ECOM20001
Econometrics 1

Lecture Note 2
Probability

A/Prof David Byrne
Department of Economics
University of Melbourne

Stock and Watson: Chapter 2

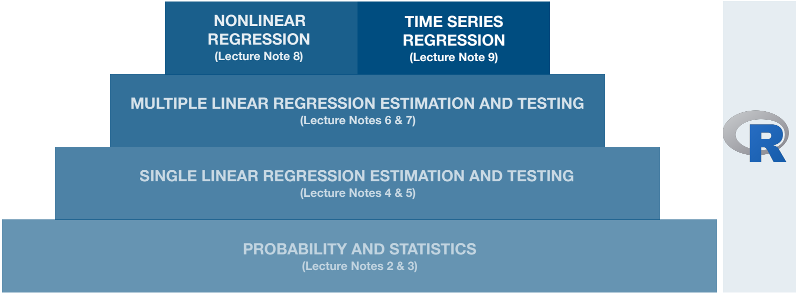
Roadmap

- Motivating Example: pokies
- One Random Variable
  - Random processes and probabilities
  - Discrete and continuous random variables
  - Probability density functions
  - Cumulative density functions
  - Describing distributions: mean, variance, skewness, kurtosis
- Two Random Variables
  - Joint, marginal, and conditional distributions
  - Covariance, correlation, and independence
- Distributions: Normal, Chi-Squared, Student t, F Distributions
- Random Sampling
  - Sample averages and population means
  - Law of Large Numbers, Central Limit Theorem

Introduction

- The next two sets of lecture notes review concepts in probability (this note) and statistics (next note)
- We will review material that is necessary for developing econometric models and estimating causal relationships
  - stats is needed for figuring out how the world works using data
- Some material we cover may be review, yet is good to refresh probability and statistics
- Work with an application to gambling and household income

Building our Econometric Toolkit



Australia Has the World's Biggest Gambling Problem

Source: The Economist, 2 February 2017

A Gambling Dataset for Victoria

- `lga`: Local Government Area (LGA)
- `avg_income`: Average Annual Individual Income per Year in an LGA($)
- `Negms_per_1000`: Number of Electronic Gaming Machines (Pokies) per 1000 People in an LGA
- Data is for the year 2010
- Fields of economics: economics of poverty, addiction and crime, illicit markets

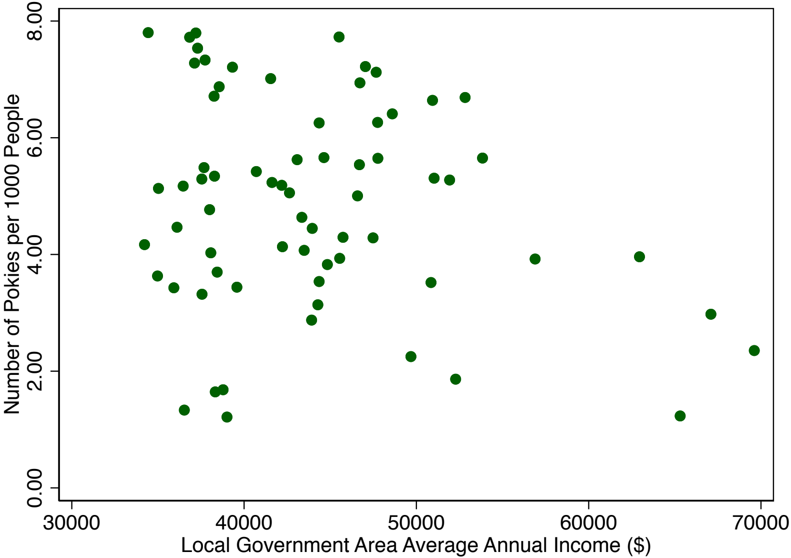
Local Government Areas in Victoria

Local Government Areas Around Melbourne

Data Set (data for first 20 LGAs shown)

lga	avg_income	Negms_1000
Alpine	36465	5.17
Ararat	36841	7.72
Ballarat	41543	7.01
Banyule	51030	5.31
Bass Coast	37736	7.33
Baw Baw	42226	4.13
Bayside	69612	2.35
Benalla	38556	6.87
Boroondara	65310	1.23
Brimbank	42636	5.06
Campaspe	38069	4.03
Cardinia	44281	3.14
Casey	44357	3.54
Central Goldfields	34434	7.80
Colac-Otway	37546	5.29
Corangamite	35924	3.43
Darebin	46722	6.94
East Gippsland	37201	7.80
Frankston	43353	4.64
Gannawarra	34222	4.17

Pokies Per 1000 People vs. LGA Average Income



Data Generating Processes

- What generates data: random processes
- Random processes involve outcomes, sample spaces, and probabilities
  - Outcome: something that happened
  - Sample Space: all the possible things that can happen
  - Probabilities: the chance that something from the sample space happens

Data Generating Processes

- Random data generating process: flipping a coin
  - example outcome: heads
  - sample space: {heads, tails}
  - probabilities: P(heads)=50% and P(tails)=50%
- Random data generating process: rolling a 6-sided dice
  - example outcome: 4
  - sample space: {1,2,3,4,5,6}
  - probabilities: P(1)=16.67%, P(2)=16.67%, P(3)=16.67%, P(4)=16.67%, P(5)=16.67%, P(6)=16.67%

Probabilities

- We can write probabilities in terms of percentages or decimals
  - 50% and 0.5 is the same
  - 21.8 and 0.218 is the same
  - 100 and 1.0 is the same
- Probabilities always sum to 1

Random Variables

- Discrete random variables: take only a discrete set of values
  - 0,1,2,3,4,5,...
  - flipping a coin, rolling a dice
- Continuous random variables: take on a continuum of values
  - all (infinitely many) values on the interval [0,5] like 0.0001, 3.242432, 4.8823
  - household income, stock market prices, pokies in an LGA

Discrete Random Variables

- Probability distribution: list of all possible outcomes the variable can take and the probability of observing each value
- Cumulative probability distribution: probability that a random variable is less than or equal to some value

	Discrete Outcome: Number of Pokies Venues in an LGA									
	1	2	3	4	5	6	7	8	9	10+
Probability Distribution	10.3	7.4	10.3	13.2	4.4	4.4	5.9	7.3	1.5	35.3
Cumulative Probability Distribution	10.3	17.7	28.0	41.2	45.6	50.0	55.9	63.2	64.7	100.0

Discrete Random Variables

	Discrete Outcome: Number of Pokies Venues in an LGA									
	1	2	3	4	5	6	7	8	9	10+
Probability Distribution	$\boxed{10.3}$	7.4	10.3	13.2	4.4	4.4	5.9	7.3	1.5	35.3
Cumulative Probability Distribution	$\boxed{10.3}$	17.7	28.0	41.2	45.6	50.0	55.9	63.2	64.7	100.0

Add up

Discrete Random Variables

	Discrete Outcome: Number of Pokies Venues in an LGA									
	1	2	3	4	5	6	7	8	9	10+
Probability Distribution	10.3	7.4	10.3	13.2	4.4	4.4	5.9	7.3	1.5	35.3
Cumulative Probability Distribution	10.3	17.7	28.0	41.2	45.6	50.0	55.9	63.2	64.7	100.0

Add up

Discrete Random Variables

	Discrete Outcome: Number of Pokies Venues in an LGA									
	1	2	3	4	5	6	7	8	9	10+
Probability Distribution	10.3	7.4	10.3	13.2	4.4	4.4	5.9	7.3	1.5	35.3
Cumulative Probability Distribution	10.3	17.7	$\boxed{28.0}$	41.2	45.6	50.0	55.9	63.2	64.7	100.0

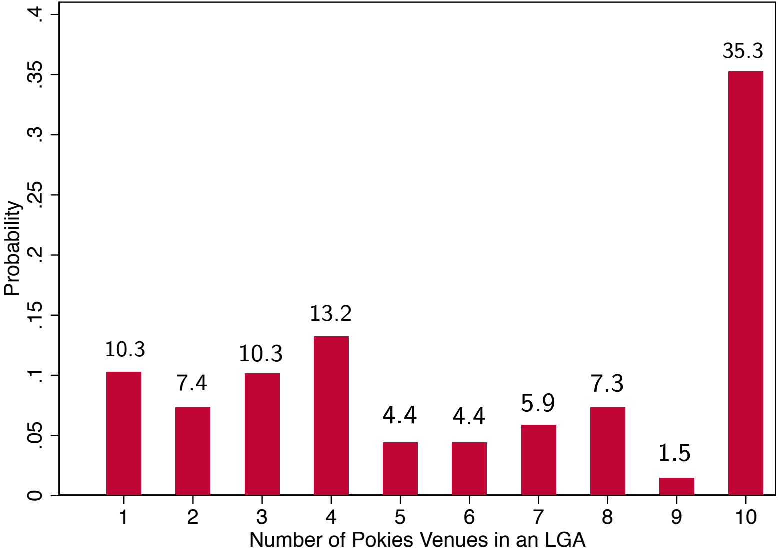
Add up

Discrete Random Variables

	Discrete Outcome: Number of Pokies Venues in an LGA									
	1	2	3	4	5	6	7	8	9	10+
Probability Distribution	10.3	7.4	10.3	13.2	4.4	4.4	5.9	7.3	1.5	35.3
Cumulative Probability Distribution	10.3	17.7	28.0	41.2	45.6	50.0	55.9	63.2	64.7	100.0

Add up

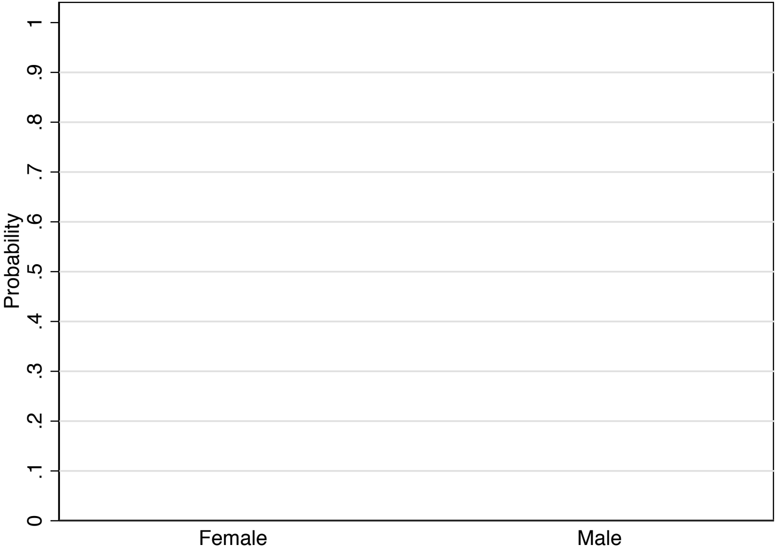
Discrete Random Probability Distribution Graphically



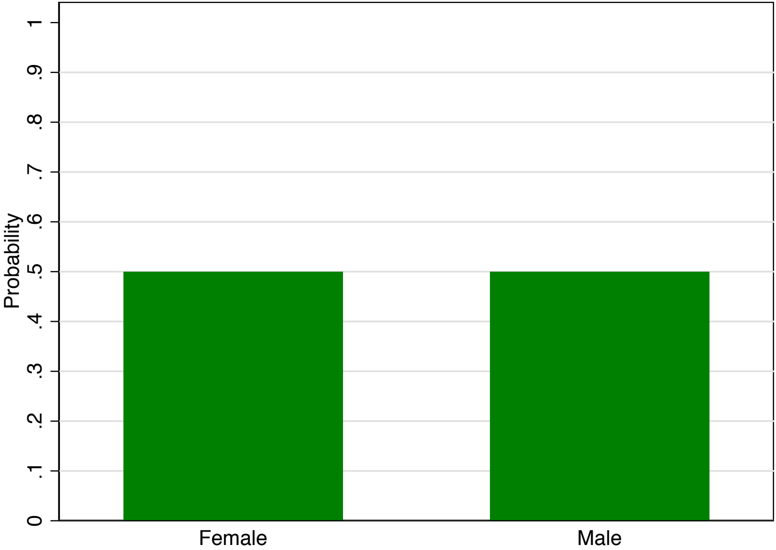
Bernoulli Random Variables

- Bernoulli random variables are a special case of a discrete random variable when the possible outcomes are either 0 or 1
- Examples
  - Female (1) or Male (0)
  - Born in Australia (1) or Born Overseas (0)
  - Love Econometrics (1), Don't Love Econometrics (0)
- Also often called binary random variables

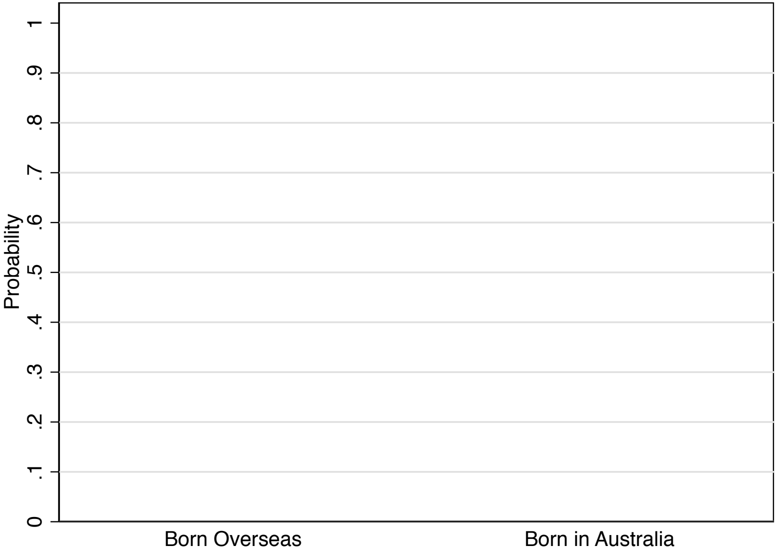
Bernoulli/Binary Random Variables Probability Distribution Graphically



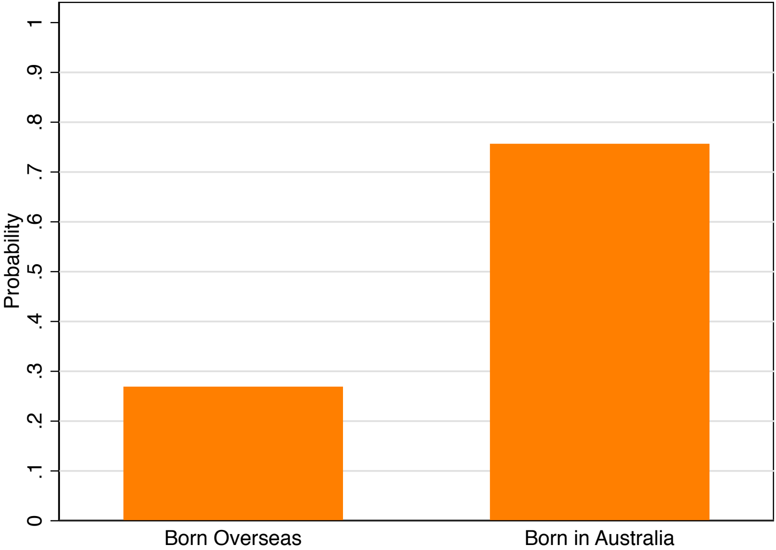
Bernoulli/Binary Random Variables Probability Distribution Graphically



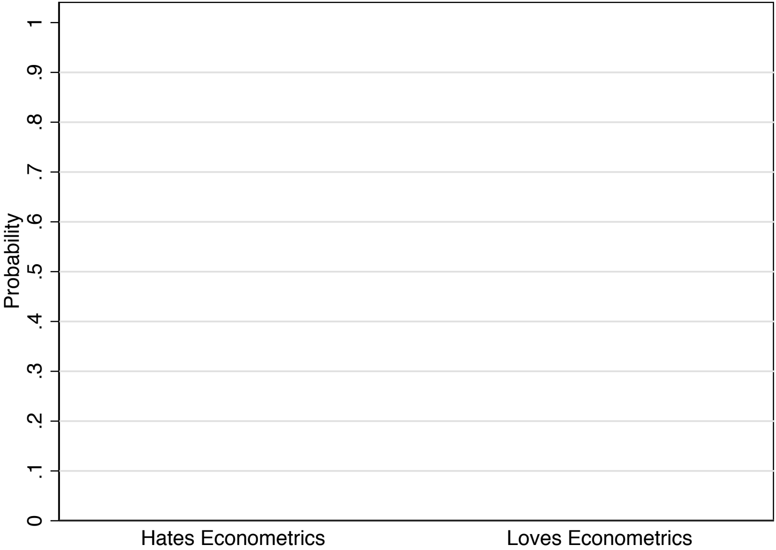
Bernoulli/Binary Random Variables Probability Distribution Graphically



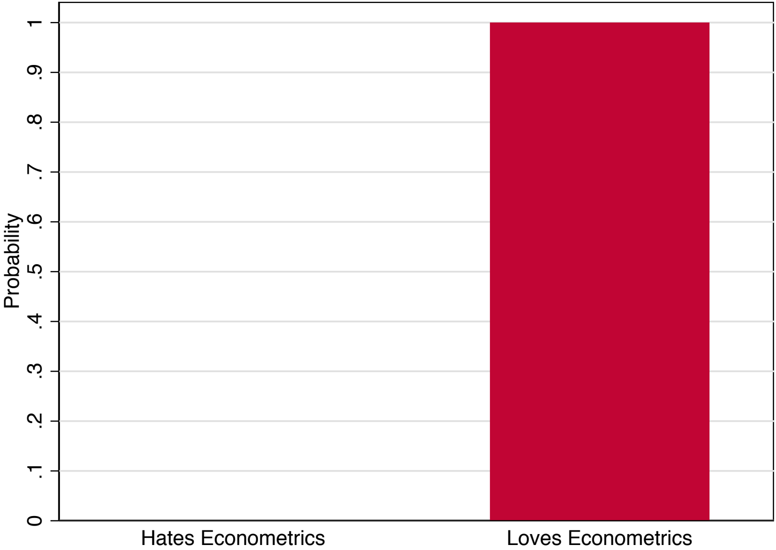
Bernoulli/Binary Random Variables Probability Distribution Graphically



Bernoulli/Binary Random Variables Probability Distribution Graphically



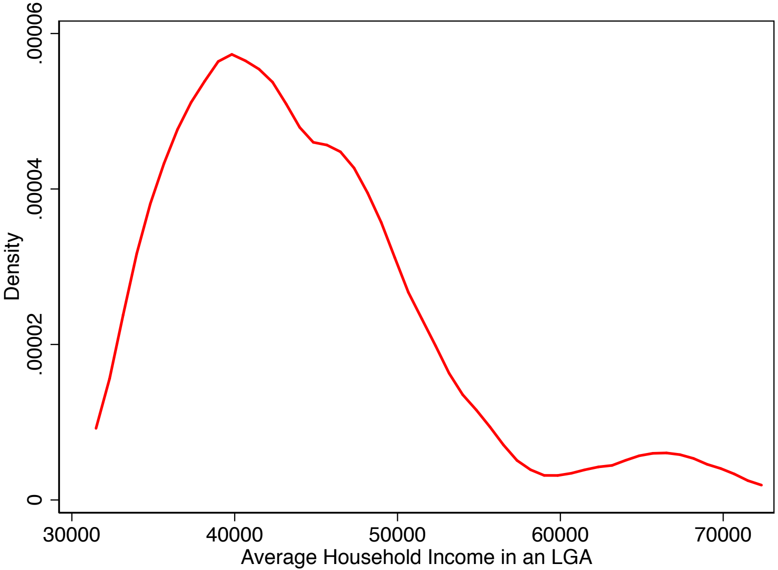
Bernoulli/Binary Random Variables Probability Distribution Graphically



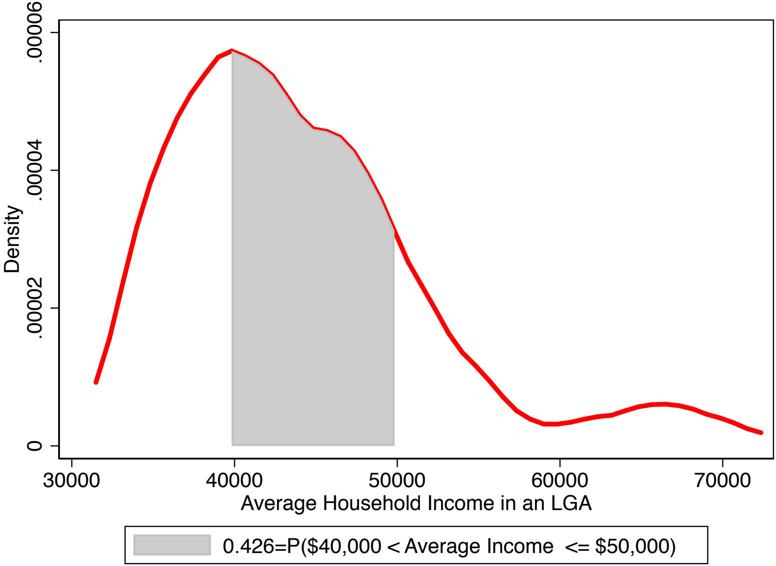
Continuous Random Variables

- Continous random variables can take on a continuum (infinite!) of possible values
- Because of this, we cannot list an individual probability for each possible value
- Instead we use a probability density function (PDF)
  - also often just called the "density function" or "density"
- A PDF defines the probability that a random variable X lies between two values a and b: $P(a \leq X \leq b)$
  - Example: probability average income (X) is between \$40,000 (a) and \$50,000 (b)
  - So, PDF defines $P(\$40,000 \leq \text{Average Income} \leq \$50,000)$

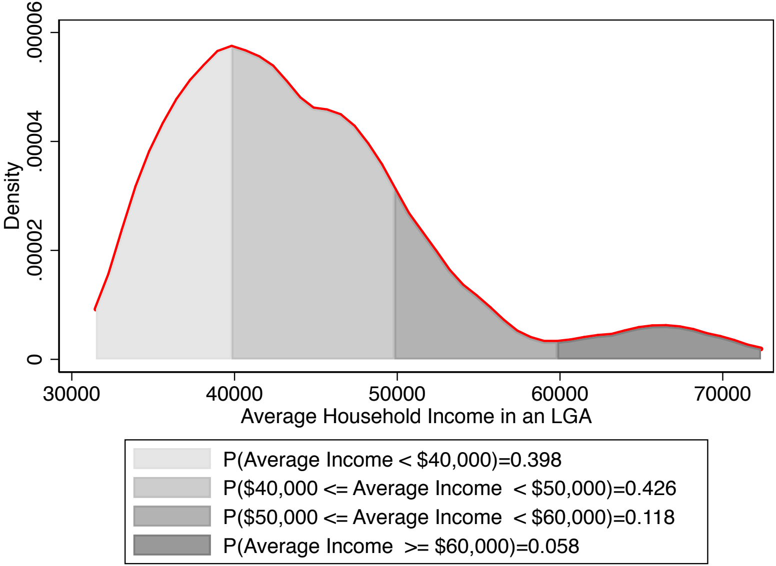
Continuous Probability Density Graphically



Continuous Probability Density Graphically



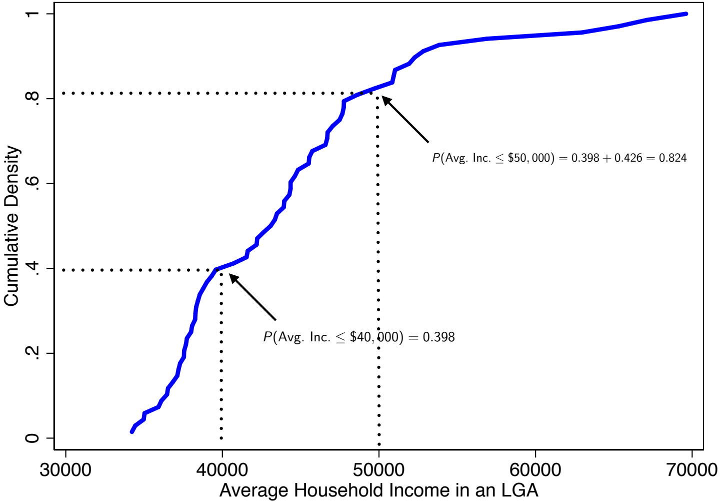
Continuous Probability Density Graphically



Cumulative Distribution Function

- The cumulative probability density is the probability that a random variable X is <u>less than or equal to</u> a particular value x
- Same definition for discrete and continuous random variables
- Cumulative probability density function (CDF) describes the cumulative density with continuous random variables
  - Ex: probability average income (X) is less than \$50,000 (x)
  - So, CDF defines $P(\text{Average Income} \leq \$50,000)$
- From the last slide, we can compute 3 points from the CDF
  - $P(\text{Avg. Inc.} \leq \$40,000) = 0.398$
  - $P(\text{Avg. Inc.} \leq \$50,000) = 0.398 + 0.426 = 0.824$
  - $P(\text{Avg. Inc.} \leq \$60,000) = 0.398 + 0.426 + 0.118 = 0.942$
- We can describe the CDF graphically by plotting it for all dollar values

Cumulative Density Function Graphically



Summation Operators

- Suppose that you have some data with n observations: $x_1, x_2, \ldots, x_n$
  - Example: x_6 could be the Avg Household Income in LGA #6
- Add up all the data points: $x_1 + x_2 +, \ldots, x_n = \sum_{i=1}^{n} x_i$
- The summation operator is $\sum_{i=1}^{n}$ which has an index i that runs from $i = 1, 2, 3, \ldots, n-1, n$
- Convenient notation to express sums of many elements
- The average of x across n data points: add all the data points (x_i) up and divide by the number of data points (n)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Rules With Summation Operators

- For a constant a:

$$\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i$$

and

$$\sum_{i=1}^{n} a = n \times a$$

- For two random variables x_i and y_i:

$$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

Expected Value or Mean of a Random Variable

- The expected value $E(X)$ or mean μ_X of a random variable describes the variable's central tendency
- For a discrete random variable X taking on $i = 1, \ldots K$ possible values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$, we compute the mean as:

$$E(X) = \mu_X = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n = \sum_{i=1}^{n} p_i x_i$$

- Example: The expected value of a 6-sided dice roll is:

$$E(X) = 0.167 \times 1 + 0.167 \times 2 + 0.167 \times 3 + 0.167 \times 4 + 0.167 \times 5 + 0.167 \times 6 = 3.5$$

Note: we will mainly focus on continuous random variables from this point forward

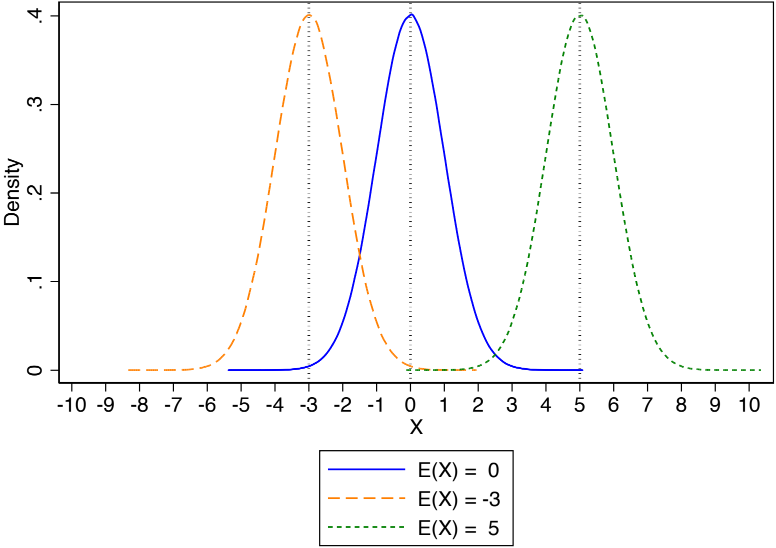
Expected Value or Mean of a Random Variable

- For a continuous random variable X which takes on infinitely many possible realisations x over the interval between x_{min} and x_{max}, we need to use an integral to compute its expected value:
$$E(X) = \mu_X = \int_{x_{min}}^{x_{max}} xf(x)dx$$
where $f(x)$ is the probability density function for X

- Examples: mean income levels, mean stock market returns

Note: we will not be working mathematically with integrals in the subject

Means Graphically



Important Rules with Expectations

- For constants a and b, and random variable X:

$$E(aX + b) = aE(X) + b$$

- If $X_1, X_2, \ldots, X_n$ are random variables and $a_1, a_2, \ldots, a_n$ are constants:

$$E(a_1X_1 + a_2X_2 + \ldots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n)$$

Variance of a Random Variable

- The variance and standard deviation of a random variable X measures the dispersion (i.e., how spread out) X's probability distribution is
- Variance is denoted $var(X)$ or σ_X^2, and for continuous random variables is calculated as:

$$var(X) = \sigma_X^2 = E[(X - \mu_X)^2]$$

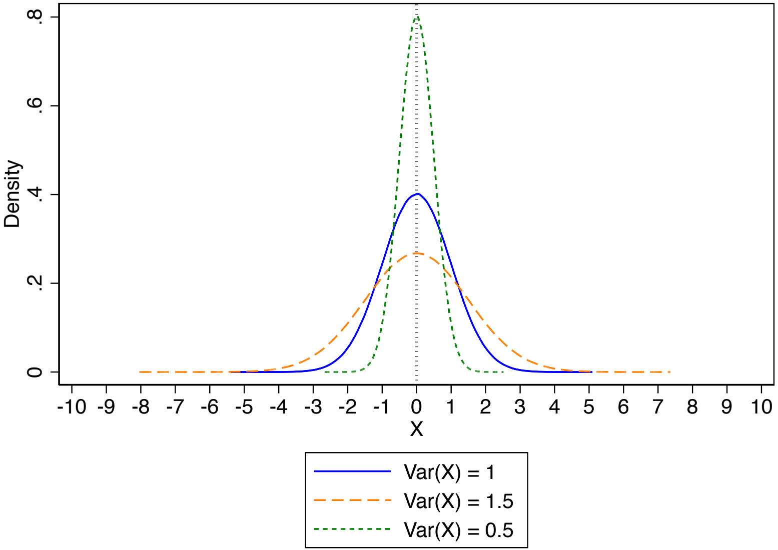
- Standard deviation is denoted $std(X)$ or σ_X and is the square root of the variance:

$$std(X) = \sigma_X = \sqrt{\sigma_X^2}$$

- $std(X)$ has the same unit as random variable X; $var(X)$ doesn't

Note: we will not study variances of discrete random variables or the Law of Iterated Expectations, see Ch.2 of the text on p.74-75 if interested. We will introduce conditional expectations/means later in the subject.

Variances Graphically



Mean and Variance of a Linear Function of a Random Variable

- We have two random variables Y (Pokies Per 1000 People) and X (Average Income in \$1000's of Dollars)
- Suppose Y is a linear function of X:

$$Y = 90 - 1.5X$$

- Interpretation: for each additional \$1,000 LGA Average Income ($\uparrow$ X by 1) there are 1.5 fewer pokies per 1000 people ($\downarrow Y$ by 1.5)
- Using our rules for expected values, $E(Y)$ is:

$$E(Y) = E(90 - 1.5X) = E(90) - 1.5E(X)$$

or

$$\mu_Y = 90 - 1.5\mu_X$$

Mean and Variance of a Linear Function of a Random Variable

- Variance of Y:

$$\begin{aligned}\sigma_Y^2 &= E[(Y-\mu_Y)^2)] \\ &= E[(\underbrace{(90-1.5X)}_{Y} - \underbrace{(90-1.5\mu_X)}_{\mu_Y})^2] \\ &= E[(-1.5(X-\mu_X))^2] \\ &= 2.25E[(X-\mu_X)^2] \\ &= 2.25\sigma_X^2\end{aligned}$$

- In sum, the variances are related as follows

$$\sigma_Y^2 = 2.25\sigma_X^2 = 1.5^2\sigma_X^2$$

- Taking the square root of both sides, the standard deviations are related as follows:

$$\sigma_Y = 1.5\sigma_X$$

Mean and Variance of a Linear Function of a Random Variable

- In general, if random variables X and Y are linearly related according to constants a (intercept) and b (slope):

$$Y = a + bX$$

then the means are related as follows:

$$\mu_Y = a + b\mu_X$$

and the variances are related as follows:

$$\sigma_Y^2 = b^2\sigma_X^2$$

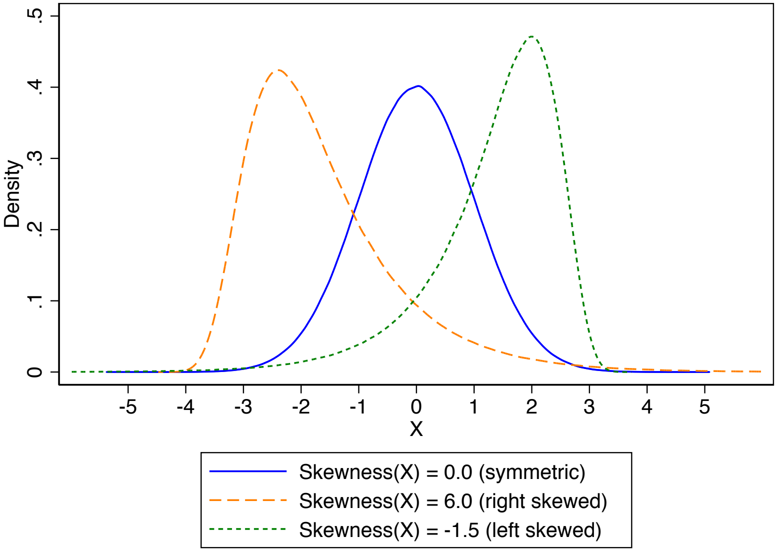
Skewness

- The skewness of the probability distribution of a random variable X measures the symmetry (or lack thereof) in a distribution
- It is computed as follows:

$$\text{Skewness} = \frac{E[(X - \mu_X)^3]}{\sigma_X^3}$$

- The larger the skewness, the “fatter” the tail of one side of the distribution relative to the centre of the distribution
- Applications of skewness: extreme income inequality (how rich are the top 1%?), stock market tail risk (how big is a very unlikely financial collapse?), climate change (how disasterous is it if all the polar ice caps melt?)

Skewness Graphically



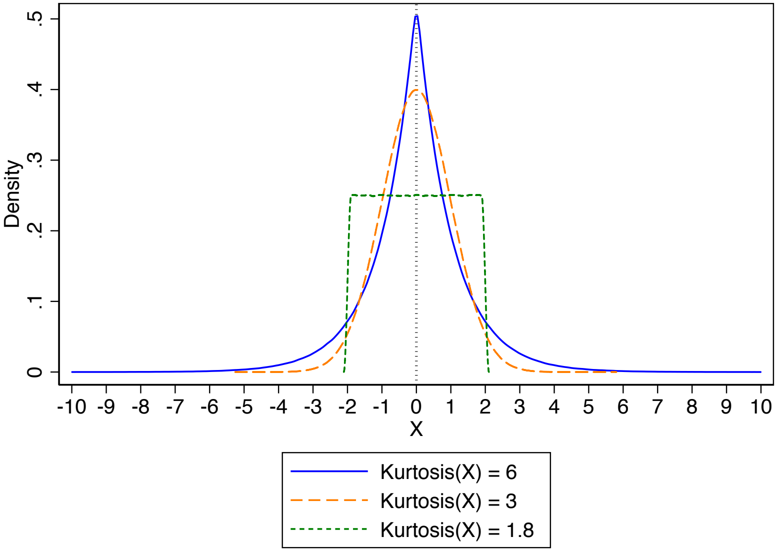
Kurtosis

- Kurtosis measures how much dispersion of a distribution is in its tails
- Tells us something about how much variability of a random variable is driven by more extreme values relative to the mean (or less extreme values)
- It is computed as:

$$\text{Kurtosis} = \frac{E[(X - \mu_X)^4]}{\sigma_X^4}$$

- The larger the kurtosis, the more likely are extreme deviations from the mean
- Applications are similar to skewness (extreme income inequality, stock market tail risk), except kurtosis captures both extreme good and bad events

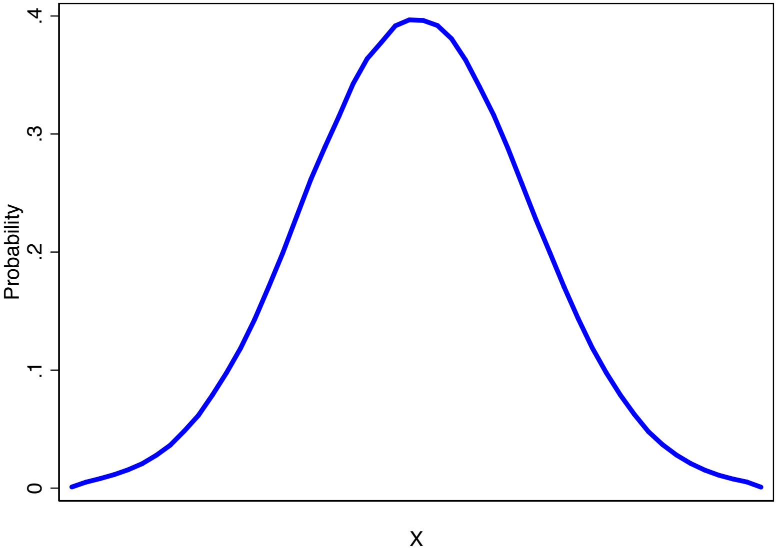
Kurtosis Graphically



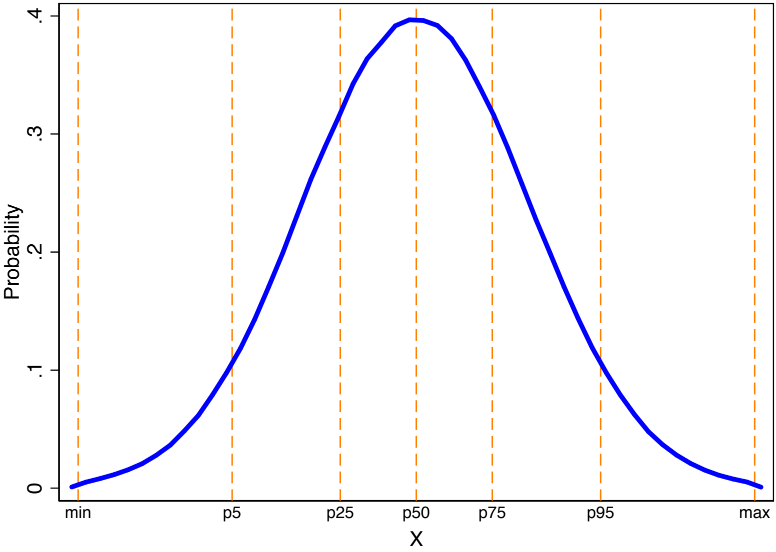
Descriptive Statistics

- We use descriptive statistics for getting a sense of what data look like
- Standard descriptive statistics include: mean, standard deviation, minimum value (min) and maximum value (max)
- Often also helpful to present different percentiles of the data of interest
  - A value p is the p^{th} percentile of a distribution if p percent of the observations are equal to or below p
- Percentiles typically presented: 5^{th}, 50^{th} (or the median), 95^{th}
  - often labelled as p5, p50, p95
- Inter-quartile Range (IQR) consists off the 25^{th}, 50^{th}, 75^{th} percentiles (p25, p50, p75)
  - IQR is less often presented, but gives a sense of the dispersion in the middle part of the distribution

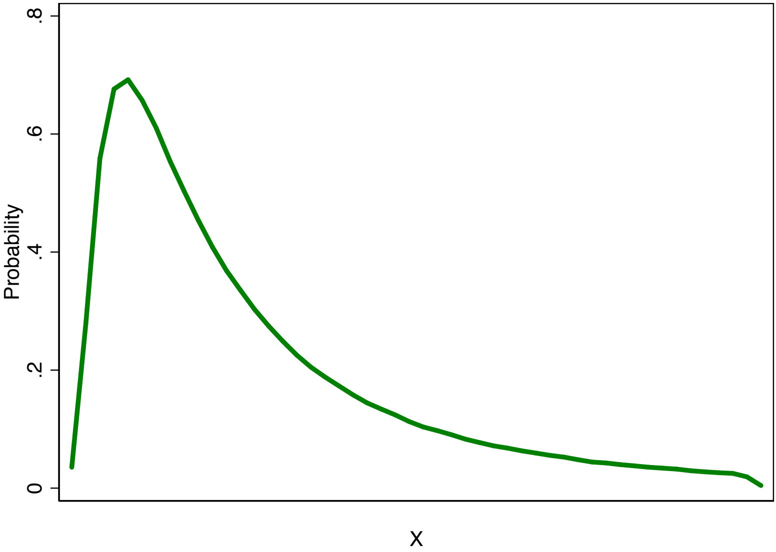
Percentiles Example: Symmetric Distribution



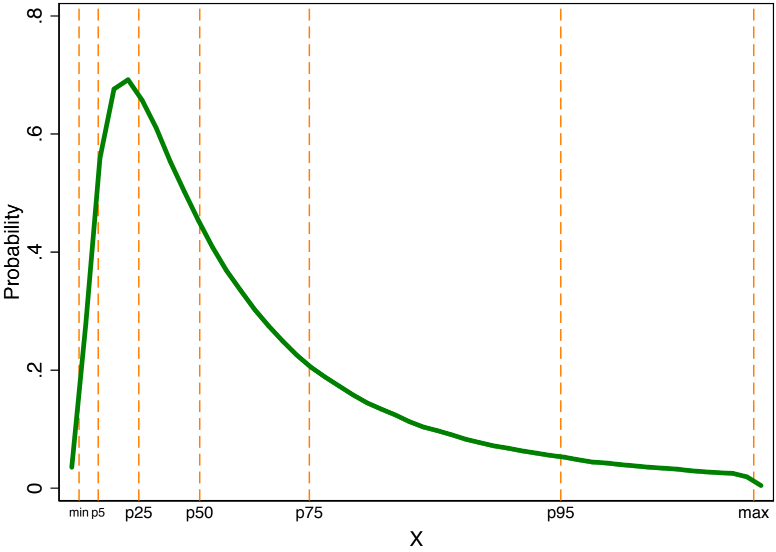
Percentiles Example: Symmetric Distribution



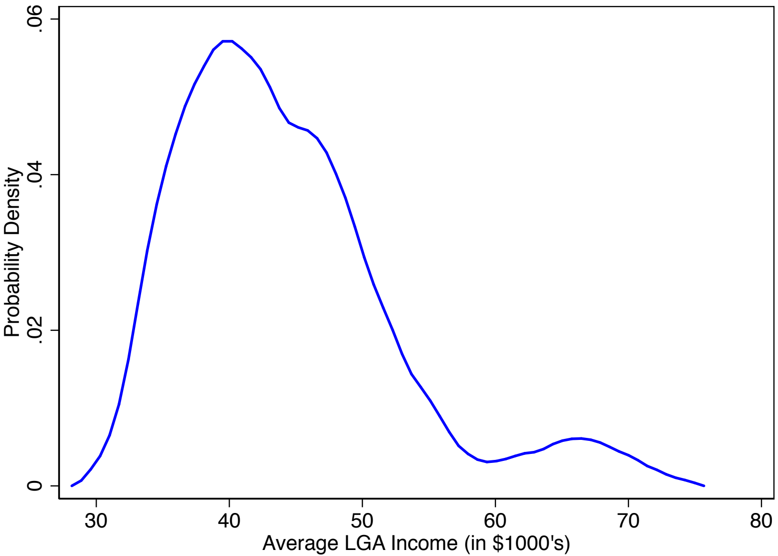
Percentiles Example: Skewed Distribution



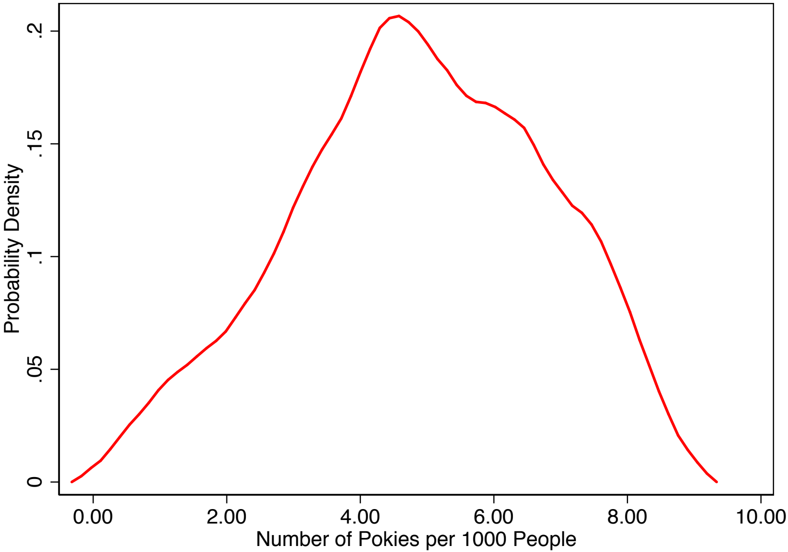
Percentiles Example: Skewed Distribution



Distribution of LGA Average Income



Distribution of Pokies per 1000 People



Descriptive Statistics Example

	Mean	Std Dev.	Min	Max	p5	p50	p95
Average Income in LGA (in $1000s)	44.1	7.9	34.2	69.6	35.0	43.2	62.9
Number of Pokies per 1000 People	4.9	1.8	1.2	7.8	1.6	5.1	7.7

- Always be consistent with the number of digits used
- Work with a scale of data that allows for 1 to 3 digits after the decimal

Two Random Variables

- In many fields of business and social science, we are very interested in empirical analyses of two random variables, Y and X
  - How do wages compare across men and women? Y=wages, X=gender
  - Do university graduates earn more income than non-graduates? Y=income, X=education
  - Does congestion fall when toll prices are increased? Y=congestion, X=toll prices
- To start investiging these many other important questions with data, we need to understand the concepts of joint, marginal, and conditional probability distributions

Joint Distributions

- The joint probability distribution of two random variables X and Y is the probability that the random variables simultaneously take on certain values, say $X = x$
  - x is a realized value of X
  - y is a realized value of Y
- The joint probability distribution can be written as

$$P(X = x, Y = y)$$

- These probabilities for all (x, y) combinations sum to 1

Joint Distributions Example

- Consider an application to weather and commuting times
  - Rain is X: $X = 0$ means rain; $X = 1$ means no rain
  - Commute is Y: $Y = 0$ means long commute; $Y = 1$ means short commute
- Table that characterizes the joint distribution of raining and commuting times with 4 possible outcomes:

	Rain (X=0)	No Rain (X=1)	Total
Long Commute (Y=0)	0.15	0.07	0.22
Short Commute (Y=1)	0.15	0.63	0.78
Total	0.30	0.70	1.00

- $P(X = 1, Y = 1) = 0.63$ implies that on 63% of days there is No Rain ($X = 1$) and a Short Commute ($Y = 1$)

Marginal Probability Distribution

- The marginal probability distribution of a random variable Y when there are two random variables is just another name for the probability distribution of Y
- Has a different name to distinguish the distribution of Y alone from the joint distribution of Y and X together
- The marginal distribution of Y, $P(Y = y)$, is computed using the joint distribution for X and Y.
- To compute $P(Y = y)$ add up the probabilities where $Y = y$ across all the K different possible values X can take on:

$$P(Y = y) = \sum_{i=1}^{K} P(X = x_i, Y = y)$$

Marginal Probability Distribution Example

	Rain (X=0)	No Rain (X=1)	Total
Long Commute (Y=0)	0.15	0.07	0.22
Short Commute (Y=1)	0.15	0.63	0.78
Total	0.30	0.70	1.00

- The marginal distribution of Y (commuting time) is:
  - Long Commute: P(Y=0)=0.15+0.07=0.22
  - Short Commute: P(Y=1)=0.15+0.63=0.78
- The marginal distribution of X (rain) is:
  - Rain: P(X=0)=0.15+0.15=0.30
  - No Rain: P(X=1)=0.07+0.63=0.70
- Notice how the marginal probabilities also add up to 1

Conditional Distributions

- The conditional distribution of Y is the distribution of Y conditional on X taking on specific value x
- We denote the conditional distribution by

$$P(Y = y|X = x)$$

  where the "|" part means "conditional" or "given"
- We compute the conditional distribution using both the joint distribution of X and Y and the marginal distribution of X:

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

- Let's go back to our example to see what this involves

Conditional Probability Distribution Example

	Rain (X=0)	No Rain (X=1)	Total
Long Commute (Y=0)	0.15	0.07	0.22
Short Commute (Y=1)	0.15	0.63	0.78
Total	0.30	0.70	1.00

- What is the probability distribution of a long commute Y conditional on raining? That is, what is $P(Y|X=0)$?
  - $P(Y=0|X=0) = \frac{P(X=0,Y=0)}{P(X=0)} = \frac{0.15}{0.30} = 0.5$
  - $P(Y=1|X=0) = \frac{P(X=0,Y=1)}{P(X=0)} = \frac{0.15}{0.30} = 0.5$
- In words: when it is raining, there is a 50/50 chance of having a long commute or short commute

Conditional Probability Distribution Example

	Rain (X=0)	No Rain (X=1)	Total
Long Commute (Y=0)	0.15	0.07	0.22
Short Commute (Y=1)	0.15	0.63	0.78
Total	0.30	0.70	1.00

- What is the probability distribution of a rain X conditional on there being a short commute? That is, what is $P(X|Y=1)$?
  - $P(X=0|Y=1) = \frac{P(X=0,Y=1)}{P(Y=1)} = \frac{0.15}{0.78} = 0.19\%$
  - $P(X=1|Y=1) = \frac{P(X=1,Y=1)}{P(Y=1)} = \frac{0.63}{0.78} = 0.81\%$
- In words: if you had a short commute, there was an 81% chance it was not raining

Independence

- If X and Y are independent random variables there are 3 important results of note

1. Their joint distribution is the product of their individual marginal probability distributions:

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

2. The conditional distribution of Y given X is just the marginal probability distribution of Y (and vice versa):

$$P(Y = y|X = x) = P(Y = y)$$

and

$$P(X = x|Y = y) = P(X = x)$$

Independence (continued)

3. The covariance of X and Y is 0:

$$cov(X, Y) = 0$$

which further implies the correlation $corr(X, Y) = 0$

Independence Example

- Let's consider an example with two binary random variables
- My son was born on a sunny day
  - $X = 1$ "Dave's kid is a boy" and $X = 0$ "Dave's kid is a girl"
  - $Y = 1$ "Sunny day" and $Y = 0$ "Cloudy day"
- Two questions to consider:
  - Would you expect the probability of my kid being a boy to depend on whether it was a sunny day on the kid's due date for birth?
  - Would you expect the weather on my kid's due date for birth to depend on my kid's gender?
- The answer to both questions is (I hope) no!

Independence Example

- The probability of having a boy does not depend on the whether on the due date
- If it is sunny ($Y = 1$) on the due date...
  - $P(X = 1|Y = 1) = 0.5$ (boy)
  - $P(X = 0|Y = 1) = 0.5$ (girl)
- If it is cloudy ($Y = 0$) on the due date...
  - $P(X = 1|Y = 0) = 0.5$ (boy)
  - $P(X = 0|Y = 0) = 0.5$ (girl)
- Notice how the conditional and marginal densities are the same:
  - $P(X = 1|Y = 1) = P(X = 1|Y = 0) = P(X = 1) = 0.5$ (boy)
  - $P(X = 0|Y = 1) = P(X = 0|Y = 0) = P(X = 0) = 0.5$ (girl)
- The distribution of X (gender) does not depend on Y (weather)

Independence Example

- Let's flip it around: the probability of having a sunny day on the due date does not depend on the gender of my kid
- If I have a boy ($X = 1$) the probability of it being sunny on the due date is ...
  - $P(Y = 1|X = 1) = 0.7$ (sunny)
  - $P(Y = 0|X = 1) = 0.3$ (cloudy)
- If I have a girl ($X = 0$) the probability of it being sunny on the due date is ...
  - $P(Y = 1|X = 0) = 0.7$ (sunny)
  - $P(Y = 0|X = 0) = 0.3$ (cloudy)
- Again, the conditional and marginal densities are the same:
  - $P(Y = 1|X = 1) = P(Y = 1|X = 0) = P(Y = 1) = 0.7$ (sunny)
  - $P(Y = 0|X = 1) = P(Y = 0|X = 0) = P(Y = 0) = 0.3$ (cloudy)
- The distribution of Y (weather) does not depend on X (gender)

Independence Example

- Joint probabilities of (gender, weather) combinations are computed as products of the respective marginal densities for gender and weather
  - Probability of having a boy AND it being sunny:
    - $P(X = 1, Y = 1) = P(X = 1)P(Y = 1) = 0.5 \times 0.7 = 0.35$
  - Probability of having a girl AND it being sunny:
    - $P(X = 0, Y = 1) = P(X = 0)P(Y = 1) = 0.5 \times 0.7 = 0.35$
  - Probability of having a boy AND it being cloudy:
    - $P(X = 1, Y = 0) = P(X = 1)P(Y = 1) = 0.5 \times 0.3 = 0.15$
  - Probability of having a girl AND it being cloudy:
    - $P(X = 0, Y = 0) = P(X = 0)P(Y = 1) = 0.5 \times 0.3 = 0.15$
- Finally, notice how the probabilities sum to 1:

$$0.35 + 0.35 + 0.15 + 0.15 = 1$$

Covariance

- Covariance measures the extent to which two random variables X and Y move together
- It is denoted $cov(X, Y) = \sigma_{XY}$ and is computed as

$$cov(X, Y) = \sigma_{XY} = E[(X - \mu_x)(Y - \mu_Y)]$$

- $cov(X, Y)$ is positive when X and Y move together, and is negative when X and Y tend to move in opposite directions.
- Covariance can take on any value.

Correlation

- Correlation also measures the extent to which X and Y together, but it is always between -1 and 1
- It is denoted $corr(X, Y)$ and is computed as:

$$corr(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- X and Y are uncorrelated if $corr(X, Y) = 0$
- $corr(X, Y) = -1$ means perfect negative correlation
- $corr(X, Y) = 1$ means perfect positive correlation;

Variance of the Sum of Two Random Variables

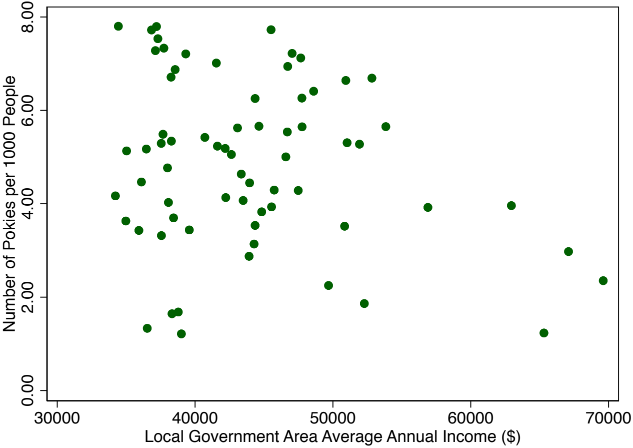
- If X and Y are random variables, then the variance of their sum is:

  $$var(X+Y) = var(X)+var(Y)+2cov(X,Y) = \sigma_X^2+\sigma_Y^2+2\sigma_{XY}$$

- If X and Y are independent, then they have 0 covariance implying that:

  $$var(X+Y) = var(X) + var(Y) = \sigma_X^2 + \sigma_Y^2$$

Pokies Per 1000 People vs. LGA Average Income



- Covariance is -3.22; Correlation is -0.237

Some Important Distributions for Doing Econometrics

- Normal Distribution
- Chi-Squared Distribution
- Student-t Distribution
- F-Distribution
- All of these distributions are foundational building blocks for developing our econometric toolkit for using data to figure out, empirically, how the world works
- If econometrics was a sports car, these distributions would represent car parts

Normal Distribution

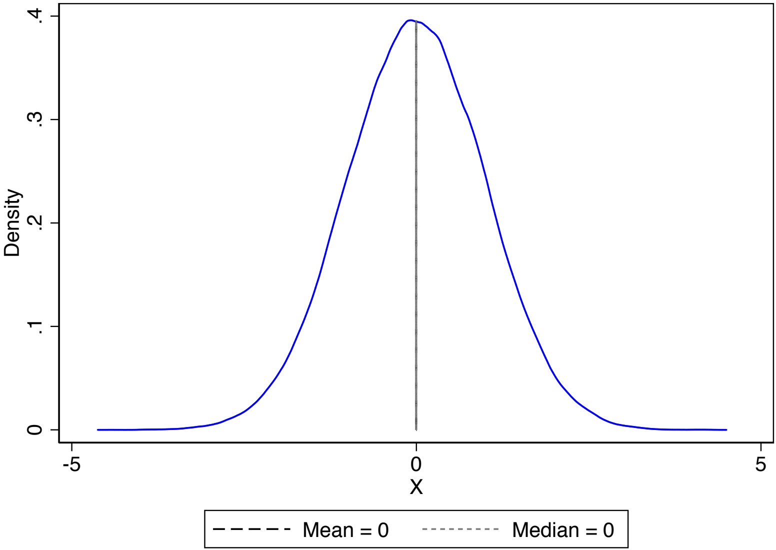
- Normal distributions have a bell-shaped probability density that is symmetric around its mean
- There are two parameters that govern the distribution's shape: mean μ and variance σ^2
- We denote a normal distribution with mean μ and variance σ^2 by $N(\mu, \sigma^2)$
- The standard normal distribution $N(0, 1)$ has mean 0 and variance 1.

Standardising Normal Random Variables

- If X is $N(\mu, \sigma^2)$ and $Z = \frac{X-\mu}{\sigma}$, then Z is $N(0, 1)$
- That is, we can always normalize a normal random variable X by subtracting its mean and dividing by its standard deviation to obtain another random variable Z that is N(0,1)
- We will use the normal distribution extensively for empirically testing hypotheses throughout the subject

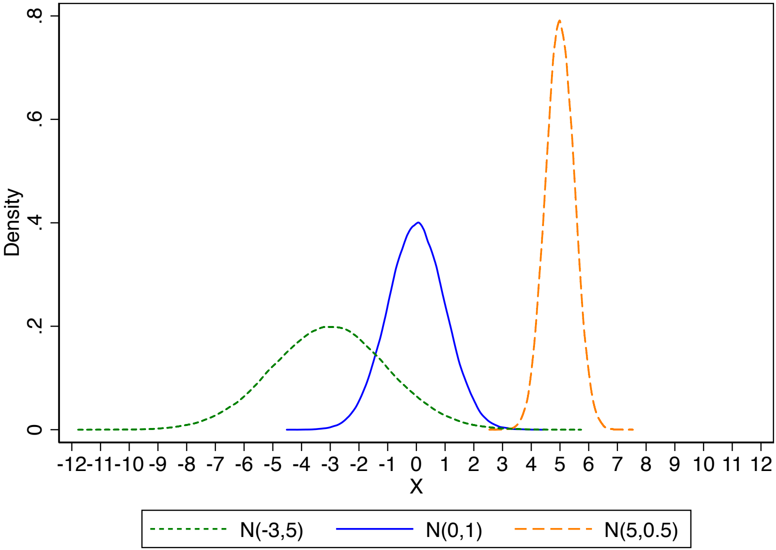
Normal Distribution Graphically

Example parameters: μ=0, $\sigma^2 = 1$



Normal Distribution Graphically

Different parameters



Chi-Squared Distribution

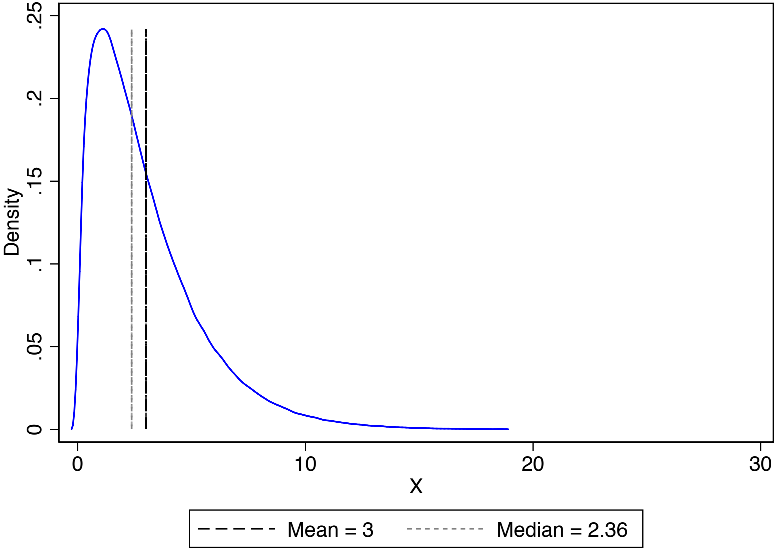
- The chi-squared distribution is also used for empirically testing hypotheses
- Formally, the chi-squared distribution is the distribution of the sum of m squared independent standard normal random variables $(Z_1, Z_2, \ldots, Z_m)$:

$$\chi^2_m = \sum_{i=1}^{m} Z_i^2$$

  where χ^2_m denotes a chi-squared random variable
- The key model parameter that governs the distribution's shape is m, the number of squared independent standard normal random variables that are being added up.
- We call m the degrees of freedom, where $E[\chi^2_m] = m$
- If Z_1, Z_2, Z_3 are 3 independent standard normal random variables, then $X = Z_1^2 + Z_2^2 + Z_3^2$ has a chi-squared distribution with 3 degrees of freedom with $E[X] = 3$

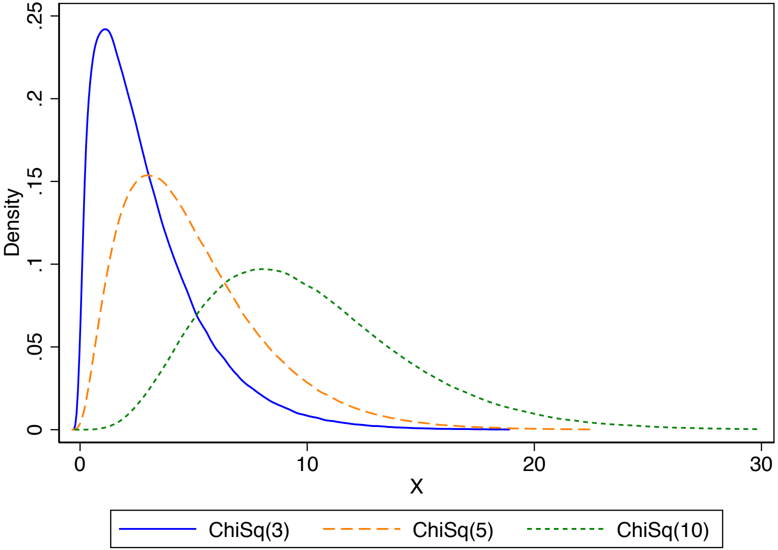
Chi-Square Distribution Graphically

Example parameters: $m=3$



Chi-Square Distribution Graphically

Different parameters



Student t Distribution

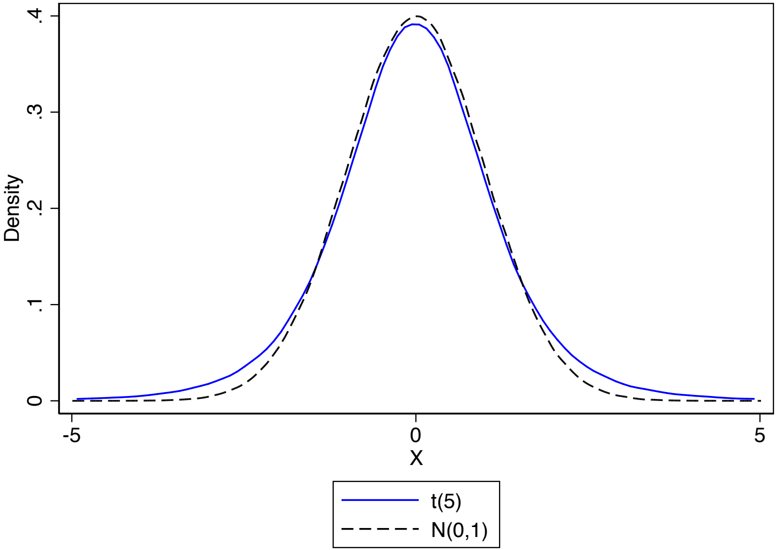
- The Student t Distribution is the ratio of a standard normal random variable (Z) divided by the square root of an independently distributed chi-squared random variable (u) divided by its degrees of freedom m:

$$t_m = \frac{Z}{\sqrt{u/m}}$$

  where t_m has a student t distribution with m degrees of freedom, which is the parameter that governs its shape
- When m is small (e.g., < 20), then the Student t distribution looks like a standard normal distribution but with fat tails
- As m gets big (e.g., > 30), the student t distribution is well-approximated by the standard normal distribution
- Student t distributions are also important for empirically testing hypotheses

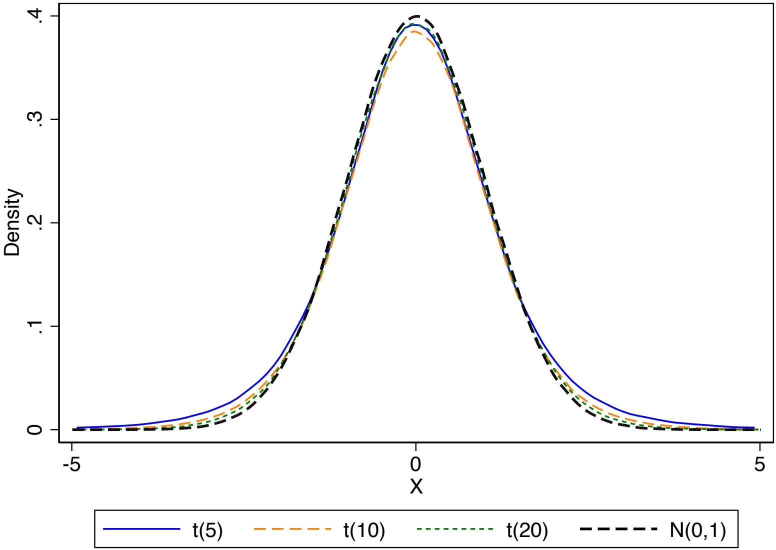
Student t Distribution Graphically

Example parameters: m=5



Student t Distribution Graphically

Different parameters



Beer, and Gosset's Student t-Distribution

F Distribution

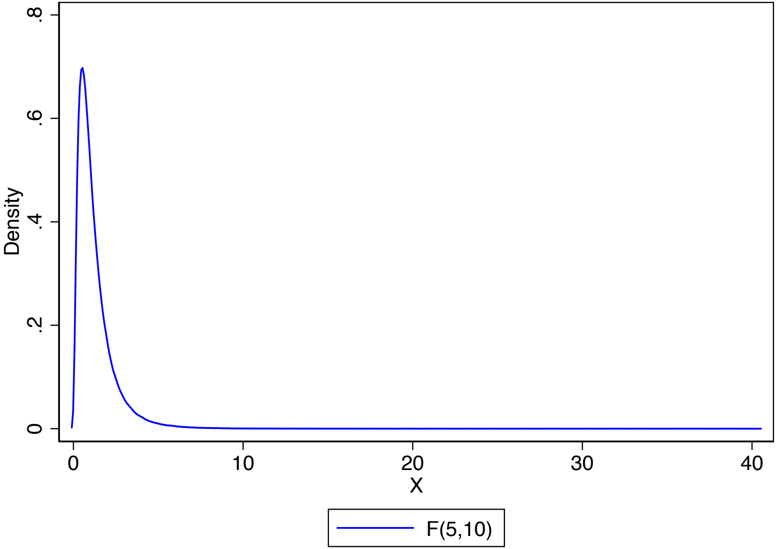
- The F distribution with m and n degrees of freedom is denoted $F_{m,n}$ and is defined as

$$F_{m,n} = \frac{u_1/m}{u_2/n}$$

  where u_1 and u_2 are independently distributed chi-square random variables with m and n degrees of freedom, respectively
- The parameters that govern the F distribution's shape are m and n
- We will also use the F distribution for empirically testing hypotheses in the subject

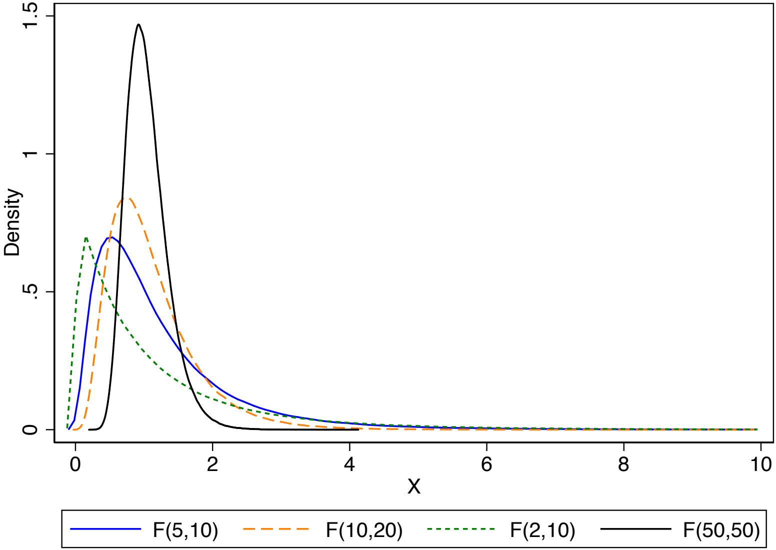
F Distribution Graphically

Example parameters: m=5, $n = 10$



F Distribution Graphically

Different parameters



Distributions Summary

- Normal, Chi-Square, t, and F distributions will show up throughout the subject
  - Normal and t distributions to test hypotheses about single parameter of an econometric model
  - Chi-Square and F distributions to jointly test hypotheses about multiple parameters of an econometric model
- Wait, in my first year statistics subject, I learned that degrees of freedom were related to sample size???
  - Yes, that is true in the specific context of statistics and hypothesis testing based on random samples
  - But we have not said anything about hypothesis testing yet; we have just focused on probability so far
- Generally, with probability distributions, degrees of freedom m in Chi-Square and t distributions, and degrees of freedom m and n in F distributions are distribution parameters just like mean μ and variance σ^2 are parameters for the Normal distribution

Random Sampling

- Almost all empirical analyses involves using samples of data
- We generate samples by randomly choosing observations from an underlying population
- And in many analyses, we interested in examining the sample average (or sample mean)
- Example: analysing income levels in Australia
  - data: random sample of 10,000 people
  - population: 24.13 million Australians
  - object of analysis: average income of the 10,000 randomly selected people
- Why are sample averages useful estimates for population means?

IID Random Draws

- Let $Y_1, Y_2, \ldots, Y_n$ be our random sample of n observations
- We assume that each of $Y_1, Y_2, \ldots, Y_n$ are drawn from the same population, which means they are all drawn from the same probability density
- Because each Y_i for $i = 1, \ldots, n$ has the same underlying density, we say they are identically distributed
- And because each Y_i for $i = 1, \ldots, n$ draw has no influence over the other values drawn in our sample, we say that $Y_1, Y_2, \ldots, Y_n$ are independently distributed.
- Putting it together, we say the random sample is independently and identically distributed, or iid for short

Sample Averages

- Random sample with iid draws: $Y_1, Y_2, \ldots, Y_n$
  - From the example: Y_i is income of household i, and there were $n = 10,000$ observations
- The sample average $\bar{Y}$ is computed as:

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

- So for a first random sample of $n = 10,000$ households, we could obtain a sample average #1 for income of

$$\bar{Y}^1 = 48,321$$

Sample Averages

- Keeping with our example, suppose we randomly selected (with replacement) another group of $n = 10,000$ people
- Unless we drew the exact same 10,000 people, we would almost surely obtain a different value for $\bar{Y}$
- Suppose we obtained a sample average #2 this time of

$$\bar{Y}^2 = \$51,991$$

- Suppose we again randomly selected (with replacement) a third sample of $n = 10,000$ people and calculated sample average #3 of

$$\bar{Y}^3 = \$46,421$$

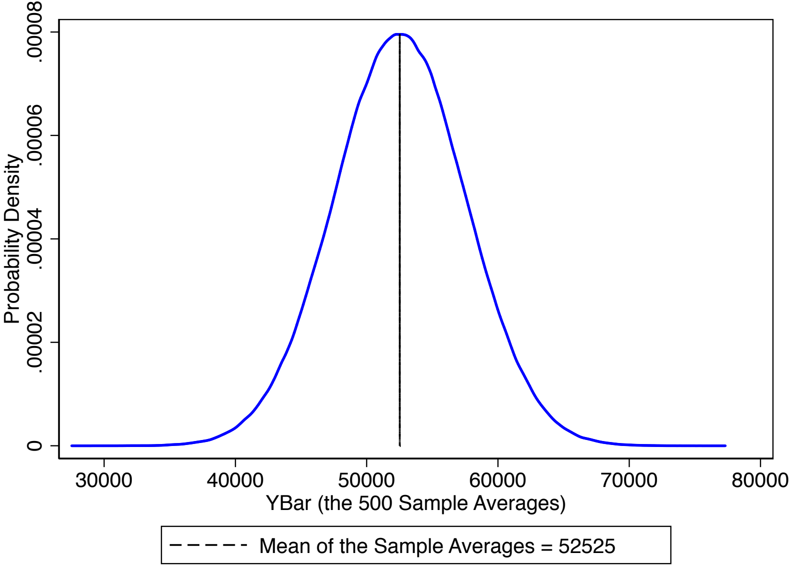
Sample Averages are Random Variables

- Pushing the example further, I could keep randomly draw samples of $n = 10,000$ people (with replacement) 500 times and construct a collection of 500 sample averages:

Random Sample	Sample Average ($\bar{Y}$)
#1	\$48,321
#2	\$51,991
#3	\$46,421
#4	\$49,444
#5	\$56,119
⋮	
#499	\$51,720
#500	\$48,848

- Just like ploting probability densities for data, we can plot the probability density function for the sample average!

Sample Averages are Random Variables



Sample Averages are Random Variables

- The upshot of this discussion is to illustrate that sample averages are random variables
- Because a sample $Y_1, Y_2, \dots, Y_n$ observations are random variables, any function of the observations is also a random variable, which includes the sample average $\bar{Y}$
- This means that the sample average is itself a draw from a distribution with an expected value and variance

Sample Averages are Random Variables

- Since the sample average is a random variable, it must have an expected value and variance.
- Expected value of the sample average $\bar{Y}$

$$E(\bar{Y}) = E\left(\frac{\sum_{i=1}^{n} Y_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \frac{1}{n}\sum_{i=1}^{n} \mu_Y = \frac{n\mu_Y}{n} = \mu_Y$$

- In words: the expected value of the sample average equals the population mean
- This is important: it means that, in expectation, the mean of a randomly-chosen sample can be used to estimate the underlying mean of the entire population
  - you can randomly sample the income of 1000 Australians and compute the sample average to estimate the average income of 26 million Australians
  - another example: polls for elections, also based on random samples, but tell us what the whole country is thinking

Sample Averages are Random Variables

- Our key result:
$$E(\bar{Y}) = \mu_Y$$
- This result is important: it means that, in expectation, the mean of a randomly-chosen sample can be used to estimate the underlying mean of the entire population
  - you can randomly sample the income of 1000 Australians and compute the sample average to estimate the average income of 26 million Australians
  - another example: polls for elections, also based on random samples, but tell us what the whole country is thinking

Sample Averages are Random Variables

- Variance of the sample average $\bar{Y}$

$$var(\bar{Y}) = \frac{\sigma_Y^2}{n}$$

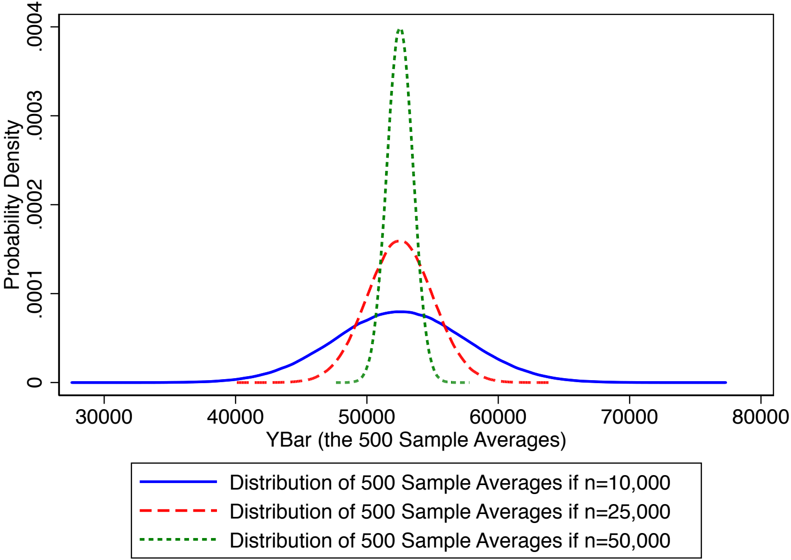
$\rightarrow$ the variance of the sample average is the population variance divided by the sample size
$\rightarrow$ see the textbook for details on the derivation

- Notice as $n \uparrow$ then $var(\bar{Y}) \downarrow$; if $n \rightarrow \infty$ then $var(\bar{Y}) \rightarrow 0$
$\rightarrow$ as the sample size becomes very large, the variance of the sample average becomes very small

- We can also compute the standard deviation of the sample average

$$std(\bar{Y}) = \frac{\sigma_Y}{\sqrt{n}}$$

Precision: how the Distribution of the Sample Average Changes With Sample Size n

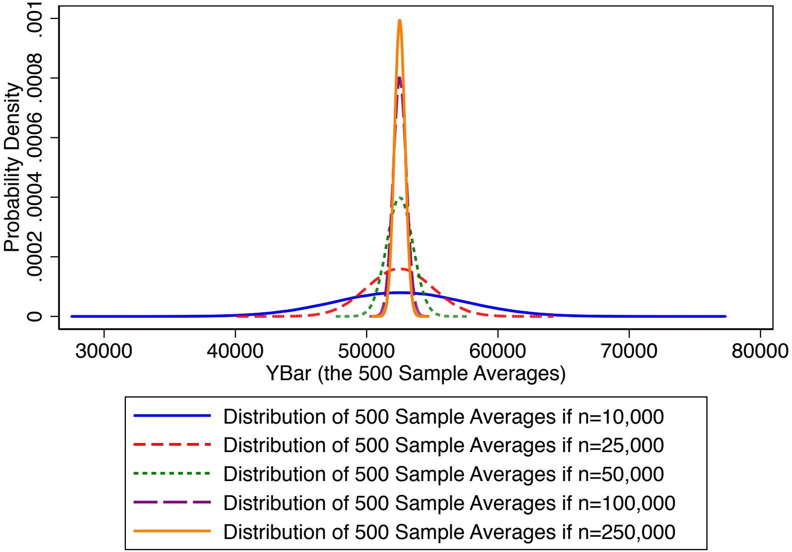


Law of Large Numbers

- Law of Large Numbers (LLN): as n gets large, the sample average $\bar{Y}$ will be very close to the population mean μ_Y with very high probability
  - if n is huge, there is still a very small chance that the sample average $\bar{Y}$ you compute will not be close to μ_Y (because $\bar{Y}$ is a random variable)
- LLN says $\bar{Y}$ is a consistent estimator of μ_Y as n gets big.
  - The sample average $\bar{Y}$ gets us the "right answer" regarding the population mean, and it is more likely to get something close to "right answer" as the sample size n grows

Law of Large Numbers Graphically

Distribution of 500 Sample Means as Sample Size *n* Gets Really Big

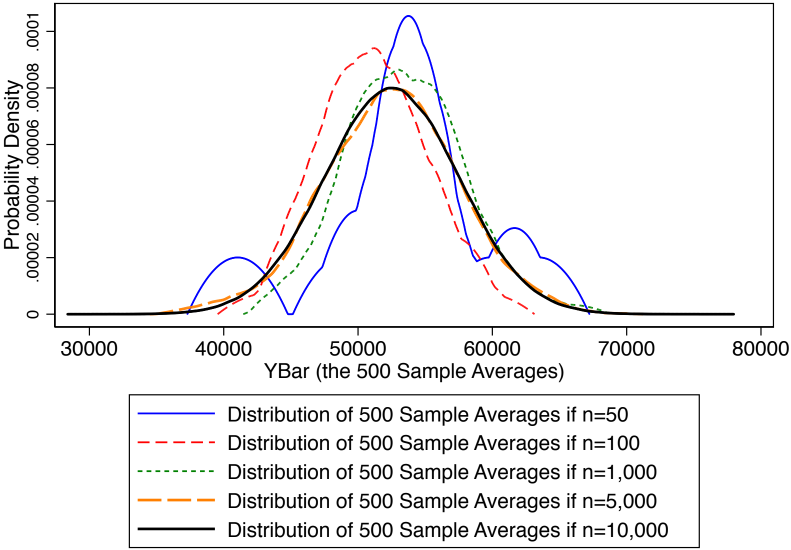


Central Limit Theorem

- Central Limit Theorem (CLT): As n gets large, the distribution of $\bar{Y}$ is approximately $N(\mu_Y, \sigma^2_{\bar{Y}})$
  - $E(\bar{Y}) = \mu_Y$ is the expected value of the sample average, which equals the population mean of Y
  - $\sigma^2_{\bar{Y}} = \frac{\sigma^2_Y}{n}$ is the variance of the sample average, which depends on the population variance of Y and the sample size n
- In words, as n gets large, the distribution of sample averages starts to look like the normal distribution with mean μ_Y

Central Limit Theorem Graphically

Distribution of 500 Sample Means as Sample Size n Goes from Really Small to Really Big



Central Limit Theorem is Critical for Econometric Analysis

- Remember that we compute the sample average $\bar{Y}$ from some underlying random sample of data $Y_1, Y_2, \ldots, Y_n$
- What is remarkable about the CLT is that we approach a normal distribution for the sample average $\bar{Y}$ as n gets big, even if the distribution of the random sample $Y_1, Y_2, \ldots, Y_n$ used to compute the sample average is nothing like a normal distribution!
- Why the CLT matters: it underpins statistical testing of hypotheses using random samples in all of econometrics.
- It is why the normal distribution is used so much for hypothesis testing in investigating whether empirical relationships between Y and X variables exist in the data