SE \leq \bar{X}_n - \mu \leq 2 \cdot SE\right) &= 0.95 \\
P\left(-2 \cdot SE - \bar{X}_n \leq -\mu \leq 2 \cdot SE - \bar{X}_n\right) &= 0.95
\end{aligned}
$$

What happens if I rearrange?

$$P\left(-2 \leq \frac{\bar{X}_n - \mu}{SE(\bar{X}_n)} \leq 2\right) = 0.95$$

$$P\left(-2 \cdot SE \leq \bar{X}_n - \mu \leq 2 \cdot SE\right) = 0.95$$

$$P\left(-2 \cdot SE - \bar{X}_n \leq -\mu \leq 2 \cdot SE - \bar{X}_n\right) = 0.95$$

$$P\left(\bar{X}_n - 2 \cdot SE \leq \mu \leq \bar{X}_n + 2 \cdot SE\right) = 0.95$$

Confidence Intervals

Confidence Interval (CI)

Range (A, B) constructed from the sample data with specified probability of containing a population parameter:

$$P(A \leq \theta_0 \leq B) = 1 - \alpha$$

Confidence Intervals

Confidence Interval (CI)

Range (A, B) constructed from the sample data with specified probability of containing a population parameter:

$$P(A \leq \theta_0 \leq B) = 1 - \alpha$$

Confidence Level

The specified probability, typically denoted $1 - \alpha$, is called the confidence level. For example, if $\alpha = 0.05$ then the confidence level is 0.95 or 95%.

Confidence Interval for Mean of Normal Population

Population Variance Known

The interval $\boxed{\bar{X}_n \pm 2\sigma/\sqrt{n}}$ has approximately 95% probability of containing the population mean μ, provided that:

$$\boxed{X_1, X_2, \ldots, X_n \sim \text{iid } N(\mu, \sigma^2)}$$

But how are we supposed to interpret this?

Confidence Interval is a Random Variable!

1. $X_1, \ldots, X_n$ are RVs $\Rightarrow \bar{X}_n$ is a RV (repeated sampling)

Confidence Interval is a Random Variable!

1. $X_1, \ldots, X_n$ are RVs $\Rightarrow \bar{X}_n$ is a RV (repeated sampling)
2. μ, σ and n are constants

Confidence Interval is a Random Variable!

1. $X_1, \ldots, X_n$ are RVs $\Rightarrow \bar{X}_n$ is a RV (repeated sampling)
2. μ, σ and n are constants
3. Confidence Interval $\bar{X}_n \pm 2\sigma/\sqrt{n}$ is also a RV!

Meaning of Confidence Interval

Formal Meaning

If we sampled many times we'd get many different sample means, each leading to a different confidence interval. Approximately 95% of these intervals will contain μ.

Rough Intuition

What values of μ are consistent with the data?

CI for Population Mean: Repeated Sampling

$$\boxed{X_1, X_2, \ldots, X_n \sim \text{iid } N(\mu, \sigma^2)}$$

CI for Population Mean: Repeated Sampling

$$X_1, X_2, \ldots, X_n \sim \text{iid } N(\mu, \sigma^2)$$

Sample 1

CI for Population Mean: Repeated Sampling

$$\begin{tikzcd}
\boxed{X_1, X_2, \ldots, X_n \sim \text{iid } N(\mu, \sigma^2)} \arrow[d] \\
\boxed{\text{Sample 1}} \arrow[d] \\
\boxed{\bar{x}_1}
\end{tikzcd}$$

CI for Population Mean: Repeated Sampling

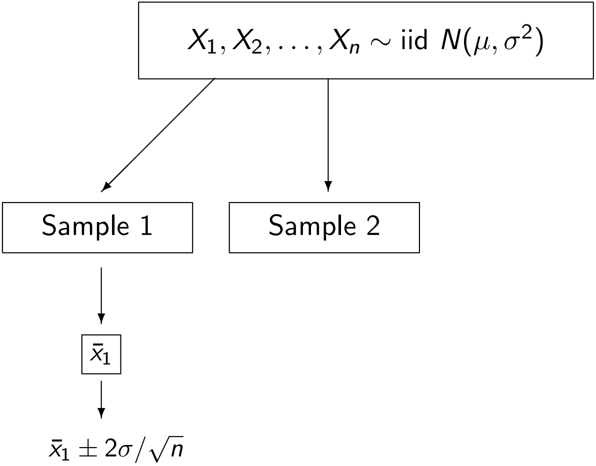
$$\boxed{X_1, X_2, \ldots, X_n \sim \text{iid } N(\mu, \sigma^2)}$$

Sample 1

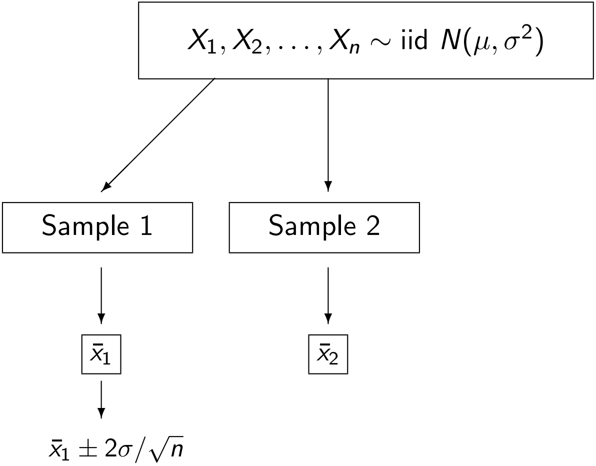
$$\boxed{\bar{x}_1}$$

$$\bar{x}_1 \pm 2\sigma/\sqrt{n}$$

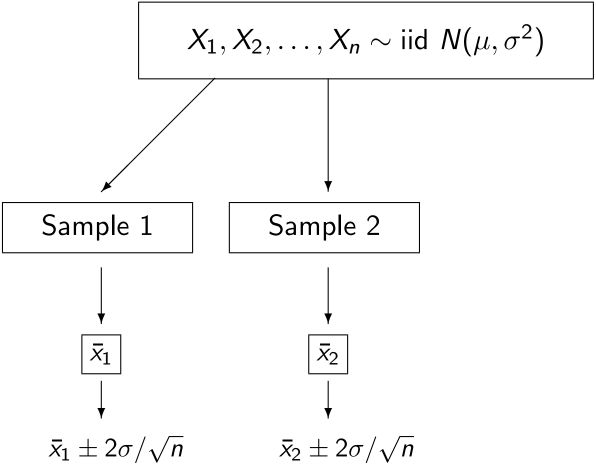
CI for Population Mean: Repeated Sampling



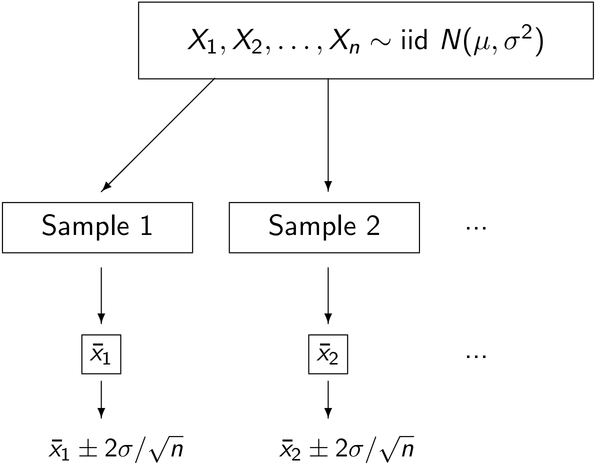
CI for Population Mean: Repeated Sampling



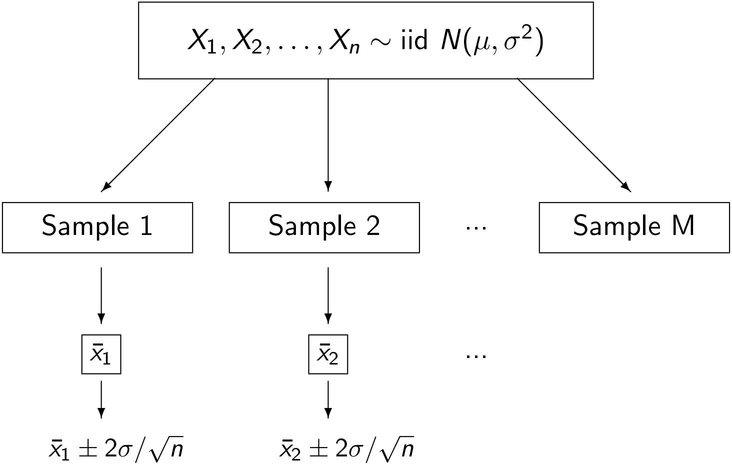
CI for Population Mean: Repeated Sampling



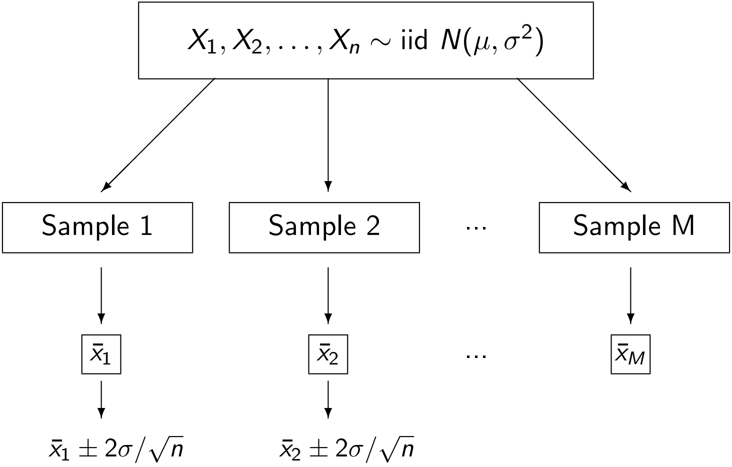
CI for Population Mean: Repeated Sampling



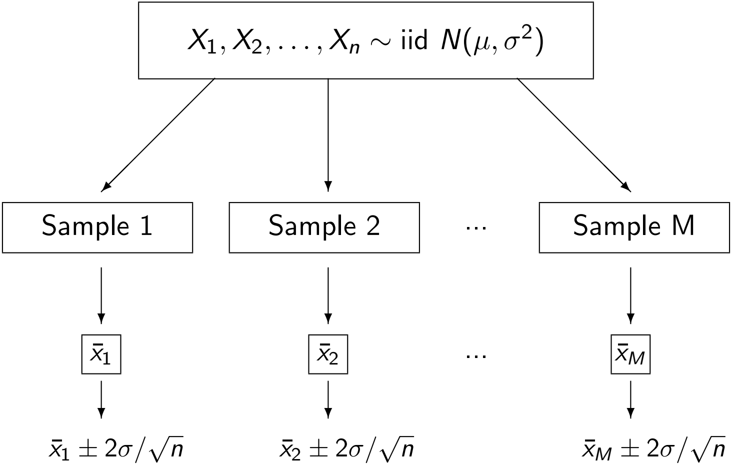
CI for Population Mean: Repeated Sampling



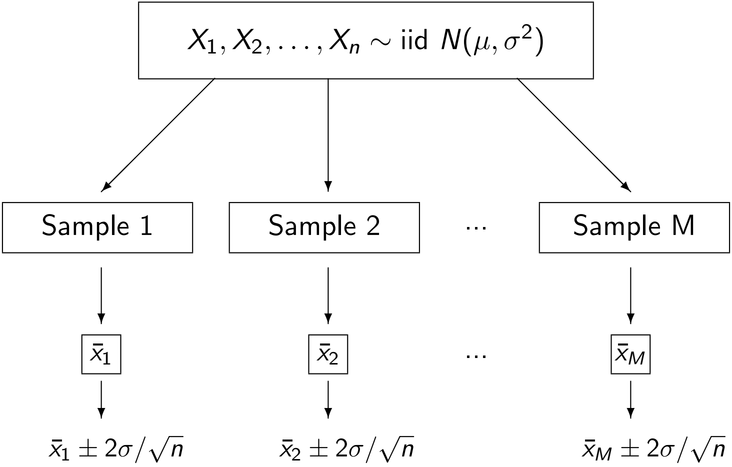
CI for Population Mean: Repeated Sampling



CI for Population Mean: Repeated Sampling

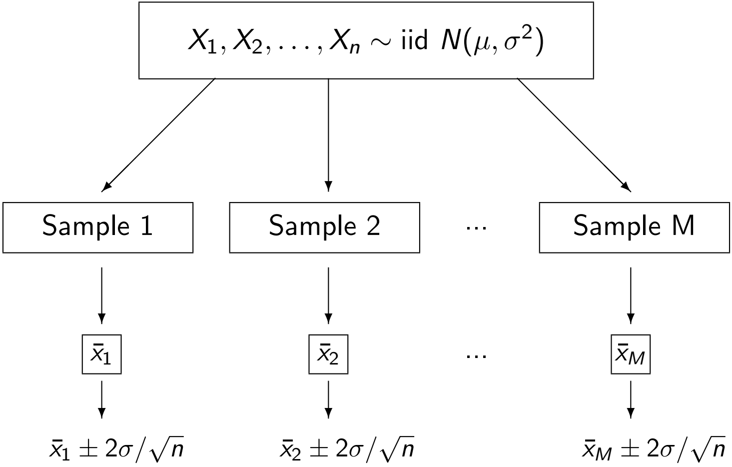


CI for Population Mean: Repeated Sampling



Repeat M times $\rightarrow$ get M different intervals

CI for Population Mean: Repeated Sampling



Repeat M times $\rightarrow$ get M different intervals

Large M $\Rightarrow$ Approx. 95% of these Intervals Contain μ

Simulation Example: $X_1, \ldots, X_5 \sim$ iid $N(0,1)$, $M = 20$

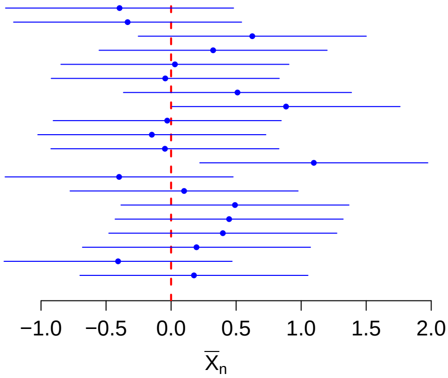


Figure: Twenty confidence intervals of the form $\bar{X}_n \pm 2\sigma/\sqrt{n}$ where $n = 5$, $\sigma^2 = 1$ and the true population mean is 0.

Meaning of Confidence Interval for θ_0

$$\boxed{P(A \leq \theta_0 \leq B) = 1 - \alpha}$$

Each time we sample we'll get a different confidence interval, corresponding to different realizations of the random variables A and B. If we sample many times, approximately $100 \times (1 - \alpha)\%$ of these intervals will contain the population parameter θ_0.

Confidence Intervals: Some Terminology

Margin of Error

When a CI takes the form $\widehat{\theta} \pm ME$, ME is the Margin of Error.

Confidence Intervals: Some Terminology

Margin of Error

When a CI takes the form $\widehat{\theta} \pm ME$, ME is the Margin of Error.

Lower and Upper Confidence Limits

The lower endpoint of a CI is the lower confidence limit (LCL), while the upper endpoint is the upper confidence limit (UCL).

Confidence Intervals: Some Terminology

Margin of Error

When a CI takes the form $\widehat{\theta} \pm ME$, ME is the Margin of Error.

Lower and Upper Confidence Limits

The lower endpoint of a CI is the lower confidence limit (LCL), while the upper endpoint is the upper confidence limit (UCL).

Width of a Confidence Interval

The distance $|\text{UCL} - \text{LCL}|$ is called the width of a CI. This means exactly what it says.

What is the Margin of Error

In the preceding example of a 95% confidence interval for the mean of a normal population when the population variance is known, which of these is the **margin of error**?

(a) $\sigma/\sqrt{n}$

(b) $\bar{X}_n$

(c) σ

(d) $2\sigma/\sqrt{n}$

(e) $1/\sqrt{n}$

What is the Margin of Error

In the preceding example of a 95% confidence interval for the mean of a normal population when the population variance is known, which of these is the **margin of error**?

(a) $\sigma/\sqrt{n}$

(b) $\bar{X}_n$

(c) σ

(d) $2\sigma/\sqrt{n}$

(e) $1/\sqrt{n}$

$2\sigma/\sqrt{n}$, since the CI is $\bar{X}_n \pm 2\sigma/\sqrt{n}$

What is the Width?

In the preceding example of a 95% confidence interval for the mean of a normal population when the population variance is known, which of these is the **width** of the interval?

(a) $\sigma/\sqrt{n}$

(b) $2\sigma/\sqrt{n}$

(c) $3\sigma/\sqrt{n}$

(d) $4\sigma/\sqrt{n}$

(e) $5\sigma/\sqrt{n}$

What is the Width?

In the preceding example of a 95% confidence interval for the mean of a normal population when the population variance is known, which of these is the **width** of the interval?

(a) $\sigma/\sqrt{n}$

(b) $2\sigma/\sqrt{n}$

(c) $3\sigma/\sqrt{n}$

(d) $4\sigma/\sqrt{n}$

(e) $5\sigma/\sqrt{n}$

$4\sigma/\sqrt{n}$, since the CI is $\bar{X}_n \pm 2\sigma/\sqrt{n}$

Example: Calculate the Margin of Error

$$\boxed{\begin{array}{l} X_1, \ldots, X_{100} \sim \text{iid } N(\mu, 1) \text{ but we don't know } \mu. \\ \text{Want to create a 95\% confidence interval for } \mu. \end{array}}$$

What is the margin of error?

Example: Calculate the Margin of Error

$$\boxed{\begin{array}{l} X_1, \ldots, X_{100} \sim \text{iid } N(\mu, 1) \text{ but we don't know } \mu. \\ \text{Want to create a 95\% confidence interval for } \mu. \end{array}}$$

What is the margin of error?

The confidence interval is $\bar{X}_n \pm 2\sigma/\sqrt{n}$ so

$$ME = 2\sigma/\sqrt{n} = 2 \cdot 1/\sqrt{100} = 2/10 = 0.2$$

Example: Calculate the Lower Confidence Limit

$X_1, \ldots, X_{100} \sim N(\mu, 1)$ but we don't know μ.
Want to create a 95% confidence interval for μ.

We found that $ME = 0.2$. The sample mean $\bar{x} = 4.9$. What is the lower confidence limit?

Example: Calculate the Lower Confidence Limit

$$\boxed{\begin{array}{l} X_1, \ldots, X_{100} \ \sim \ N(\mu, 1) \text{ but we don't know } \mu. \\ \text{Want to create a 95\% confidence interval for } \mu. \end{array}}$$

We found that $ME = 0.2$. The sample mean $\bar{x} = 4.9$. What is the lower confidence limit?

$$\text{LCL} = \bar{x} - ME = 4.9 - 0.2 = 4.7$$

Example: Similarly for the Upper Confidence Limit...

$$\boxed{\begin{array}{l} X_1, \ldots, X_{100} \ \sim \ N(\mu, 1) \text{ but we don't know } \mu. \\ \text{Want to create a 95\% confidence interval for } \mu. \end{array}}$$

We found that $ME = 0.2$. The sample mean $\bar{x} = 4.9$. What is the upper confidence limit?

Example: Similarly for the Upper Confidence Limit. . .

$$\boxed{\begin{array}{l} X_1, \ldots, X_{100} \sim N(\mu, 1) \text{ but we don't know } \mu. \\ \text{Want to create a 95\% confidence interval for } \mu. \end{array}}$$

We found that $ME = 0.2$. The sample mean $\bar{x} = 4.9$. What is the upper confidence limit?

$$\text{UCL} = \bar{x} + ME = 4.9 + 0.2 = 5.1$$

Example: 95% CI for Normal Mean, Popn. Var. Known

$$\boxed{X_1, \ldots, X_{100} \sim N(\mu, 1) \text{ but we don't know } \mu.}$$

95% CI for $\mu = [4.7, 5.1]$

What values of μ are plausible?

Example: 95% CI for Normal Mean, Popn. Var. Known

$$\boxed{X_1, \ldots, X_{100} \sim N(\mu, 1) \text{ but we don't know } \mu.}$$

95% CI for $\mu = [4.7, 5.1]$

What values of μ are plausible?

The data actually came from a $N(5, 1)$ Distribution.

Want to be more certain? Use higher confidence level.

What value of c should we use to get a $100\times(1-\alpha)\%$ CI for μ?

$$P\left(-c \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq c\right) \;=\; 1-\alpha$$

Want to be more certain? Use higher confidence level.

What value of c should we use to get a $100\times(1-\alpha)\%$ CI for μ?

$$P\left(-c \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq c\right) \;\; = \;\; 1-\alpha$$

$$P\left(\bar{X}_n - c\sigma/\sqrt{n} \leq \mu \leq \bar{X}_n + c\sigma/\sqrt{n}\right) \;\; = \;\; 1-\alpha$$

Want to be more certain? Use higher confidence level.

What value of c should we use to get a $100 \times (1-\alpha)\%$ CI for μ?

$$P\left(-c \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq c\right) = 1-\alpha$$

$$P\left(\bar{X}_n - c\sigma/\sqrt{n} \leq \mu \leq \bar{X}_n + c\sigma/\sqrt{n}\right) = 1-\alpha$$

Take $c = \texttt{qnorm}(1-\alpha/2)$

Want to be more certain? Use higher confidence level.

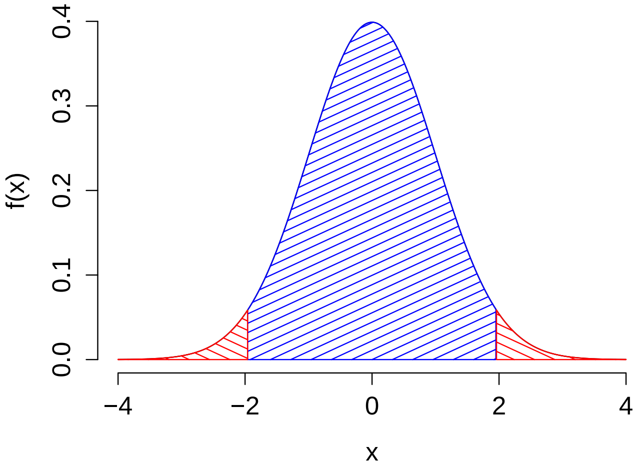
What value of c should we use to get a $100\times(1-\alpha)\%$ CI for μ?

$$P\left(-c \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq c\right) \;=\; 1-\alpha$$

$$P\left(\bar{X}_n - c\sigma/\sqrt{n} \leq \mu \leq \bar{X}_n + c\sigma/\sqrt{n}\right) \;=\; 1-\alpha$$

Take $c = \texttt{qnorm}(1-\alpha/2)$

$$\bar{X}_n \pm \texttt{qnorm}(1-\alpha/2) \times \sigma/\sqrt{n}$$



0.0
0.1
0.2
0.3
0.4
f(x)
−4
−2
0
2
4
x

What Affects the Margin of Error?

$$\boxed{\bar{X}_n \pm \texttt{qnorm}(1-\alpha/2) \times \sigma/\sqrt{n}}$$

Sample Size n

ME decreases with n: bigger sample $\implies$ tighter interval

Population Std. Dev. σ

ME increases with σ: more variable population $\implies$ wider interval

Confidence Level $1-\alpha$

ME increases with $1-\alpha$: higher conf. level $\implies$ wider interval

What Affects the Margin of Error?

$$\boxed{\bar{X}_n \pm \texttt{qnorm}(1-\alpha/2) \times \sigma/\sqrt{n}}$$

Sample Size n

ME decreases with n: bigger sample $\implies$ tighter interval

Population Std. Dev. σ

ME increases with σ: more variable population $\implies$ wider interval

Confidence Level $1-\alpha$

ME increases with $1-\alpha$: higher conf. level $\implies$ wider interval

Conf. Level	90%	95%	99%
α	0.1	0.05	0.01
$\texttt{qnorm}(1-\alpha/2)$	1.64	1.96	2.56

Lecture #16 – Confidence Intervals II

Comparing intervals with different confidence levels

What if the population is normal but σ is unknown?

What if the population isn't normal? – The Central Limit Theorem

CI for a Proportion Using the Central Limit Theorem

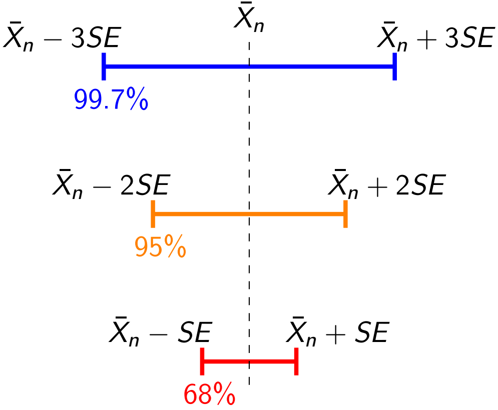


Figure: Each CI gives a range of "plausible" values for the population mean μ, centered at the sample mean $\bar{X}_n$. Values near the middle are "more plausible" in the sense that a small reduction in confidence level gives a much shorter interval centered in the same place. This is because the sample mean is unlikely to take on values far from the population mean in repeated sampling.

Assume that: $X_1, \ldots, X_n \sim \text{iid } \mathrm{N}(\mu, \sigma^2)$

σ Known

$$P\left[-\texttt{qnorm}(1-\alpha/2) \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq \texttt{qnorm}(1-\alpha/2)\right] = 1-\alpha$$

$\Longrightarrow$ Confidence Interval: $\bar{X}_n \pm \texttt{qnorm}(1-\alpha/2) \times \sigma/\sqrt{n}$

Assume that: $X_1, \ldots, X_n \sim$ iid $\mathrm{N}(\mu, \sigma^2)$

σ Known

$$P\left[-\texttt{qnorm}(1-\alpha/2) \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq \texttt{qnorm}(1-\alpha/2)\right] = 1 - \alpha$$

$\implies$ Confidence Interval: $\bar{X}_n \pm \texttt{qnorm}(1-\alpha/2) \times \sigma/\sqrt{n}$

σ Unknown

Idea: estimate σ with S. Unfortunately:

$$\frac{\bar{X}_n - \mu}{S/\sqrt{n}} \quad \text{IS NOT A NORMAL RV!}$$

50000 Simulation replications: $X_1, \ldots, X_5 \sim$ iid $\mathrm{N}(\mu, \sigma^2)$

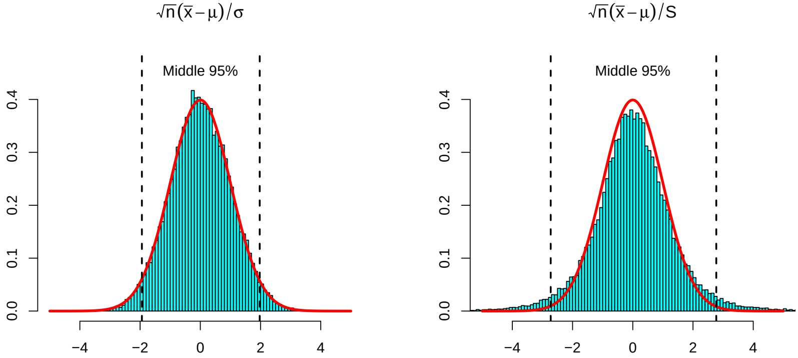


Figure: In each plot the red curve is the pdf of the standard normal RV.
At left: the sampling distribution of $\sqrt{5}(\bar{X}_5 - \mu)/\sigma$ is standard normal.
At right: the sampling distribution of $\sqrt{5}(\bar{X}_5 - \mu)/S$ clearly isn't!

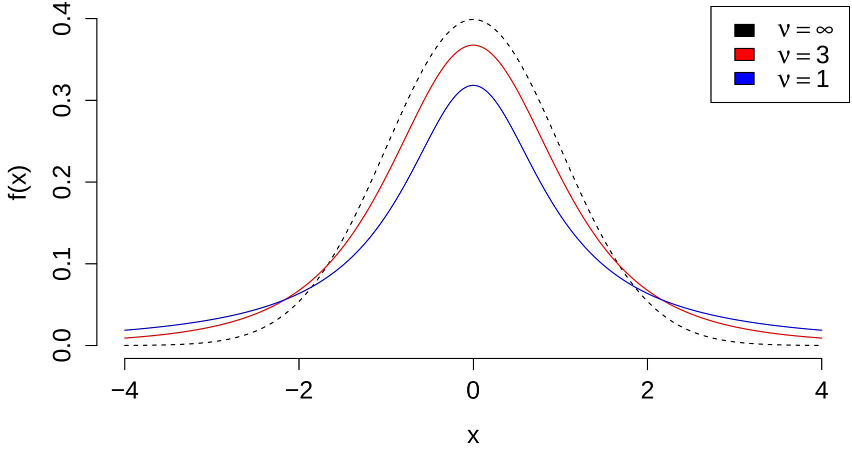
Student-t Random Variable

If $X_1, \ldots, X_n \sim$ iid $N(\mu, \sigma^2)$, then

$$\boxed{\frac{\bar{X}_n - \mu}{S/\sqrt{n}} \sim t(n-1)}$$

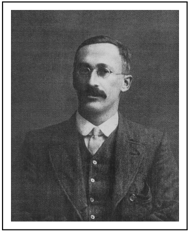
- Parameter: $\nu = n - 1$ "degrees of freedom"
- Support $= (-\infty, \infty)$
- Symmetric around zero, but mean and variance may not exist!
- Degrees of freedom ν control "thickness of tails"
- As $\nu \to \infty$, $t \to$ Standard Normal.

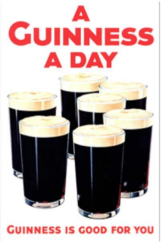
Student-t PDFs



Who was "Student?"

"Guinnessometrics: The Economic Foundation of Student's t"





"Student" is the pseudonym used in 19 of 21 published articles by William Sealy Gosset, who was a chemist, brewer, inventor, and self-trained statistician, agronomer, and designer of experiments ... [Gosset] worked his entire adult life ... as an experimental brewer for one employer: Arthur Guinness, Son & Company, Ltd., Dublin, St. Jamess Gate. Gosset was a master brewer and rose in fact to the top of the top of the brewing industry: Head Brewer of Guinness.

CI for Mean of Normal Distribution, Popn. Var. Unknown

Same argument as we used when the variance was known, except with $t(n-1)$ rather than standard normal distribution:

$$P\left(-c \leq \frac{\bar{X}_n - \mu}{S/\sqrt{n}} \leq c\right) = 1 - \alpha$$

$$P\left(\bar{X}_n - c\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X}_n + c\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

$c = \texttt{qt}(1-\alpha/2, \texttt{df} = n-1)$

$$\boxed{\bar{X}_n \pm \texttt{qt}(1-\alpha/2, \texttt{df} = n-1)\ \frac{S}{\sqrt{n}}}$$

Comparison of CIs for Mean of Normal Distribution

$100 \times (1-\alpha)\%$ Confidence Level

$$\boxed{X_1, \ldots, X_n \sim \text{iid } N(\mu, \sigma^2)}$$

Known Population Std. Dev. (σ)

$$\bar{X}_n \pm \texttt{qnorm}(1-\alpha/2)\, \frac{\sigma}{\sqrt{n}}$$

Unknown Population Std. Dev. (σ)

$$\bar{X}_n \pm \texttt{qt}(1-\alpha/2, \texttt{df} = n-1)\, \frac{S}{\sqrt{n}}$$

Comparison of Normal and t CIs

Table: Values of $\texttt{qt}(1-\alpha/2, \texttt{df} = n-1)$ for various choices of n and α.

n	1	5	10	30	100	∞
$\alpha = 0.10$	6.31	2.02	1.81	1.70	1.66	1.64
$\alpha = 0.05$	12.71	2.57	2.23	2.04	1.98	1.96
$\alpha = 0.01$	63.66	4.03	3.17	2.75	2.63	2.58

As $n \to \infty$, $t(n-1) \to N(0,1)$

In a sense, using the t-distribution involves making a "small-sample correction." In other words, it is only when n is fairly small that this makes a practical difference for our confidence intervals.

Am I Taller Than The Average American Male?

Source: Centers for Disease Control (pg. 16)

Sample Mean	69 inches
Sample Std. Dev.	6 inches
Sample Size	5647
My Height	73 inches

$$\begin{aligned} \widehat{SE}(\bar{X}_n) &= s/\sqrt{n} \\ &= 6/\sqrt{5647} \\ &\approx 0.08 \end{aligned}$$

Assuming the population is normal,

$$\boxed{\bar{X}_n \pm \texttt{qt}(1-\alpha/2, \texttt{df} = n-1)\ \widehat{SE}(\bar{X}_n)}$$

What is the approximate value of

`qt(1-0.05/2, df = 5646)`?

Am I Taller Than The Average American Male?

Source: Centers for Disease Control (pg. 16)

Sample Mean	69 inches
Sample Std. Dev.	6 inches
Sample Size	5647
My Height	73 inches

$$\begin{aligned} \widehat{SE}(\bar{X}_n) &= s/\sqrt{n} \\ &= 6/\sqrt{5647} \\ &\approx 0.08 \end{aligned}$$

Assuming the population is normal,

$$\boxed{\bar{X}_n \pm \texttt{qt}(1 - \alpha/2, \texttt{df} = n - 1)\ \widehat{SE}(\bar{X}_n)}$$

What is the approximate value of

`qt(1-0.05/2, df = 5646)`?

For large n, $t(n-1) \approx N(0,1)$, so the answer is approximately 2

Am I Taller Than The Average American Male?

Source: Centers for Disease Control (pg. 16)

Sample Mean	69 inches
Sample Std. Dev.	6 inches
Sample Size	5647
My Height	73 inches

$$\begin{aligned}\widehat{SE}(\bar{X}_n) &= s/\sqrt{n} \\ &= 6/\sqrt{5647} \\ &\approx 0.08\end{aligned}$$

Assuming the population is normal,

$$\boxed{\bar{X}_n \pm \ \mathtt{qt}(1-\alpha/2, \mathtt{df} = n-1)\ \widehat{SE}(\bar{X}_n)}$$

What is the approximate value of

`qt(1-0.05/2, df = 5646)`?

For large n, $t(n-1) \approx N(0,1)$, so the answer is approximately 2

What is the ME for the 95% CI?

Am I Taller Than The Average American Male?

Source: Centers for Disease Control (pg. 16)

Sample Mean	69 inches
Sample Std. Dev.	6 inches
Sample Size	5647
My Height	73 inches

$$\begin{aligned}\widehat{SE}(\bar{X}_n) &= s/\sqrt{n} \\ &= 6/\sqrt{5647} \\ &\approx 0.08\end{aligned}$$

Assuming the population is normal,

$$\boxed{\bar{X}_n \pm \mathtt{qt}(1-\alpha/2, \mathtt{df} = n-1)\ \widehat{SE}(\bar{X}_n)}$$

What is the approximate value of `qt(1-0.05/2, df = 5646)`?

For large n, $t(n-1) \approx N(0,1)$, so the answer is approximately 2

What is the ME for the 95% CI?

$ME \approx 0.16 \implies 69 \pm 0.16$

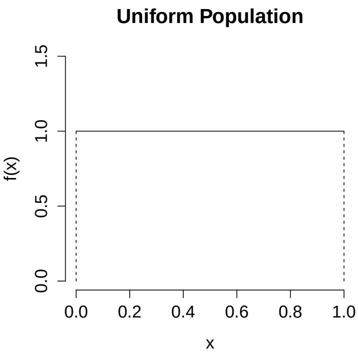
The Central Limit Theorem

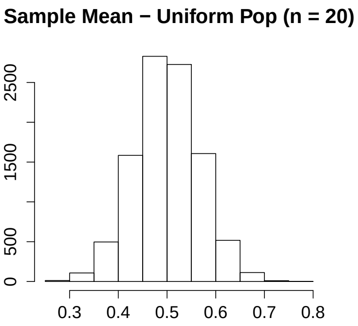
Suppose that $X_1, \ldots, X_n$ are a random sample from a some population that is not necessarily normal and has an unknown mean μ. Then, provided that n is *sufficiently large*,

$$\frac{\bar{X}_n - \mu}{S/\sqrt{n}} \approx N(0,1)$$

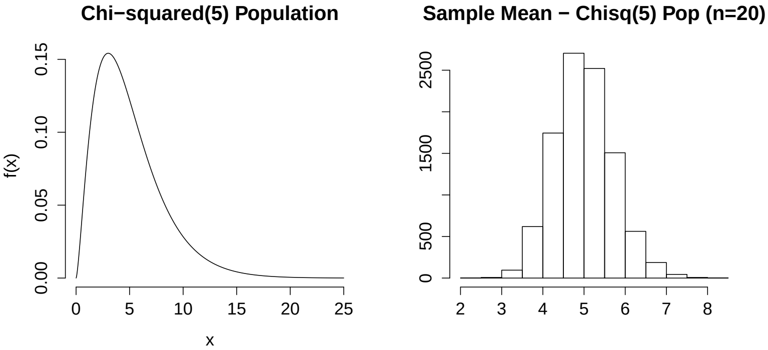
We will use this fact to create *approximate* CIs for population mean even if we know *nothing* about the population.

Example: Uniform(0,1) Population, $n = 20$

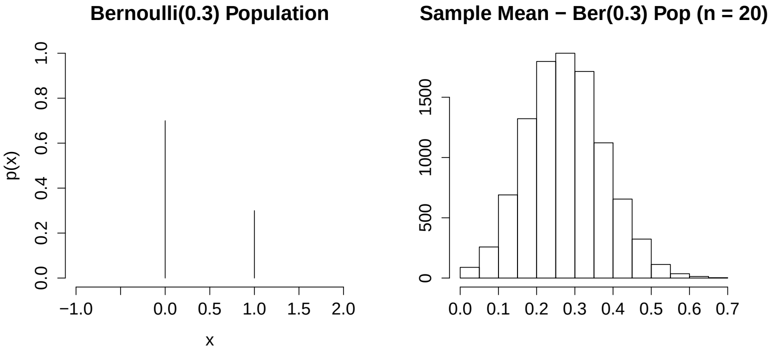




Example: $\chi^2(5)$ Population, $n = 20$



Example: Bernoulli(0.3) Population, $n = 20$



Are US Voters Really That Ignorant?

Pew: "What Voters Know About Campaign 2012"

The Data

Of 771 registered voters polled, only 39% correctly identified John Roberts as the current chief justice of the US Supreme Court.

Research Question

Is the majority of voters unaware that John Roberts is the current chief justice, or is this just sampling variation?

Assume Random Sampling...

Confidence Interval for a Proportion

What is the appropriate probability model for the sample?

$X_1, \ldots, X_n \sim$ iid Bernoulli(p), $1 =$ Know Roberts is Chief Justice

What is the parameter of interest?

$p =$ Proportion of voters *in the population* who know Roberts is Chief Justice.

What is our estimator?

Sample Proportion: $\widehat{p} = (\sum_{i=1}^{n} X_i)/n$

Sample Proportion *is* the Sample Mean!

$X_1, \ldots, X_n \sim$ iid Bernoulli(p)

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$$

Sample Proportion *is* the Sample Mean!

$X_1, \ldots, X_n \sim$ iid Bernoulli(p)

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$$

$$E[\widehat{p}] \;\; = \;\; E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{np}{n} = p$$

Sample Proportion *is* the Sample Mean!

$X_1, \ldots, X_n \sim$ iid Bernoulli(p)

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$$

$$\begin{aligned} E[\widehat{p}] &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{np}{n} = p \\ Var(\widehat{p}) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \end{aligned}$$

Sample Proportion *is* the Sample Mean!

$X_1, \ldots, X_n \sim$ iid Bernoulli(p)

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$$

$$\begin{aligned}
E[\widehat{p}] &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{np}{n} = p \\
Var(\widehat{p}) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \\
SE(\widehat{p}) &= \sqrt{Var(\widehat{p})} = \sqrt{\frac{p(1-p)}{n}}
\end{aligned}$$

Sample Proportion *is* the Sample Mean!

$X_1, \ldots, X_n \sim$ iid Bernoulli(p)

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$$

$$\begin{aligned}
E[\widehat{p}] &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{np}{n} = p \\
Var(\widehat{p}) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \\
SE(\widehat{p}) &= \sqrt{Var(\widehat{p})} = \sqrt{\frac{p(1-p)}{n}} \\
\widehat{SE}(\widehat{p}) &= \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}
\end{aligned}$$

Central Limit Theorem Applied to Sample Proportion

Central Limit Theorem: Intuition

Sample means are approximately normally distributed provided the sample size is large even if the population is non-normal.

CLT For Sample Mean

$$\frac{\bar{X}_n - \mu}{S/\sqrt{n}} \approx N(0,1)$$

CLT for Sample Proportion

$$\frac{\widehat{p} - p}{\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}} \approx N(0,1)$$

In this example, the population is Bernoulli(p) rather than normal.

The sample mean is $\widehat{p}$ and the population mean is p.

Approximate 95% CI for Population Proportion

$$\frac{\widehat{p}-p}{\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}} \approx N(0,1)$$

$$P\left(-2 \leq \frac{\widehat{p}-p}{\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}} \leq 2\right) \approx 0.95$$

$$P\left(\widehat{p}-2\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \leq p \leq \widehat{p}+2\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right) \approx 0.95$$

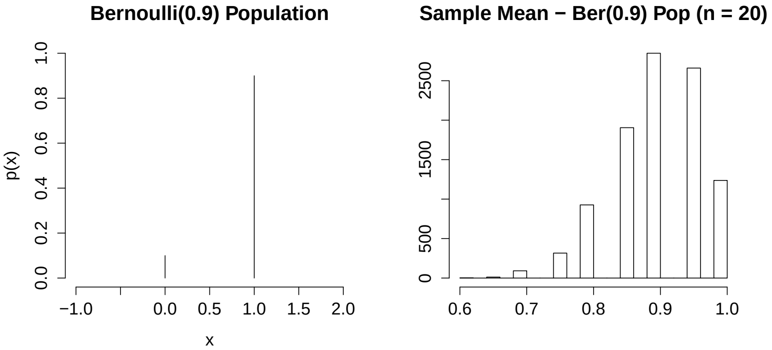
$100 \times (1 - \alpha)$ CI for Population Proportion (p)

$X_1, \ldots, X_n \sim$ iid Bernoulli(p)

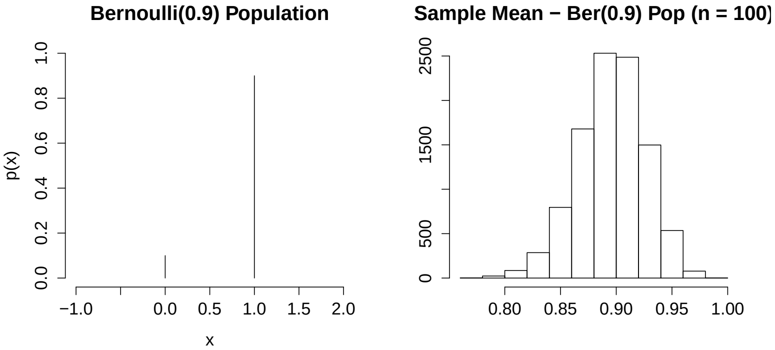
$$\widehat{p} \pm \texttt{qnorm}(1 - \alpha/2)\sqrt{\frac{\widehat{p}(1 - \widehat{p})}{n}}$$

Approximation based on the CLT. Works well provided n is large and p isn't too close to zero or one.

Example: Bernoulli(0.9) Population, $n = 20$



Example: Bernoulli(0.9) Population, $n = 100$



Approximate 95% CI for Population Proportion

39% of 771 Voters Polled Correctly Identified Chief Justice Roberts

$$\begin{aligned} \widehat{SE}(\widehat{p}) &= \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = \sqrt{\frac{(0.39)(0.61)}{771}} \\ &\approx 0.018 \end{aligned}$$

What is the ME for an approximate 95% confidence interval?

Approximate 95% CI for Population Proportion

39% of 771 Voters Polled Correctly Identified Chief Justice Roberts

$$\begin{aligned}\widehat{SE}(\widehat{p}) &= \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = \sqrt{\frac{(0.39)(0.61)}{771}} \\ &\approx 0.018\end{aligned}$$

What is the ME for an approximate 95% confidence interval?

$$ME \approx 2 \times \widehat{SE}(\bar{X}_n) \approx 0.04$$

Approximate 95% CI for Population Proportion

39% of 771 Voters Polled Correctly Identified Chief Justice Roberts

$$
\begin{aligned}
\widehat{SE}(\widehat{p}) &= \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = \sqrt{\frac{(0.39)(0.61)}{771}} \\
&\approx 0.018
\end{aligned}
$$

What is the ME for an approximate 95% confidence interval?

$$ME \approx 2 \times \widehat{SE}(\bar{X}_n) \approx 0.04$$

What can we conclude?

$$\boxed{\text{Approximate 95\% CI: } (0.35, 0.43)}$$