# Introduction to Econometrics

Dr. Yang Liang

Department of Economics
San Diego State University

## Hello! I am Yang Liang!

My research:

I study labor, health, and international economics, focusing on two primary areas:

1. The economics of risky behaviors among adolescents and young adults. I examine their responses to tobacco control policies and marijuana legalization. What are the intended and unintended consequences of vape curbing policies? How do consumers react to the legalization of recreational marijuana?

2. The labor market consequences of technological change and globalization. I analyze how firms and workers respond to import penetration from China, as well as the impact of AI and robotics.

# Agenda

- Administrative stuff
- Introductions and class description
- Begin recap: probablity and statistics

# Admin Stuff

- Course website (Canvas)
  - handouts/slides
  - notes
  - problem sets
- Textbooks: Wooldrige 6e; Stock and Watson (4e)
- Programming in **Stata**, R
- Problem sets: homework combining econometrics and Stata tasks
- Exams: midterm and final
- Office hours
- CHEPS seminars: `https://cheps.sdsu.edu/seminar`

# Our Goals for this Class

- Learn core tools of **econometrics** (theory and empirical)
- You will also learn to use '`Stata`' programming language.

# Our Goals for this Class

- Learn core tools of **econometrics** (theory and empirical)
- You will also learn to use '`Stata`' programming language.
- ***What is econometrics***?
  - Econometrics: A set of *techniques and methods* to answer (economic) questions with data.
  - Okay so what are we doing here exactly?
  - Economists work with data (a lot)
  - Statisticians also work with data
  - so do data scientists
  - the *goals* for these groups differ considerably

# Causality and Prediction

- Data scientists are generally concerned with prediction
- They want to use the data at hand to predict what comes next
- They generally don't care why they're making the prediction they are
- This can be really handy for certain tasks - "is this picture a cat or a dog?" "what's the probability that a customer with qualities X, Y, and Z will end up purchasing our good?" "do you have diabetes?"

# Causality and Prediction

- Econometricians, on the other hand, care almost exclusively about why
- Data scientists want to minimize prediction error
- Econometricians want to minimize *inference* and *identification* error
- We want to correctly understand the underlying data generating process (mechanisms, like Doctors)
- some examples...

# Answering Important Questions with Econometrics

**Does immigration *lead* to lower wages and/or higher unemployment for locals?**

**Does raising the minimum wage *reduce* employment for the low-skilled?**

**Does getting a college degree *afford* higher wages?**

**Do higher public debt levles *lead* to lower economic growth?**

**Does import penetration from China *destroy* manufacturing jobs?**

# Inference Error and Randomness

- One big problem we face when trying to figure out how variables relate to each other is randomness
- This is simply the fact that, when we gather data, we can only possibly get a subsample of all the data
- So, just by random chance, the relationship we get in our data might not be quite the same as the true relationship

# Inference Error and Randomness

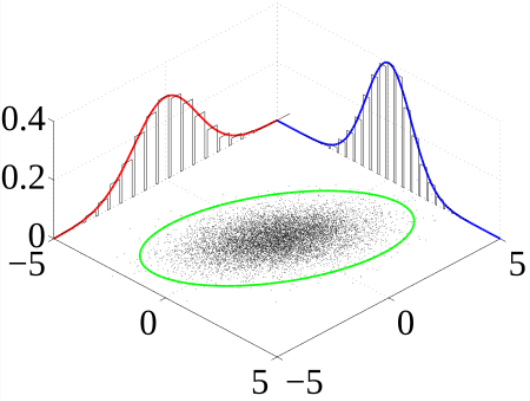
- One big problem we face when trying to figure out how variables relate to each other is randomness
- This is simply the fact that, when we gather data, we can only possibly get a subsample of all the data
- So, just by random chance, the relationship we get in our data might not be quite the same as the true relationship

- So if we look in a data set and see that $X$ and $Y$ appear to be positively related to each other...
- Are they actually positively related, or is that just random chance?
- Even if they are positively related, maybe we're understating or overstating how positively related

# Inference Error and Randomness

- If the true relationship is 0, then in the data we'll see a positive relationship half the time, and a negative relationship half the time
- Even though the truth is 0!
- How can we properly make an inference about whether the relationship is 0 or not (or positive, or negative, or how positive or negative), taking into account this randomness?
- That's being careful about inference. The statisticians teach us all about this!

# Data Generating Process, Example



# Stata Example , two independent normal variables

```
clear
set obs 10000
gen x = rnormal()
gen y = 2 * rnormal()
// graph start
scatter y x, mcolor(gray) msize(medium) msymbol(oh) ///
graphregion(color(white)) ///
xtitle("Variable X - N(0,1)") ytitle("Variable Y - N(0,1)") ///
title("Scatter Plot of Two Independent Normally Distributed Variables")
```

# Stata Example

```
// with fitted a line
// graph start
twoway ///
(scatter y x, mcolor(gray) msize(medium) msymbol(oh)) ///
(lfit y x), ///
graphregion(color(white)) ///
xtitle("Variable X - N(0,1)") ytitle("Variable Y - N(0,1)") ///
title("Scatter Plot of Two Independent Normally Distributed Variables")
```

# Redo with Selection

```
// with fitted a line
// some data selection
gen z=0
replace z=1 if int(y)==int(1*x)
keep if z==1
// graph start
twoway ///
(scatter y x, mcolor(gray) msize(medium) msymbol(oh)) ///
(lfit y x), ///
graphregion(color(white)) ///
xtitle("Variable X - N(0,1)") ytitle("Variable Y - N(0,1)") ///
title("Scatter Plot of Two Independent Normally Distributed Variables")
```

# Identification Error

- What is identification error?
- Identification is how you link the result you see with the conclusion you draw from it
- For example, say you observe that kids who play video games are more aggressive in everyday life (result), and you conclude from that result that video games make kids more aggressive (conclusion)
- If seeing that result is actually evidence for that conclusion, then we are properly identified (a broken clock is right twice a day)

# Identification Error

- But if there's another reason why we might see that result, i.e. if the same result could give us a different conclusion, like kids who are aggressive play more video games or people notice aggression more when kids play video games, then we have made an identification error - our result was not identified!
- Identification error is when your result in the data doesn't actually have a clear theoretical ("why" or "because") interpretation
- For example, if you observe that people tend to wear more shorts on days they eat more ice cream, and you conclude that eating ice cream makes you put on shorts, you have committed an identification error

# Identification Error

- Other examples
- Returns to education: A classic causal question in education literature, to what extent does educational attainment affect an individual's earnings?
- Minimum wage and employment: How does an increase in the minimum wage affect employment?
- UI and job search: Unemployment benefits and job search: How do unemployment benefits influence duration of employment?
- Labor market discrimination: How does discrimination in the labor market affect wages and employment opportunities?
- Trade/AI and job loss: How does the import penetration from China or adoption of automation affect the manufacturing employment?
- Workplace flexibility and productivity: Does WFH reduce employee productivity?

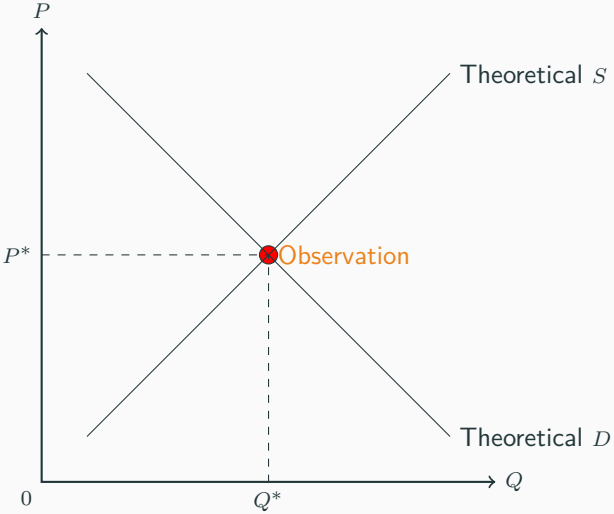
# Data Generating Process (DGP)

- To avoid identification error, economists think closely about the data generating process (DGP)
- What is a data generating process?
- The data generating process is the true set of laws that determine where our data comes from
- For example, if you hold a rock and drop it, it falls to the floor
- What is the data we observe? (Hold the rock & Rock is up) and (Let go & Rock is down)
- What is the data generating process? Gravity makes the rock fall down when you drop it

# Econ DGP Example

- Another example is a model of supply and demand
- We observe prices and quantities in a competitive market
- What led to those being the prices and quantities we see?
- The supply and demand model and its equilibrium, theory is important!

# Data Generating Process

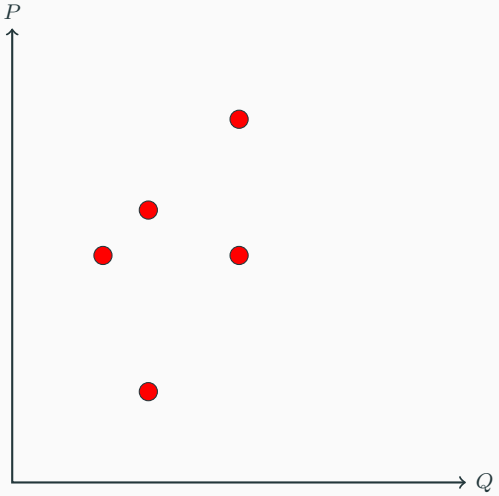


# DGP

- The prices that we observe come from that theoretical construct
- When we see the prices and quantities moving, according to our theory, it's because the S and D lines are moving But we can't see the S and D lines
- Our goal: use the observations we do see to infer what the theoretical model (data generating process) is

# DGP

- But It's hard! What inference about S and D can we draw from these observations?



# Causality

- A data generating process can be described by a series of equations that describe where the data comes from. For example:

$$X = \gamma_0 + \gamma_1 \epsilon + \nu$$
$$Y = \beta_0 + \beta_1 X + \epsilon$$

- This says $X$ is caused by $\epsilon$ and $\nu$, and $Y$ is caused by $X$ and $\epsilon$
- the truth is that an increase in $X$ causally increaes $Y$ by $\beta_1$
- the goal of econometrics is to be able to estimate what $\beta_1$ is accurately

# Causality in Directed Acyclic Graph (DAG)

- What can also represent this set of relationships as a graph (DAG), with arrows telling you what variables cause each other

$$\begin{tikzcd}
\nu \arrow[d] & & \epsilon \arrow[dl] \arrow[dr] & \\
X \arrow[rr] & & & Y
\end{tikzcd}$$

# Causality

- We do this because most of the relationships we're interested in are causal - we want to know, if we could reach in and manipulate $X$, would $Y$ change as a results, and how much?
- *recall* our previous examples
- You mostly only observe *correlation*
- And *correlation* is not the same thing as *causation*
- "Credibility Revolution" in econometrics over the past 30 years (2022 Economics Nobel awarded to some of the main protagonists of this revolution)

# Causality

- A given correlation, (e.g, negative, minimum wage and employment; positive, returns to education), can be consistent with a number of different causal relationships
- As econometricians, we need to figure out which one it is!
- How can we narrow it down?
- How many possible diagrams?

# Causality in Directed Acyclic Graph (DAG), again

$$\begin{tikzcd}
 & \epsilon \arrow[dl] \arrow[dr] & \\
X \arrow[rr] & & Y
\end{tikzcd}$$

$$\begin{tikzcd}
 & \epsilon \arrow[dl] & \\
X \arrow[rr] & & Y
\end{tikzcd}$$

$$\begin{tikzcd}
 & \epsilon & \\
X \arrow[rr] & & Y
\end{tikzcd}$$

$$\begin{tikzcd}
 & \epsilon \arrow[dl] \arrow[dr] & \\
X & & Y
\end{tikzcd}$$

$$\begin{tikzcd}
 & \epsilon \arrow[dr] & \\
X \arrow[rr] & & Y
\end{tikzcd}$$

# Causality

- Which one is the worst? (X has nothing to do with Y)
- Which one(s) will lead to bias? why? Which direction?
- The key is in $\epsilon$
- $\epsilon$ is everything that determines $Y$ other than $X$
- Perhaps the health of the economy, or the policies that area has chosen, many other things
- **So we almost certainly have a graph with** $\epsilon \to Y$
- If those things also affect $X$, we are in the first DAG.

# Endogeneity

- So "correlation isn't causation" isn't quite complete
- It's more "only certain correlations are causal"
- Many correlations are beset by these problems like *endogeneity*, i.e., the presence of another variable in $\epsilon$ related to both X and Y, giving the effect a "back door" (ommitted variable bias)
- so the correlation reflects both the causal effect and the influence of $\epsilon$ (bias part)

# Next Semester, Random Experiments

- The golden approach around this problem is to run a randomized control trial/experiment (RCT)
- If you can randomly assign $X$, then we know we're not in the first DAG, because our graph becomes

$$
\begin{tikzcd}
\text{Random Assignment} \arrow[d] & \epsilon \arrow[dr] & \\
X \arrow[rr] & & Y
\end{tikzcd}
$$

# RCT

- Although they have their own problems, (external validity - your sample might not represent the population, statistical mistakes, respondents react differently knowing they're in an experiment, etc...)
- Many questions for which we *can't* run an experiment (impossible, infeasible, immoral)
- So one way we can think about solving this endogeneity problem with econometrics is to use our *observational* data in such a way that it behaves as though there were an experiment being run
- You'll need the solid foundation from this course to pull off the cool tricks later.

# Questions

- What does it mean to say that X has a causal effect on Y?
- Why might the relationship between X and Y in data not be the same as the causal effect?
- Give me an example of observational data?
- Does getting an MBA make you a better manager? What are X and Y here? What would be in the error term? What are the likely endogeneity problems here?

# Suprious Correlations

- Let's visit this site all about "spurious" correlations (i.e. correlations that almost certainly do not reveal a true effect of one variable on the other): `https://tylervigen.com/spurious-correlations`
- Take a look at how easy it is to find variables that are related statistically, even though clearly neither causes the other
- Do you think this correlation is an example of inferential error (just random chance) or identification error (truly related, but not because one causes the other)? Why? which direction? Why?

# Next Time

- We'll begin with Statistics Review using some Stata (so make sure you have access to it)

# Probability and Statistics Review

# Probability

- Let's begin by introducing the notions of probability, randomness, and random variables....
- The use of probability to measure uncertainty and variability began hundreds of years ago with the "study" of gambling.
- Generally speaking, probability is the *chance* that something (an event) will happen.
- The probability of an event or outcome is the proportion of the time it occurs in the long run - this is called the frequency interpretation of probability.
- For an event A, the number $\Pr(A)$ indicates the probability that A will occur.

# Probability

- In order to satisfy the mathematical definition of probability, $\Pr(A)$ must satisfy certain conditions (axioms).
  - For every event A, $\Pr(A) \geq 0$.
  - If an event A is certain to occur, then $\Pr(A) = 1$.
  - If A and B are *mutually exclusive*, then $\Pr(A \text{ or } B) = \Pr(A) + \Pr(B)$.
- From these axioms, one can derive a number of properties like

$$0 \leq \Pr(A) \leq 1$$

and

$$\Pr(A^c) = 1 - \Pr(A)$$

# Random Variables and Distributions

- A random variable is simply a numerical summary of a random outcome (an outcome determined by an experiment).
- A random variable is a real valued function, defined on the set of possible outcomes (i.e. the sample space), that assigns a real number to every possible outcome.
  - e.g the number of heads appearing in 10 flips of a coin
- Said in a different way, a random variable is a variable whose value is subject to variations due to chance.
- A random variable can take on a set of possible different values, each with an associated probability
- The collection of all these probabilities is the distribution of the random variable.

# Random Variables and Distributions

- There are two major classes of random variables and distributions
  - Discrete and Continuous
- A **discrete** random variable takes on only a discrete set of values.
- A **continuous** random variable takes on a continuum of values.

# Discrete Random Variables

- We characterize or describe a **discrete** random variable $X$ with a probability function (pf), or probability distribution.
- A pf lists the probability for each possible discrete outcome.
- The pf of a discrete random variable is defined as the function $f$ such that for every real number $x$,

$$f(x) = \Pr(X = x)$$

  where $X$ represents a random variable and $x$ represents a *realization* of that random variable.

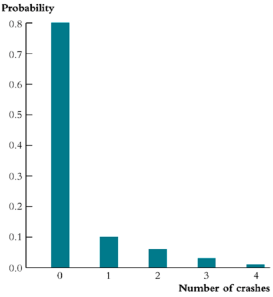
# Probability Function

**TABLE 2.1 Probability of Your Computer Crashing *M* Times**

| | Outcome (number of crashes) | | | | |
|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 |
| Probability distribution | 0.80 | 0.10 | 0.06 | 0.03 | 0.01 |
| Cumulative probability distribution | 0.80 | 0.90 | 0.96 | 0.99 | 1.00 |

**FIGURE 2.1 Probability Distribution of the Number of Computer Crashes**

The height of each bar is the probability that the computer crashes the indicated number of times. The height of the first bar is 0.80, so the probability of 0 computer crashes is 80%. The height of the second bar is 0.1, so the probability of 1 computer crash is 10%, and so forth for the other bars.



# Cumulative Distribution Function (CDF)

- Another way of characterizing the distribution of a random variable is with a cumulative distribution function (cdf)
- The cdf lists the probability that a random variable is less than or equal to a specific value

$$F(x) = \Pr(X \leq x)$$

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\Pr(X \leq x)$ | 0.8 | 0.9 | 0.96 | 0.99 | 1.0 |

# Continuous Random Variables

- A random variable $Y$ that can take on any real value within some range is a continuous random variable.
  - Time, temperature, height...
- We typically speak of interval probabilities (i.e. the probability that $Y$ will take on some subset of values)

$$\Pr(a \leq Y \leq b)$$

- For continuous random variables, the probability of a particular value occurring is equal to zero

$$\Pr(Y = y) = 0$$

- Note that probability zero does not mean impossible.

# Probability Density Function (pdf)

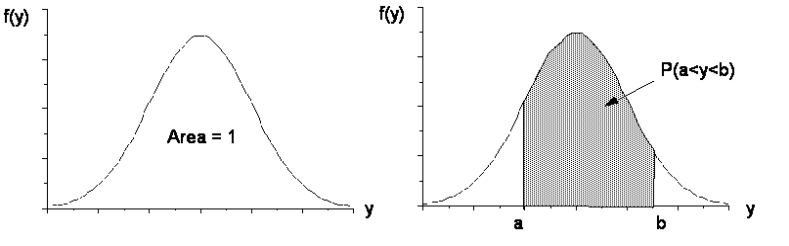
- The probabilities associated with a continuous random variable $Y$ are determined by the pdf of $Y$.
- The pdf of $Y$, denoted $f(y)$, has the following properties:
  1. $f(y) \geq 0$, for all $y$.
  2. The probability that the uncertain quantity $Y$ will fall in the interval $(a, b)$ is equal to the area under $f(y)$ between $a$ and $b$ :

$$P(a < y < b) = \int_a^b f(y)dy.$$

  3. The total area under the entire curve of $f(y)$ is equal to 1

$$\int_{-\infty}^{\infty} f(y)dy = 1$$

# Probability Density Function (pdf)



# CDF of Continuous Random Variables

- The definition of a cdf for a continuous random variable is the same as that of a discrete random variable

$$F(y) = \Pr(Y \leq y)$$

- With a continuous random variable, the cdf is a continuous function over the entire real line, so we can write down a formula

$$F(y) = \Pr(Y \leq y) = \int_{-\infty}^{y} f(t)dt$$

# CDF of Continuous Random Variables

- We can easily see that

$$\Pr(Y > y) = 1 - F(y)$$

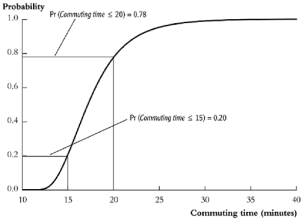
and

$$\Pr(y_1 < Y \leq y_2) = F(y_2) - F(y_1)$$

# CDF of Continuous Random Variables

- In practice, the CDF allows us to calculate the probability for any interval.

**FIGURE 2.2** **Cumulative Distribution and Probability Density Functions of Commuting Time**




**(a)** Cumulative distribution function of commuting time

Figure 2.2a shows the cumulative probability distribution (or c.d.f.) of commuting times. The probability that a commuting time is less than 15 minutes is 0.20 (or 20%), and the probability it is less than 20 minutes is 0.78 (78%). Figure 2.2b shows the probability density function (or p.d.f.) of commuting times. Probabilities are given by areas under the p.d.f. The probability that a commuting time is between 15 and 20 minutes is 0.58 (58%), and is given by the area under the curve between 15 and 20 minutes.

- $\Pr(15 < CT \leq 20) = \Pr(CT \leq 20) - \Pr(CT \leq 15) = F(20) - F(15) = .58$
- $\Pr(CT > 20) = 1 - \Pr(CT \leq 20) = 1 - F(20) = 0.22$

# Measures of Central Tendency for Distributions

Mode

- The *mode* is the value that occurs with the greatest probability.
  - Example: What is the modal age of students in this class?

# Measures of Central Tendency for Distributions

Median

- The *median* is the value such that the probability of the random variable being less than or equal to that value is at least 50% and the probability of the random variable being greater than or equal to that value is at least 50%.
  - Example: What is the median age of students in this class?

# Measures of Central Tendency for Distributions

Median

- The *median* is the value such that the probability of the random variable being less than or equal to that value is at least 50% and the probability of the random variable being greater than or equal to that value is at least 50%.
  - Example: What is the median age of students in this class?
  - Case 1

| Age | 19 | 20 | 21 | 22 |
|---|---|---|---|---|
| % | .4 | .2 | .2 | .2 |

# Measures of Central Tendency for Distributions

Median

- The *median* is the value such that the probability of the random variable being less than or equal to that value is at least 50% and the probability of the random variable being greater than or equal to that value is at least 50%.
  - Example: What is the median age of students in this class?
  - Case 1

| Age | 19 | 20 | 21 | 22 |
|---|---|---|---|---|
| % | .4 | .2 | .2 | .2 |

  - Case 2

| Age | 19 | 20 | 21 | 22 |
|---|---|---|---|---|
| % | .4 | .05 | .35 | .2 |

# Mean (Expected Value)

- The *mean* or expected value of a random variable $X$ is the **weighted average** of all its possible outcomes, weighted by the probabilities of those outcomes.
- The expected value or mean of $X$ is typically denoted by $E(X)$ or $\mu_X$.
- Unlike the mode, it's unique (like a center of gravity).
- However, one drawback of the mean is it's sensitivity to outliers

# Mean (Expected Value)

- The *mean* or expected value of a random variable $X$ is the **weighted average** of all its possible outcomes, weighted by the probabilities of those outcomes.
- The expected value or mean of $X$ is typically denoted by $E(X)$ or $\mu_X$.
- Unlike the mode, it's unique (like a center of gravity).
- However, one drawback of the mean is it's sensitivity to outliers
  - When Bill Gates walks into a bar...

# Mean (Expected Value)

- For a discrete random variable $X$,

$$E(X) = x_1 \Pr(x_1) + ... + x_n \Pr(x_n) = \sum_{i=1}^{n} x_i \Pr(x_i) = \sum_{i=1}^{n} x_i f(x_i)$$

- Example: Expected value of throwing a die:

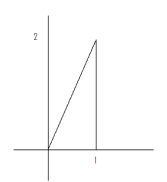
$$\begin{aligned} E(X) &= \frac{1}{6}1 + \frac{1}{6}2 + \frac{1}{6}3 + \frac{1}{6}4 + \frac{1}{6}5 + \frac{1}{6}6 \\ &= \frac{1}{6}(1+2+3+4+5+6) = 3.5 \end{aligned}$$

- For a continuous random variable $X$ with probability density function $f(x)$, the mean of $X$ is defined as

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

# Example

$f(x) = 2x$ for $0 < x < 1$ (and 0 otherwise)




The expected value is

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = \int_{0}^{1} x(2x)\,dx = \int_{0}^{1} 2x^2 dx = \tfrac{2}{3}x^3 \mid_0^1 = \tfrac{2}{3}$$

Question: what's $F(x)$?

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{x} 2t dt = t^2 \mid_0^x = x^2$$

What about the median?

$\frac{2}{3}$? $F(\frac{2}{3}) = \frac{4}{9} < \frac{1}{2}$ Nope. $F(\frac{1}{\sqrt{2}}) = \frac{1}{2}$

# Expected Value of a Function

- We know that if $X$ is a random variable with pdf $f(x)$ we can calculate $E(X)$ as either

$$E(X) = \sum_{i=1}^{n} x_i f(x_i) \quad \text{or} \quad E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

- But what if we want to calculate $E(X^2)$ or $E(\ln(X))$?

# Expected Value of a Function

- In general, what's the expected value of a function $g(\cdot)$ of $X$?

$$E(g(X)) = \sum_{i=1}^{n} g(x_i) f(x_i) \quad \text{or} \quad E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

- For example, for a continuous distribution

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$$

# Measures of Dispersion for Distributions

- The *variance* of a random variable measures the spread or dispersion of the variable around its mean

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

- In the discrete case, this is simply the weighted average of the squared deviations of $X$ from its mean

$$Var(X) = \sum_{i=1}^{n} [x_i - E(X)]^2 \Pr(x_i)$$

- For a continuous random variable $X$ with probability density function $f(x)$, the variance of $X$ is

$$Var(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 f(x) dx.$$

# Variance & Moments

- Another (equivalent) formula that can be used in either case is:

$$Var(X) = E(X^2) - (E(X))^2$$

where the **second moment** $E(X^2)$ is defined as

$$E(X^2) = (x_1)^2 \Pr(x_1) + (x_2)^2 \Pr(x_2) + ... + (x_n)^2 \Pr(x_n)$$

or

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$$

- Other (higher order) moments are defined similarly.

# Standard Deviation

- The variance of $X$ is denoted $Var(X)$ or $\sigma_X^2$.
- The standard deviation of $X$, denoted $\sigma_X$, is the square root of $Var(X)$.
- The standard deviation is useful because it has the same units of $X$, while the units of the variance are the units of the square of $X$.
- A large standard deviation and variance means that the probability distribution is quite spread out: a large difference between the outcome and the expected value is anticipated.

# Example

| $x$ | $\Pr(x)$ | $x \Pr(x)$ | $(x)^2 \Pr(x)$ |
|---|---|---|---|
| 1 | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |
| 2 | $\frac{1}{6}$ | $\frac{1}{3}$ | $\frac{2}{3}$ |
| 3 | $\frac{1}{6}$ | $\frac{1}{2}$ | $\frac{3}{2}$ |
| 4 | $\frac{1}{6}$ | $\frac{2}{3}$ | $\frac{8}{3}$ |
| 5 | $\frac{1}{6}$ | $\frac{5}{6}$ | $\frac{25}{6}$ |
| 6 | $\frac{1}{6}$ | $1$ | $6$ |
| | | $\downarrow$ | $\downarrow$ |
| | | $E(X) = 3.5$ | $E(X^2) = 15\frac{1}{6}$ |

- $E(X) = \sum_{i=1}^{n} x_i \Pr(x_i) = 3.5$
- $E(X^2) = \sum_{i=1}^{n} (x_i)^2 \Pr(x_i) = 15\frac{1}{6}$
- $Var(X) = E(X^2) - (E(X))^2 = 15\frac{1}{6} - 3.5^2 = 2.92$
- $\sigma_X = 1.71$

# Example

$f(x) = 2x$ for $0 < x < 1$ (and 0 otherwise)

We already know $E(X) = \frac{2}{3}$. What's $Var(X)$?

$$Var(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 f(x)dx = \int_0^1 \left(x - \frac{2}{3}\right)^2 2xdx$$

$$= \int_0^1 \left(2x^3 - \frac{8}{3}x^2 + \frac{8}{9}x\right) dx = \left[\left(\frac{1}{2}x^4 - \frac{8}{9}x^3 + \frac{4}{9}x^2\right)\right]_0^1 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$

or (using the other formula)

$$Var(X) = E(X^2) - (E(X))^2 = \int_0^1 x^2 (2x)\, dx - (\frac{2}{3})^2$$

$$= \int_0^1 2x^3 dx - \frac{4}{9} = \left[\frac{1}{2}x^4\right]_0^1 - \frac{4}{9} = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$

# Expected Value and Variance of a Linear Function

- If $Y = a + bX$, then $E(Y) = a + bE(X)$

  Example: Suppose $E(X) = 5, Y = 3X - 5, Z = -3X + 15$
  $E(Y) = 3 \cdot E(X) - 5 = 3 \cdot 5 - 5 = 10$
  $E(Z) = -3 \cdot E(X) + 15 = -3 \cdot 5 + 15 = 0$

- If $Y = a + bX$, then $Var(Y) = b^2 Var(X)$

  Example: $Var(X) = 5, Y = 3X - 5, Z = -3X + 15$
  $Var(Y) = 9 \cdot Var(X) = 9 \cdot 5 = 45$
  $Var(Z) = 9 \cdot Var(X) = 9 \cdot 5 = 45$

These properties can be easily proven using the definitions of expectation and variance...

# Relationships Between Two Random Variables

- Often we are interested in the relationship between a pair (or more) of variables
  - education and salary
  - quantity and price
  - price and quality
  - stock price and earnings
- We need to introduce some more tools to allow us to analyze these relationships

# Joint and Marginal Distributions

- The joint probability distribution is the probability that two random variables ($X$ and $Y$) simultaneously take on particular values ($x$ and $y$)
- The joint probability distribution function can be written as $P(X = x, Y = y) = f(x, y)$

**TABLE 2.2 Joint Distribution of Weather Conditions and Commuting Times**

| | Rain ($X = 0$) | No Rain ($X = 1$) |
|---|---|---|
| Long Commute ($Y = 0$) | 0.15 | 0.07 |
| Short Commute ($Y = 1$) | 0.15 | 0.63 |

- What if we also track what day of the week it is?

# Joint and Marginal Distributions

- Can we recover the distribution of $X$ or $Y$ alone from their joint distribution?
- The **marginal** probability distribution of a r.v. $Y$ is just its own probability distribution.
- It can be computed from the joint probability distribution of $X$ and $Y$ as:

$$P(Y = y) = \sum_{\text{all values of } X} P(X = x_i, Y = y)$$

or (more compactly)

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

# Joint and Marginal Distributions

**TABLE 2.2 Joint Distribution of Weather Conditions and Commuting Times**

| | 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 |

Rain:

$$P(X = 0) = P(X = 0, Y = 0) + P(X = 0, Y = 1)$$

$$= .15 + .15 = .30$$

Long Commute:

$$P(Y = 0) = P(X = 0, Y = 0) + P(X = 1, Y = 0)$$

$$= .15 + .07 = .22$$

# Conditional Distribution

- Sometimes knowing something about one random variable tells us something about another
  - Knowing it's going to rain tells you something about your expected commute time
- The distribution of a random variable $Y$ conditional on another random variable $X$ taking on a specific value is the **conditional** distribution of $Y$ given $X$
- Formally,

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

# Conditional Distribution: Example

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

**TABLE 2.3** Joint and Conditional Distributions of Computer Crashes ($M$) and Computer Age ($A$)

**A. Joint Distribution**

| | $M = 0$ | $M = 1$ | $M = 2$ | $M = 3$ | $M = 4$ | **Total** |
|---|---|---|---|---|---|---|
| Old computer ($A = 0$) | 0.35 | 0.065 | 0.05 | 0.025 | 0.01 | 0.50 |
| New computer ($A = 1$) | 0.45 | 0.035 | 0.01 | 0.005 | 0.00 | 0.50 |
| **Total** | 0.8 | 0.1 | 0.06 | 0.03 | 0.01 | 1.00 |

**B. Conditional Distributions of $M$ given $A$**

| | $M = 0$ | $M = 1$ | $M = 2$ | $M = 3$ | $M = 4$ | **Total** |
|---|---|---|---|---|---|---|
| $\Pr(M \mid A = 0)$ | 0.70 | 0.13 | 0.10 | 0.05 | 0.02 | 1.00 |
| $\Pr(M \mid A = 1)$ | 0.90 | 0.07 | 0.02 | 0.01 | 0.00 | 1.00 |

$$P(M = 1 \mid A = 1) = \frac{P(M{=}1, A{=}1)}{P(A{=}1)} = \frac{.035}{.5} = .07$$

# Conditional Expectation

- We can also calculate the expected value of one random variable, conditional on the realized value of another random variable
- These *conditional expectations* are just...expectations! But we must be careful to use conditional probabilities instead of marginal probabilities

$$E[Y \mid X = x] = \sum_{i=1}^{n} y_i P\left(Y = y_i \mid X = x\right)$$

- Example:

$$E(M \mid A = 1) = 0 \cdot .9 + 1 \cdot .07 + 2 \cdot .02 + 3 \cdot .01 + 4 \cdot 0 = .14$$

- Similarly

$$E(M \mid A = 0) = .56$$

# Conditional Expectation

**TABLE 2.2 Joint Distribution of Weather Conditions and Commuting Times**

| | 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 |

- Suppose a long commute takes 25 minutes and a short commute 10 minutes.
- What's the expected commute time if it doesn't rain?

$$E(Y \mid X = 1) = 10P(Y = 10 \mid X = 1) + 25P(Y = 25 \mid X = 1)$$

$$= 10\frac{P(Y=10,X=1)}{P(X=1)} + 25\frac{P(Y=25,X=1)}{P(X=1)} = 10\left(\frac{.63}{.7}\right) + 25\left(\frac{.07}{.7}\right)$$

$$= 10\,(.9) + 25\,(.1) = 11.5 \text{ minutes}$$

# Conditional Variance

- Similarly, the variance of $Y$ conditional on $X$ is the variance of the conditional distribution of $Y$ given $X$

$$Var\left(Y \mid X = x\right) = \sum_{i=1}^{n} \left[y_i - E[Y \mid X = x]\right]^2 P\left(Y = y_i \mid X = x\right)$$

- Example:

$Var(M \mid A = 1) = (0 - .14)^2 \cdot .9 + (1 - .14)^2 \cdot .07 + (2 - .14)^2 \cdot .02 + (3 - .14)^2 \cdot .01 + (4 - .14)^2 \cdot 0 = .22$

**B. Conditional Distributions of *M* given *A***

| | $M = 0$ | $M = 1$ | $M = 2$ | $M = 3$ | $M = 4$ | **Total** |
|---|---|---|---|---|---|---|
| $\Pr(M \mid A = 0)$ | 0.70 | 0.13 | 0.10 | 0.05 | 0.02 | 1.00 |
| $\Pr(M \mid A = 1)$ | 0.90 | 0.07 | 0.02 | 0.01 | 0.00 | 1.00 |

# The Law of Iterated Expectations (L.I.E.)

- One quite powerful result from probability theory is called the **Law of Iterated Expectations** (LIE)

$$E\left[E[(Y \mid X)\right] = E\left[Y\right]$$

- The LIE states that the expected value of $Y$ is the weighted average of the conditional expectation of $Y$ given $X$, weighted by the probability distribution of $X$.
- For example, the mean income of US citizens is the weighted average of the mean income of men and the mean income of women, weighted by the proportion of men and women.
- The LIE is very useful since if we know that $E\left(Y \mid X\right) = 0$ then we can conclude that $E\left(Y\right) = 0$.
- Let's prove the LIE.

# Independence

- $X$ and $Y$ are said to be **independent** if for all $x$ and $y$

$$P(Y = y \mid X = x) = P(Y = y)$$

  - "Knowing $X$ tells you nothing about $Y$"

# Independence

- $X$ and $Y$ are said to be **independent** if for all $x$ and $y$

$$P(Y = y \mid X = x) = P(Y = y)$$

  - "Knowing $X$ tells you nothing about $Y$"
- Here's an alternative (and equivalent) definition: if $X$ and $Y$ are independent, then

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

  - "The joint distribution is the product of the marginals"

# Independence

How do I know the joint distribution is the product of the marginals?

Well, under independence,

$$P(Y = y \mid X = x) = P(Y = y)$$

# Independence

How do I know the joint distribution is the product of the marginals?

Well, under independence,

$$P(Y = y \mid X = x) = P(Y = y)$$

and

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

so

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

which yields

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

# Are $X$ and $Y$ independent in this example?

**TABLE 2.2** Joint Distribution of Weather Conditions and Commuting Times

| | 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 |

Let's check for $X = 0, Y = 0$ :

# Are $X$ and $Y$ independent in this example?

**TABLE 2.2** Joint Distribution of Weather Conditions and Commuting Times

| | 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 |

Let's check for $X = 0, Y = 0$ :

$$P(Y = 0 \mid X = 0) = \frac{P(X=0, Y=0)}{P(X=0)} = \frac{.15}{.30} = .5 \neq P(Y = 0) = .22$$

# Are $X$ and $Y$ independent in this example?

**TABLE 2.2** Joint Distribution of Weather Conditions and Commuting Times

| | 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 |

Let's check for $X = 0, Y = 0$ :

$P(Y = 0 \mid X = 0) = \frac{P(X=0,Y=0)}{P(X=0)} = \frac{.15}{.30} = .5 \neq P(Y = 0) = .22$

This is what the joint distribution would look like under independence.

| | X=0 | X=1 | Total |
|---|---|---|---|
| Y=0 | 0.066 | 0.154 | 0.22 |
| Y=1 | 0.234 | 0.546 | 0.78 |
| Total | 0.3 | 0.7 | 1 |

Now, for $X = 0, Y = 0$ :

$P(Y = 0 \mid X = 0) = \frac{P(X=0,Y=0)}{P(X=0)} = \frac{.066}{.30} = .22 = P(Y = 0)$

# Measuring the Strength of the Relation

- **Variance** of $X$: measures how $X$ alone varies.

$$\begin{aligned}
\sigma_X^2 &= E[X - \mu_X]^2 \\
\sigma_X^2 &= \sum_{i=1}^{n} (x_i - \mu_X)^2 P(X = x_i) \\
\sigma_X^2 &= \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x) dx \\
\sigma_X^2 &= E(X^2) - [E(X)]^2
\end{aligned}$$

# Measuring the Strength of the Relation

- **Variance** of $X$: measures how $X$ alone varies.

$$\begin{aligned}
\sigma_X^2 &= E[X - \mu_X]^2 \\
\sigma_X^2 &= \textstyle\sum_{i=1}^{n} (x_i - \mu_X)^2 P(X = x_i) \\
\sigma_X^2 &= \textstyle\int_{-\infty}^{\infty} (x - \mu_X)^2 f(x) dx \\
\sigma_X^2 &= E(X^2) - [E(X)]^2
\end{aligned}$$

- **Covariance** of $X$ and $Y$: measures how $X$ and $Y$ vary together.

$$\begin{aligned}
cov(X, Y) &= E\left[(X - \mu_X)(Y - \mu_Y)\right] \\
cov(X, Y) &= \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i - \mu_X)(y_j - \mu_Y) P(X = x_i, Y = y_j) \\
cov(X, Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y) dx dy \\
cov(X, Y) &= E(XY) - E(X)E(Y)
\end{aligned}$$

# Correlation

- $cov(X, Y)$ is in units of $X \times Y$, which can be tricky to interpret.
- The **correlation** of $X$ and $Y$ is the covariance rescaled to be between $-1$ and $1$ and unit free:

$$\rho = \frac{cov(X, Y)}{\sigma_X \sigma_Y}$$

# Correlation

- $cov(X, Y)$ is in units of $X \times Y$, which can be tricky to interpret.
- The **correlation** of $X$ and $Y$ is the covariance rescaled to be between $-1$ and $1$ and unit free:

$$\rho = \frac{cov(X, Y)}{\sigma_X \sigma_Y}$$

- Note that $-1 \leq \rho \leq 1$
- $\rho = 1$ : perfect positive relation
- $\rho = -1$ : perfect negative relation
- The size in absolute value $|\rho|$, reflects the strength of the **linear** association between the two variables

# Correlation



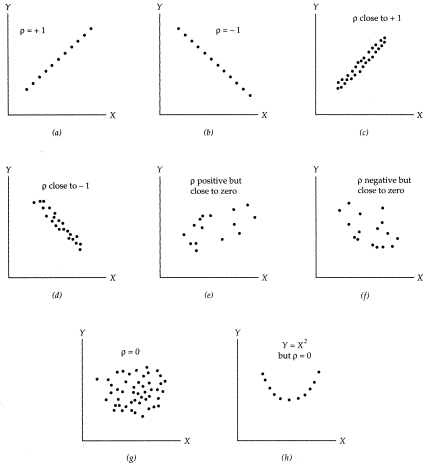
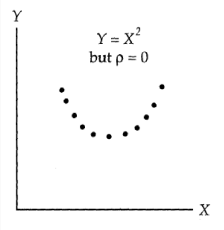



**FIGURE 2-7**
Some typical patterns of the correlation coefficient, ρ.

# Independence vs. Uncorrelation




- Independent random variables are also uncorrelated, but the reverse is NOT true!
- **Intuition**: the definition of independence of $X$ and $Y$ refers to *all possible values* of $X$ and $Y$, while uncorrelation simply refers to expected values!

# Example of two uncorrelated but not independent R.V.

| | | $Y$ | | |
|---|---|---|---|---|
| | | -1 | 0 | 1 |
| $X$ | -1 | 0 | 0.2 | 0 |
| | 0 | 0.3 | 0 | 0.3 |
| | 1 | 0 | 0.2 | 0 |

$X$ and $Y$ are clearly not independent, but $Cov(X, Y) = 0$ so $\rho = 0$.

Do you know how to show this?

Hint:

$cov(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i - \mu_X)(y_j - \mu_Y) P(X = x_i, Y = y_j)$

but note that $\mu_X = 0$ & $\mu_Y = 0$ so

$cov(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i)(y_j) P(X = x_i, Y = y_j)$

You should be able to argue that this is zero without performing any calculations...

# Covariance Continued

- If the conditional mean of $Y$ does not depend on $X$, then $Y$ and $X$ are uncorrelated

$$E(Y \mid X) = E(Y) \Longrightarrow cov(X, Y) = 0 \text{ and } corr(X, Y) = 0$$

# Covariance Continued

- If the conditional mean of $Y$ does not depend on $X$, then $Y$ and $X$ are uncorrelated

$$E(Y \mid X) = E(Y) \Longrightarrow cov(X, Y) = 0 \text{ and } corr(X, Y) = 0$$

- If $X_1, X_2, ..., X_k$ are independent – the probability distribution for one does not depend on the others, then

$$Var(X_1 + ... + X_n) = Var(X_1) + Var(X_2) + ... + Var(X_n).$$

- In general

$$Var(X_1 + ... + X_n) = \sum_{i=1}^{n} Var(X_i) + 2 \sum \sum_{i<j} Cov(X_i, X_j)$$

- Example

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

# Some Useful Results

**Means, Variances, and Covariances of Sums of Random Variables**

Let $X$, $Y$, and $V$ be random variables, let $\mu_X$ and $\sigma_X^2$ be the mean and variance of $X$, let $\sigma_{XY}$ be the covariance between $X$ and $Y$ (and so forth for the other variables), and let $a$, $b$, and $c$ be constants. The following facts follow from the definitions of the mean, variance, and covariance:

$$E(a + bX + cY) = a + b\mu_X + c\mu_Y, \tag{2.29}$$

$$\text{var}(a + bY) = b^2\sigma_Y^2, \tag{2.30}$$

$$\text{var}(aX + bY) = a^2\sigma_X^2 + 2ab\sigma_{XY} + b^2\sigma_Y^2, \tag{2.31}$$

$$E(Y^2) = \sigma_Y^2 + \mu_Y^2, \tag{2.32}$$

$$\text{cov}(a + bX + cV, Y) = b\sigma_{XY} + c\sigma_{VY}, \text{ and} \tag{2.33}$$

$$E(XY) = \sigma_{XY} + \mu_X\mu_Y. \tag{2.34}$$

$$|\text{corr}(X, Y)| \leq 1 \text{ and } |\sigma_{XY}| \leq \sqrt{\sigma_X^2\sigma_Y^2} \text{ (correlation inequality).} \tag{2.35}$$

# Some Useful Results

**Means, Variances, and Covariances of Sums of Random Variables**

Let $X$, $Y$, and $V$ be random variables, let $\mu_X$ and $\sigma_X^2$ be the mean and variance of $X$, let $\sigma_{XY}$ be the covariance between $X$ and $Y$ (and so forth for the other variables), and let $a$, $b$, and $c$ be constants. The following facts follow from the definitions of the mean, variance, and covariance:

$$E(a + bX + cY) = a + b\mu_X + c\mu_Y, \tag{2.29}$$

$$\text{var}(a + bY) = b^2\sigma_Y^2, \tag{2.30}$$

$$\text{var}(aX + bY) = a^2\sigma_X^2 + 2ab\sigma_{XY} + b^2\sigma_Y^2, \tag{2.31}$$

$$E(Y^2) = \sigma_Y^2 + \mu_Y^2, \tag{2.32}$$

$$\text{cov}(a + bX + cV, Y) = b\sigma_{XY} + c\sigma_{VY}, \text{ and} \tag{2.33}$$

$$E(XY) = \sigma_{XY} + \mu_X\mu_Y. \tag{2.34}$$

$$|\text{corr}(X, Y)| \leq 1 \text{ and } |\sigma_{XY}| \leq \sqrt{\sigma_X^2\sigma_Y^2} \text{ (correlation inequality).} \tag{2.35}$$

Q: What is $Cov\,(X, X)$?

# Some Useful Results

**Means, Variances, and Covariances of Sums of Random Variables**

Let $X$, $Y$, and $V$ be random variables, let $\mu_X$ and $\sigma_X^2$ be the mean and variance of $X$, let $\sigma_{XY}$ be the covariance between $X$ and $Y$ (and so forth for the other variables), and let $a$, $b$, and $c$ be constants. The following facts follow from the definitions of the mean, variance, and covariance:

$$E(a + bX + cY) = a + b\mu_X + c\mu_Y, \tag{2.29}$$

$$\text{var}(a + bY) = b^2\sigma_Y^2, \tag{2.30}$$

$$\text{var}(aX + bY) = a^2\sigma_X^2 + 2ab\sigma_{XY} + b^2\sigma_Y^2, \tag{2.31}$$

$$E(Y^2) = \sigma_Y^2 + \mu_Y^2, \tag{2.32}$$

$$\text{cov}(a + bX + cV, Y) = b\sigma_{XY} + c\sigma_{VY}, \text{ and} \tag{2.33}$$

$$E(XY) = \sigma_{XY} + \mu_X\mu_Y. \tag{2.34}$$

$$|\text{corr}(X,Y)| \le 1 \text{ and } |\sigma_{XY}| \le \sqrt{\sigma_X^2\sigma_Y^2} \text{ (correlation inequality).} \tag{2.35}$$

Q: What is $Cov\,(X, X)$?

Q: (one of the MOST common mistakes) $Var\,(X - Y) = ?????$

# Some Useful Results

**Means, Variances, and Covariances of Sums of Random Variables**

Let $X$, $Y$, and $V$ be random variables, let $\mu_X$ and $\sigma_X^2$ be the mean and variance of $X$, let $\sigma_{XY}$ be the covariance between $X$ and $Y$ (and so forth for the other variables), and let $a$, $b$, and $c$ be constants. The following facts follow from the definitions of the mean, variance, and covariance:

$$E(a + bX + cY) = a + b\mu_X + c\mu_Y, \tag{2.29}$$

$$\text{var}(a + bY) = b^2\sigma_Y^2, \tag{2.30}$$

$$\text{var}(aX + bY) = a^2\sigma_X^2 + 2ab\sigma_{XY} + b^2\sigma_Y^2, \tag{2.31}$$

$$E(Y^2) = \sigma_Y^2 + \mu_Y^2, \tag{2.32}$$

$$\text{cov}(a + bX + cV, Y) = b\sigma_{XY} + c\sigma_{VY}, \text{ and} \tag{2.33}$$

$$E(XY) = \sigma_{XY} + \mu_X\mu_Y. \tag{2.34}$$

$$|\text{corr}(X,Y)| \le 1 \text{ and } |\sigma_{XY}| \le \sqrt{\sigma_X^2\sigma_Y^2} \text{ (correlation inequality).} \tag{2.35}$$

Q: What is $Cov\,(X, X)$?

Q: (one of the MOST common mistakes) $Var\,(X - Y) = ?????$

$$Var(X - Y) = 1^2 Var(X) + 2(1)(-1)Cov(X, Y) + 1^2 Var(Y) = Var(X) + Var(Y) - 2Cov(X, Y)$$

Does this make intuitive sense to you?

# Some Useful Distributions

## Bernoulli Distribution

- A Bernoulli random variable takes on the value 1 with probability $p$ & 0 with probability $(1-p)$.

$$\begin{aligned} P(X &= x) = p^x(1-p)^{1-x} \text{ for } X = 0, 1 \\ P(X &= x) = 0 \text{ otherwise} \end{aligned}$$

- $P(X = 1) = p, P(X = 0) = (1-p)$
- $E(X) = \sum_{i=1}^{n} x_i \Pr(x_i) = 1 \cdot p + 0 \cdot (1-p) = p$
- $E(X^2) = \sum_{i=1}^{n} x_i^2 \Pr(x_i) = 1^2 \cdot p + 0^2 \cdot (1-p) = p$
- $Var(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1-p)$

# Binomial Distribution

The sum of $n$ independent Bernoulli random variables has a Binomial distribution. For $Y = 0, 1, 2, ..., n$

$$P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}$$

- $\binom{n}{y} = \frac{n!}{y!(n-y)!}$ (this is called a Binomial coefficient)
- $n! = n \cdot (n-1) \cdot (n-2) \cdot ... \cdot 1$ (note: $0! = 1$)
- Example 1: $n = 2, p = \frac{1}{2}$

$P(Y = 0) = \frac{2!}{0!2!}(\frac{1}{2})^0(\frac{1}{2})^2 = 1 \cdot 1 \cdot (\frac{1}{2})^2 = \frac{1}{4}$

$P(Y = 1) = \frac{2!}{1!1!}(\frac{1}{2})^1(\frac{1}{2})^1 = 2 \cdot (\frac{1}{2}) \cdot (\frac{1}{2}) = \frac{1}{2}$

$P(Y = 2) = \frac{2!}{2!0!}(\frac{1}{2})^2(\frac{1}{2})^0 = 1 \cdot (\frac{1}{2})^2 \cdot 1 = \frac{1}{4}$

# Binomial Distribution

$$P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}$$

- Example 2: $n = 10, p = \frac{1}{2}$

$P(Y = 0) = \frac{10!}{0!10!}(\frac{1}{2})^0(\frac{1}{2})^{10} \approx .001$

$P(Y = 5) = \frac{10!}{5!5!}(\frac{1}{2})^5(\frac{1}{2})^5 \approx .25$

## Mean and Variance

- Given that $Y = X_1 + X_2 + ... + X_n$, where the $X_i$ are all independent Bernoulli trials:
- $E(Y) = E(X_1) + E(X_2) + ... + E(X_n) = np$
- $Var(Y) = Var(X_1) + Var(X_2) + ... + Var(X_n) = np(1-p)$

# Some Useful Distributions

## Uniform Distribution

- Uniform distribution over $[a, b]$ : $f(x) = \frac{1}{b-a}$
- Example: there is a bus every 10 minutes. You arrive at the stop at a random time. What is the probability that the bus will appear within $X$ minutes?

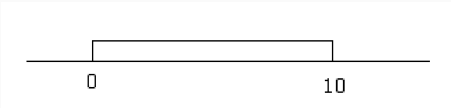
Mean of a uniform distribution

$$E(X) = \int_a^b \frac{1}{b-a} x dx = \frac{1}{b-a}[\frac{1}{2}x^2]_a^b = \frac{1}{2(b-a)}(b^2 - a^2) = \frac{a+b}{2}$$

Example: $U[0, 10]$

$f(x) = \frac{1}{10}$

$E(X) = 5$



Variance of a uniform distribution:

$$Var(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 f(x)dx = \int_a^b \left[x - \frac{a+b}{2}\right]^2 \frac{1}{b-a} dx = \left[\frac{1}{3(b-a)} \left[x - \frac{a+b}{2}\right]^3\right]_a^b =$$

$$\left[\frac{1}{3(b-a)} \left[b - \frac{a+b}{2}\right]^3 - \frac{1}{3(b-a)} \left[a - \frac{a+b}{2}\right]^3\right]$$

$$= \left[\frac{1}{3(b-a)} \left[\frac{b-a}{2}\right]^3 - \frac{1}{3(b-a)} \left[\frac{a-b}{2}\right]^3\right]$$

$$= \left[\frac{1}{3(b-a)} \left[\frac{b-a}{2}\right]^3 + \frac{1}{3(b-a)} \left[\frac{b-a}{2}\right]^3\right]$$

$$= \frac{2}{3(b-a)} \left[\frac{b-a}{2}\right]^3 = \frac{(b-a)^2}{12}$$

Variance of a uniform distribution:

$$Var(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 f(x)dx = \int_a^b \left[x - \frac{a+b}{2}\right]^2 \frac{1}{b-a} dx = \left[\frac{1}{3(b-a)} \left[x - \frac{a+b}{2}\right]^3\right]_a^b =$$

$$\left[\frac{1}{3(b-a)} \left[b - \frac{a+b}{2}\right]^3 - \frac{1}{3(b-a)} \left[a - \frac{a+b}{2}\right]^3\right]$$

$$= \left[\frac{1}{3(b-a)} \left[\frac{b-a}{2}\right]^3 - \frac{1}{3(b-a)} \left[\frac{a-b}{2}\right]^3\right]$$

$$= \left[\frac{1}{3(b-a)} \left[\frac{b-a}{2}\right]^3 + \frac{1}{3(b-a)} \left[\frac{b-a}{2}\right]^3\right]$$

$$= \frac{2}{3(b-a)} \left[\frac{b-a}{2}\right]^3 = \frac{(b-a)^2}{12}$$

For $a = 0, b = 10 : Var(X) = \frac{(10)^2}{12} = 8.33, SD(X) = 2.89$

Variance of a uniform distribution:

$$Var(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 f(x)dx = \int_{a}^{b} \left[x - \frac{a+b}{2}\right]^2 \frac{1}{b-a} dx = \left[\frac{1}{3(b-a)} \left[x - \frac{a+b}{2}\right]^3\right]_a^b =$$

$$\left[\frac{1}{3(b-a)} \left[b - \frac{a+b}{2}\right]^3 - \frac{1}{3(b-a)} \left[a - \frac{a+b}{2}\right]^3\right]$$

$$= \left[\frac{1}{3(b-a)} \left[\frac{b-a}{2}\right]^3 - \frac{1}{3(b-a)} \left[\frac{a-b}{2}\right]^3\right]$$

$$= \left[\frac{1}{3(b-a)} \left[\frac{b-a}{2}\right]^3 + \frac{1}{3(b-a)} \left[\frac{b-a}{2}\right]^3\right]$$

$$= \frac{2}{3(b-a)} \left[\frac{b-a}{2}\right]^3 = \frac{(b-a)^2}{12}$$

For $a = 0, b = 10 : Var(X) = \frac{(10)^2}{12} = 8.33, SD(X) = 2.89$

What about the CDF?

$$F(x) = \int_{-\infty}^{x} \frac{1}{b-a} du = \frac{1}{b-a} u \mid_a^x = \frac{1}{b-a}(x-a)$$

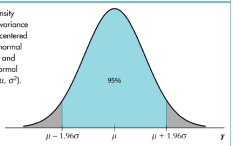
For $a = 0, b = 10,$ then $F(5) = \frac{1}{10}(5-0) = \frac{1}{2}$

# Normal Distribution

- Lots of things turn out to be approximately normal.
- Hence, the normal distribution is the most important and most frequently used probability density function in econometrics.
- Many estimators have a normal distribution for $n \to \infty$
- The normal distribution:
  - is completely defined by its mean and standard deviation (2 parameters);
  - is single peaked;
  - is symmetric about the mean;
  - can be standardized to have mean 0 and variance 1

**FIGURE 2.3 The Normal Probability Density**

The normal probability density function with mean $\mu$ and variance $\sigma^2$ is a bell-shaped curve, centered at $\mu$. The area under the normal p.d.f. between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$ is 0.95. The normal distribution is denoted $N(\mu, \sigma^2)$.



# Normal Distribution

- Letting $Y$ represent a r.v. that has a normal distribution with mean $\mu$ and standard deviation $\sigma$, the density function for $Y$ is

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2} \qquad -\infty < y < \infty$$

- The CDF is given by

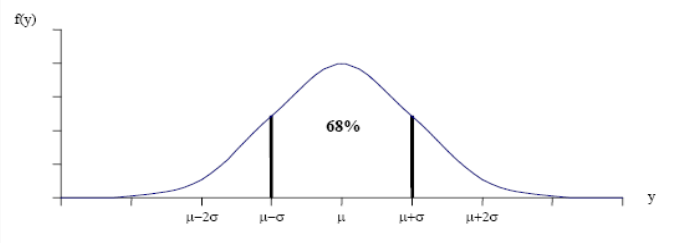
$$F(y) = \int_{-\infty}^{y} f(t)dt$$

- This integral can't be evaluated in closed form, that's why you need a table of values (or a numerical approximation).

# Normal Distribution

- Rules of thumb
  - The probability within a given number of standard deviations from the mean is constant for all normal distributions.
  - This allows us to construct some convenient rules of thumb.
  - Approximately 68% of the observations lie within one standard deviation of the mean:

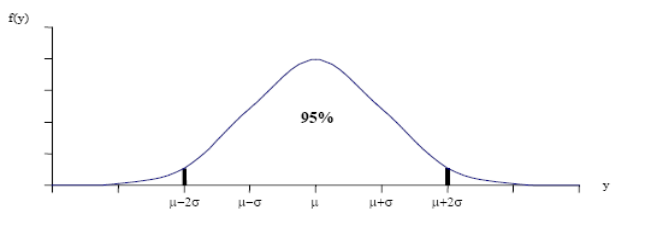
$$\Pr(\mu - \sigma \leq Y \leq \mu + \sigma) \cong 0.68.$$



# Normal Distribution

- Approximately 95% of the observations lie within two (really 1.96) standard deviations of the mean:

$$\Pr(\mu - 2\sigma \leq Y \leq \mu + 2\sigma) \cong 0.95.$$



# Standard Normal Distribution

- Recall the normal density and CDF

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2} \quad \text{and} \quad F(y) = \int_{-\infty}^{y} f(t)dt$$

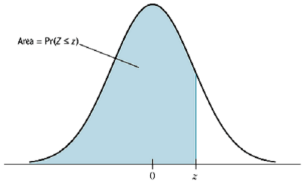
- A normal distribution with mean 0 and standard deviation 1 is called a standard normal distribution.
- A standard normal random variable is typically denoted by $Z$, its pdf by $\phi(z)$, and its df by $\Phi(z)$ where:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad \text{and} \quad \Phi(z) = \int_{-\infty}^{z} \phi(t)\, dt$$

# Standard Normal Distribution

- Any normal r.v. $Y \sim N(\mu, \sigma^2)$ may be transformed to the standard normal r.v. $Z \sim N(0, 1)$ by subtracting the mean and dividing by the standard deviation.
- If $Y$ is normal with mean $\mu$ and standard deviation $\sigma$, then $Z = \frac{Y-\mu}{\sigma}$ is normal with mean 0 and standard deviation 1.
- Therefore, you only need one set of tables to handle all normally distributed variables!

**TABLE 1 The Cumulative Standard Normal Distribution Function, $\Phi(z) = \Pr(Z \leq z)$**




| | Second Decimal Value of z | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| z | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| −2.9 | 0.0019 | 0.0018 | 0.0018 | 0.0017 | 0.0016 | 0.0016 | 0.0015 | 0.0015 | 0.0014 | 0.0014 |
| −2.8 | 0.0026 | 0.0025 | 0.0024 | 0.0023 | 0.0023 | 0.0022 | 0.0021 | 0.0021 | 0.0020 | 0.0019 |
| −2.7 | 0.0035 | 0.0034 | 0.0033 | 0.0032 | 0.0031 | 0.0030 | 0.0029 | 0.0028 | 0.0027 | 0.0026 |
| −2.6 | 0.0047 | 0.0045 | 0.0044 | 0.0043 | 0.0041 | 0.0040 | 0.0039 | 0.0038 | 0.0037 | 0.0036 |
| −2.5 | 0.0062 | 0.0060 | 0.0059 | 0.0057 | 0.0055 | 0.0054 | 0.0052 | 0.0051 | 0.0049 | 0.0048 |
| −2.4 | 0.0082 | 0.0080 | 0.0078 | 0.0075 | 0.0073 | 0.0071 | 0.0069 | 0.0068 | 0.0066 | 0.0064 |
| −2.3 | 0.0107 | 0.0104 | 0.0102 | 0.0099 | 0.0096 | 0.0094 | 0.0091 | 0.0089 | 0.0087 | 0.0084 |
| −2.2 | 0.0139 | 0.0136 | 0.0132 | 0.0129 | 0.0125 | 0.0122 | 0.0119 | 0.0116 | 0.0113 | 0.0110 |
| −2.1 | 0.0179 | 0.0174 | 0.0170 | 0.0166 | 0.0162 | 0.0158 | 0.0154 | 0.0150 | 0.0146 | 0.0143 |
| −2.0 | 0.0228 | 0.0222 | 0.0217 | 0.0212 | 0.0207 | 0.0202 | 0.0197 | 0.0192 | 0.0188 | 0.0183 |
| −1.9 | 0.0287 | 0.0281 | 0.0274 | 0.0268 | 0.0262 | 0.0256 | 0.0250 | 0.0244 | 0.0239 | 0.0233 |
| −1.8 | 0.0359 | 0.0351 | 0.0344 | 0.0336 | 0.0329 | 0.0322 | 0.0314 | 0.0307 | 0.0301 | 0.0294 |
| −1.7 | 0.0446 | 0.0436 | 0.0427 | 0.0418 | 0.0409 | 0.0401 | 0.0392 | 0.0384 | 0.0375 | 0.0367 |
| −1.6 | 0.0548 | 0.0537 | 0.0526 | 0.0516 | 0.0505 | 0.0495 | 0.0485 | 0.0475 | 0.0465 | 0.0455 |
| −1.5 | 0.0668 | 0.0655 | 0.0643 | 0.0630 | 0.0618 | 0.0606 | 0.0594 | 0.0582 | 0.0571 | 0.0559 |
| −1.4 | 0.0808 | 0.0793 | 0.0778 | 0.0764 | 0.0749 | 0.0735 | 0.0721 | 0.0708 | 0.0694 | 0.0681 |
| −1.3 | 0.0968 | 0.0951 | 0.0934 | 0.0918 | 0.0901 | 0.0885 | 0.0869 | 0.0853 | 0.0838 | 0.0823 |
| −1.2 | 0.1151 | 0.1131 | 0.1112 | 0.1093 | 0.1075 | 0.1056 | 0.1038 | 0.1020 | 0.1003 | 0.0985 |
| −1.1 | 0.1357 | 0.1335 | 0.1314 | 0.1292 | 0.1271 | 0.1251 | 0.1230 | 0.1210 | 0.1190 | 0.1170 |
| −1.0 | 0.1587 | 0.1562 | 0.1539 | 0.1515 | 0.1492 | 0.1469 | 0.1446 | 0.1423 | 0.1401 | 0.1379 |
| −0.9 | 0.1841 | 0.1814 | 0.1788 | 0.1762 | 0.1736 | 0.1711 | 0.1685 | 0.1660 | 0.1635 | 0.1611 |

**TABLE 1** *(continued)*

| | Second Decimal Value of $z$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $z$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| −0.8 | 0.2119 | 0.2090 | 0.2061 | 0.2033 | 0.2005 | 0.1977 | 0.1949 | 0.1922 | 0.1894 | 0.1867 |
| −0.7 | 0.2420 | 0.2389 | 0.2358 | 0.2327 | 0.2296 | 0.2266 | 0.2236 | 0.2206 | 0.2177 | 0.2148 |
| −0.6 | 0.2743 | 0.2709 | 0.2676 | 0.2643 | 0.2611 | 0.2578 | 0.2546 | 0.2514 | 0.2483 | 0.2451 |
| −0.5 | 0.3085 | 0.3050 | 0.3015 | 0.2981 | 0.2946 | 0.2912 | 0.2877 | 0.2843 | 0.2810 | 0.2776 |
| −0.4 | 0.3446 | 0.3409 | 0.3372 | 0.3336 | 0.3300 | 0.3264 | 0.3228 | 0.3192 | 0.3156 | 0.3121 |
| −0.3 | 0.3821 | 0.3783 | 0.3745 | 0.3707 | 0.3669 | 0.3632 | 0.3594 | 0.3557 | 0.3520 | 0.3483 |
| −0.2 | 0.4207 | 0.4168 | 0.4129 | 0.4090 | 0.4052 | 0.4013 | 0.3974 | 0.3936 | 0.3897 | 0.3859 |
| −0.1 | 0.4602 | 0.4562 | 0.4522 | 0.4483 | 0.4443 | 0.4404 | 0.4364 | 0.4325 | 0.4286 | 0.4247 |
| −0.0 | 0.5000 | 0.4960 | 0.4920 | 0.4880 | 0.4840 | 0.4801 | 0.4761 | 0.4721 | 0.4681 | 0.4641 |
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9699 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 | 0.9767 |
| 2.0 | 0.9772 | 0.9778 | 0.9783 | 0.9788 | 0.9793 | 0.9798 | 0.9803 | 0.9808 | 0.9812 | 0.9817 |
| 2.1 | 0.9821 | 0.9826 | 0.9830 | 0.9834 | 0.9838 | 0.9842 | 0.9846 | 0.9850 | 0.9854 | 0.9857 |
| 2.2 | 0.9861 | 0.9864 | 0.9868 | 0.9871 | 0.9875 | 0.9878 | 0.9881 | 0.9884 | 0.9887 | 0.9890 |
| 2.3 | 0.9893 | 0.9896 | 0.9898 | 0.9901 | 0.9904 | 0.9906 | 0.9909 | 0.9911 | 0.9913 | 0.9916 |
| 2.4 | 0.9918 | 0.9920 | 0.9922 | 0.9925 | 0.9927 | 0.9929 | 0.9931 | 0.9932 | 0.9934 | 0.9936 |
| 2.5 | 0.9938 | 0.9940 | 0.9941 | 0.9943 | 0.9945 | 0.9946 | 0.9948 | 0.9949 | 0.9951 | 0.9952 |
| 2.6 | 0.9953 | 0.9955 | 0.9956 | 0.9957 | 0.9959 | 0.9960 | 0.9961 | 0.9962 | 0.9963 | 0.9964 |
| 2.7 | 0.9965 | 0.9966 | 0.9967 | 0.9968 | 0.9969 | 0.9970 | 0.9971 | 0.9972 | 0.9973 | 0.9974 |
| 2.8 | 0.9974 | 0.9975 | 0.9976 | 0.9977 | 0.9977 | 0.9978 | 0.9979 | 0.9979 | 0.9980 | 0.9981 |
| 2.9 | 0.9981 | 0.9982 | 0.9982 | 0.9983 | 0.9984 | 0.9984 | 0.9985 | 0.9985 | 0.9986 | 0.9986 |

This table can be used to calculate $\Pr(Z \leq z)$ where $Z$ is a standard normal variable. For example, when $z = 1.17$, this probability is 0.8790, which is the table entry for the row labeled 1.1 and the column labeled 7.

You can use a standard normal table to find probabilities for any normal random variable:

$$\Pr(Y \leq a \mid \mu, \sigma) = \Pr\left(Z \leq \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right).$$

- Example: $\mu = 4.20$ years, $\sigma = 0.6$ years
- What is $\Pr($light will last between 3 and 5 years$)$?

$\Pr(3 < Y < 5) = \Pr\left(\frac{3-4.2}{.6} < Z < \frac{5-4.2}{.6}\right) = \Pr(-2 < Z < 1.33)$

$= \Pr(Z < 1.33) - \Pr(Z < -2) = \Phi(1.33) - \Phi(-2)$

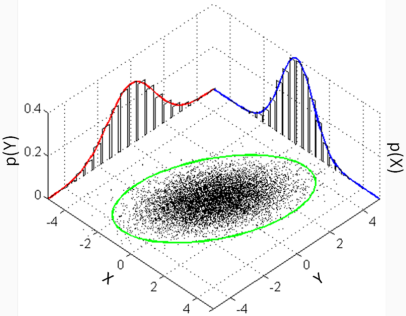
$= .9082 - .0228 = .8854$

- How about $\Pr($light will last more than 3 years$)$?

$\Pr(3 < Y) = \Pr\left(\frac{3-4.2}{.6} < Z\right) = \Pr(-2 < Z) =$

$1 - \Pr(Z < -2) = 1 - .0028 = .9772$

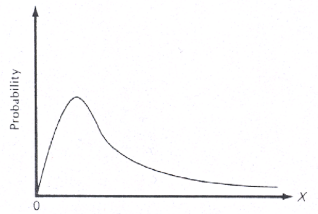
# Multivariate Normal




- The normal distribution can be generalized to describe the joint distribution of several random variables.
- Some useful properties
  - If $n$ variables have a multivariate normal distribution, any linear combination of them is also normally distributed.
  - The marginals of a multi-variate normal are also normal.
  - 0 covariance implies independence.

# Chi-Squared, $F$ and $t$ Distributions

- A lot of other useful distributions can be constructed from normals (we'll see why later)
- $\chi^2_m$ is the sum of $m$ squared standard normals ($m$ is called the degrees of freedom)
- $F_{n,m} = \frac{\frac{\chi^2_n}{n}}{\frac{\chi^2_m}{m}}$ (where the $\chi^2$'s are independent)
  - Both are always positive (why?)
- $F_{m,\infty} = \frac{\chi^2_m}{m}$ (the average of $m$ squared normals) will be important for asymptotic results



# Student's $t$ distribution

- Student's $t$ distribution (when we know $z$ is standard normal and independent from $\chi^2_m$)

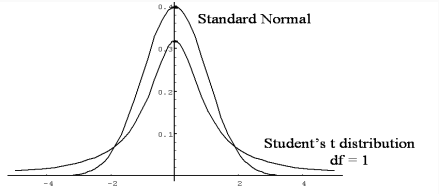
$$t_m = \frac{z}{\sqrt{\frac{\chi^2_m}{m}}} = \frac{z}{\sqrt{F_{m,\infty}}}$$

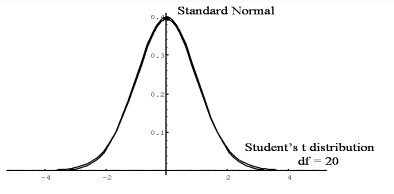
# Student's $t$ distribution

- Student's $t$ distribution (when we know $z$ is standard normal and independent from $\chi^2_m$)

$$t_m = \frac{z}{\sqrt{\frac{\chi^2_m}{m}}} = \frac{z}{\sqrt{F_{m,\infty}}}$$

- When $m \to \infty$ then $t \xrightarrow{d} N(0,1)$

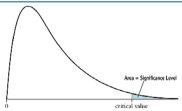


- In the back of the textbook, there are tables that list the "critical values" of the $\chi^2, F$ and $t$ distributions
  - For example, you can find that for the $F_{8,\infty}$ distribution

$$
\begin{aligned}
P(Y &< 1.67) = .9 \\
P(Y &< 1.94) = .95 \\
P(Y &< 2.51) = .99
\end{aligned}
$$

**TABLE 4 Critical Values for the $F_{8,\infty}$ Distribution**




| Degrees of Freedom | Significance Level 10% | 5% | 1% |
|---|---|---|---|
| 1 | 2.71 | 3.84 | 6.63 |
| 2 | 2.30 | 3.00 | 4.61 |
| 3 | 2.08 | 2.60 | 3.78 |
| 4 | 1.94 | 2.37 | 3.32 |
| 5 | 1.85 | 2.21 | 3.02 |
| 6 | 1.77 | 2.10 | 2.80 |
| 7 | 1.72 | 2.01 | 2.64 |
| 8 | 1.67 | 1.94 | 2.51 |
| 9 | 1.63 | 1.88 | 2.41 |
| 10 | 1.60 | 1.83 | 2.32 |
| 11 | 1.57 | 1.79 | 2.25 |
| 12 | 1.55 | 1.75 | 2.18 |
| 13 | 1.52 | 1.72 | 2.13 |
| 14 | 1.50 | 1.69 | 2.08 |
| 15 | 1.49 | 1.67 | 2.04 |
| 16 | 1.47 | 1.64 | 2.00 |
| 17 | 1.46 | 1.62 | 1.97 |
| 18 | 1.44 | 1.60 | 1.93 |
| 19 | 1.43 | 1.59 | 1.90 |
| 20 | 1.42 | 1.57 | 1.88 |
| 21 | 1.41 | 1.56 | 1.85 |
| 22 | 1.40 | 1.54 | 1.83 |
| 23 | 1.39 | 1.53 | 1.81 |
| 24 | 1.38 | 1.52 | 1.79 |
| 25 | 1.38 | 1.51 | 1.77 |
| 26 | 1.37 | 1.50 | 1.76 |
| 27 | 1.36 | 1.49 | 1.74 |
| 28 | 1.35 | 1.48 | 1.72 |
| 29 | 1.35 | 1.47 | 1.71 |
| 30 | 1.34 | 1.46 | 1.70 |

This table contains the 90th, 95th, and 99th percentiles of the $F_{8,\infty}$ distribution. These serve as critical values for tests with significance levels of 10%, 5%, and 1%.