Maximum Likelihood
Week 1

POLI 6003

Rich Frank

September 6, 2012

The Linear Regression Model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

OLS Assumptions

1. **<u>Linear</u>**

-In the parameters:

--Change in $\hat{Y}$ is always β, not $f(\beta)$

-In the variables:

--Change in $\hat{Y}$ is always a function of x, not $f(x)$

OLS Assumptions

2. <u>Unbiased</u>

-The average, or expected, value of β is equal to the true value.

$$E(\hat{\beta}) = \beta$$

OLS Assumptions

3. <u>Efficient</u>

-If there are two unbiased estimators of β ($\hat{\beta}_1$ and $\hat{\beta}_2$), and the variance of $\hat{\beta}_1$ is smaller than or equal to $\hat{\beta}_2$ then $\hat{\beta}_1$ is the most efficient estimator.

OLS Assumptions

4. <u>Consistent</u>

As the sample size moves to infinity, the variance of the least square estimators approaches zero.

$$\text{As } N \rightarrow \infty, \quad \sigma^2 \rightarrow 0$$

OLS Assumptions

5. <u>Random Sampling</u>

-You are not sampling from a biased sample (e.g. college students).

OLS Assumptions

6. <u>Zero-conditional mean of ε</u>

-In your sample the positive and negative errors balance each other out.

$$E(\varepsilon_i \mid \mathbf{x}_i) = 0$$

OLS Assumptions

7. <u>No perfect collinearity</u>

-the x's are linearly independent.

OLS Assumptions

8. <u>Homoskedastic</u> disturbances

-For a given x, the errors have a constant variance.

$$\mathrm{Var}\,(\varepsilon_i \,|\, \mathrm{x}_i) = \sigma^2 \text{ for all } i$$

OLS Assumptions

9. <u>Uncorrelated</u> disturbances

-The independent variables are not correlated.

--A strong assumption, especially in CSTS data.

If these assumptions are met then OLS is the <u>best linear unbiased estimator</u> (BLUE).

Estimating β

$$\hat{\beta} = (\mathbf{X'X})^{-1}\,\mathbf{X'y}$$

We will not be trying to do this by hand today.

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$$

Time Series

1. Different data-sets have different characteristics you need to take into account.

2. A detailed analysis of time series is outside this class's scope.

3. There are many excellent and in-depth treatments out there.

4. I do want to mention several points about data structure and analysis.

Time Series

1. Cross-sectional

2. Cross-temporal

3. Shallow and wide

4. Narrow and long

Variable distributions

Variable distributions

- Social science rarely has experimental data, therefore we rely on historical data to test hypotheses.

- We have some uncertainty as to how the data were generated.

- We have to be clear as to what extent we think the sample represents the population as well as our uncertainty about the data generation process.

Variable distributions

- We often rely on the rules of probability to make statements about our uncertainty.

- Different outcomes have different probabilities associated with them.

  – Dying
  – Winning the lottery

- The complete list of possible outcomes and their associated probabilities is called a **probability distribution**.

- The probability that an event X happens $P(X)$ is the proportion of times X will take place in n trials of an experiment.

1. $P(X)$ is bounded by zero and one.

$$0 \leq P(X) \leq 1$$

2. The sum of probabilities for an exhaustive set of mutually exclusive events, $X, Y, Z\ldots$, is equals to one.

$$P(X) + P(Y) + \ldots + P(k) = 1$$

3. If $X, Y, Z \ldots$ are mutually exclusive events, then $P(X + Y + \ldots + k) = P(X) + P(Y) + \ldots + P(k)$.

- These events thus have two qualities:
  - They are mutually **exclusive**.
  - They are **exhaustive**.

Types of Outcomes

Dependent variables are some part of the social system that we are trying to explain.

Random variables

1. Discrete

2. Continuous

Discrete random variables

- A variable with a finite (or countably infinite) number of integer values
  – Not indefinitely divisible.
- Examples:
  – Number of wars
  – Outcome of coin toss
  – Roll of two dice
  – Home runs in a season

Continuous Random Variables

- Infinite number of values, each can have a zero probability of occurring.

- Examples:
  - Income
  - Weight
  - Temperature
  - Percentage of a vote

- These random variables are characterized by probability distributions.

- They can be summarized either by the Probability Density Function (PDF) or by the Cumulative Density Function (CDF).

PDF Definition

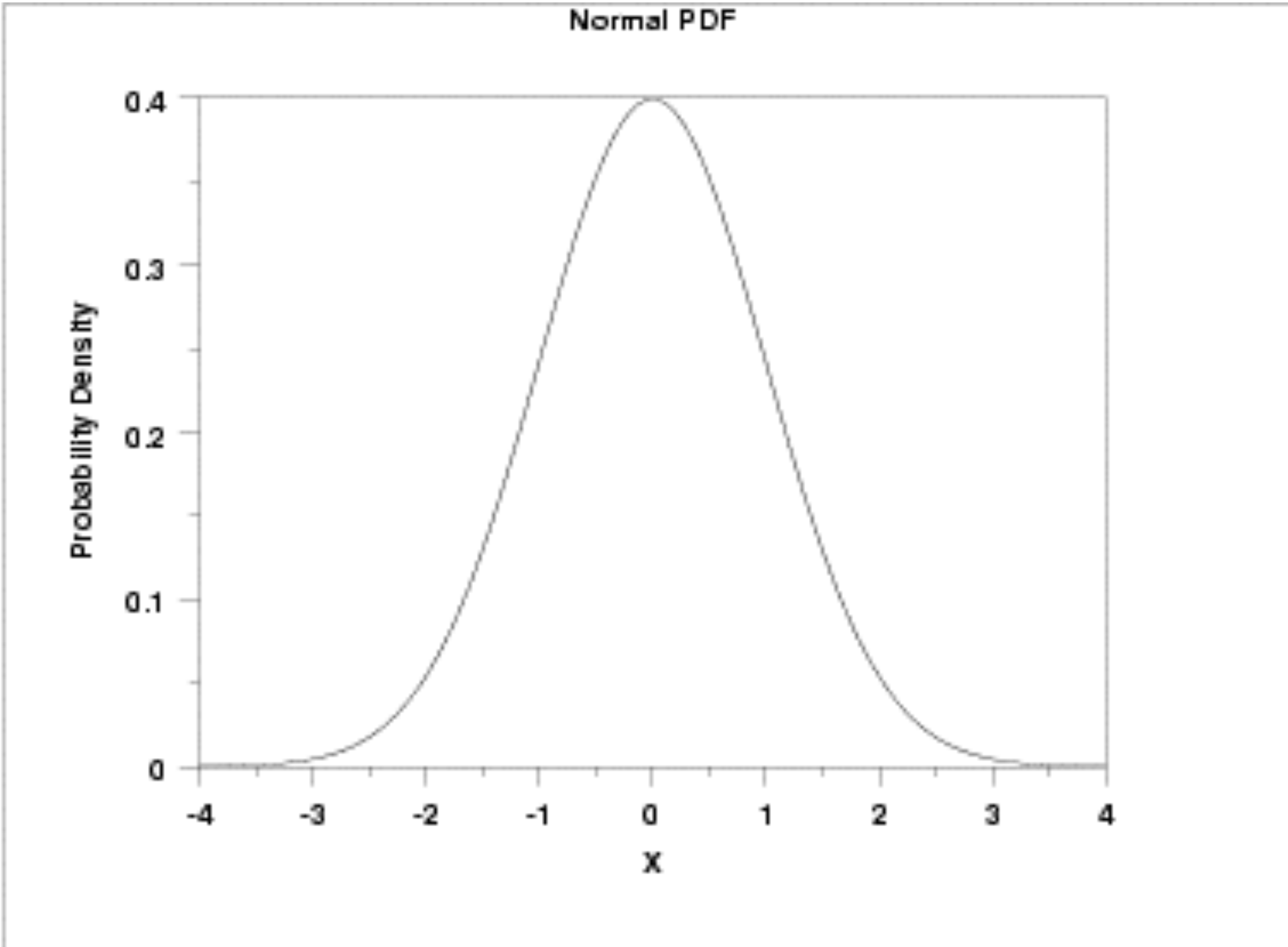
For a random variable, Y, the probability Y is equal to some particular value y in the range of Y defines the probability density function.

- It is possible to define a PDF of discrete events but not continuous ones.

- It is possible to define a PDF for a range of a continuous variable.

  – e.g. families earning between \$30,000 and \$40,000.

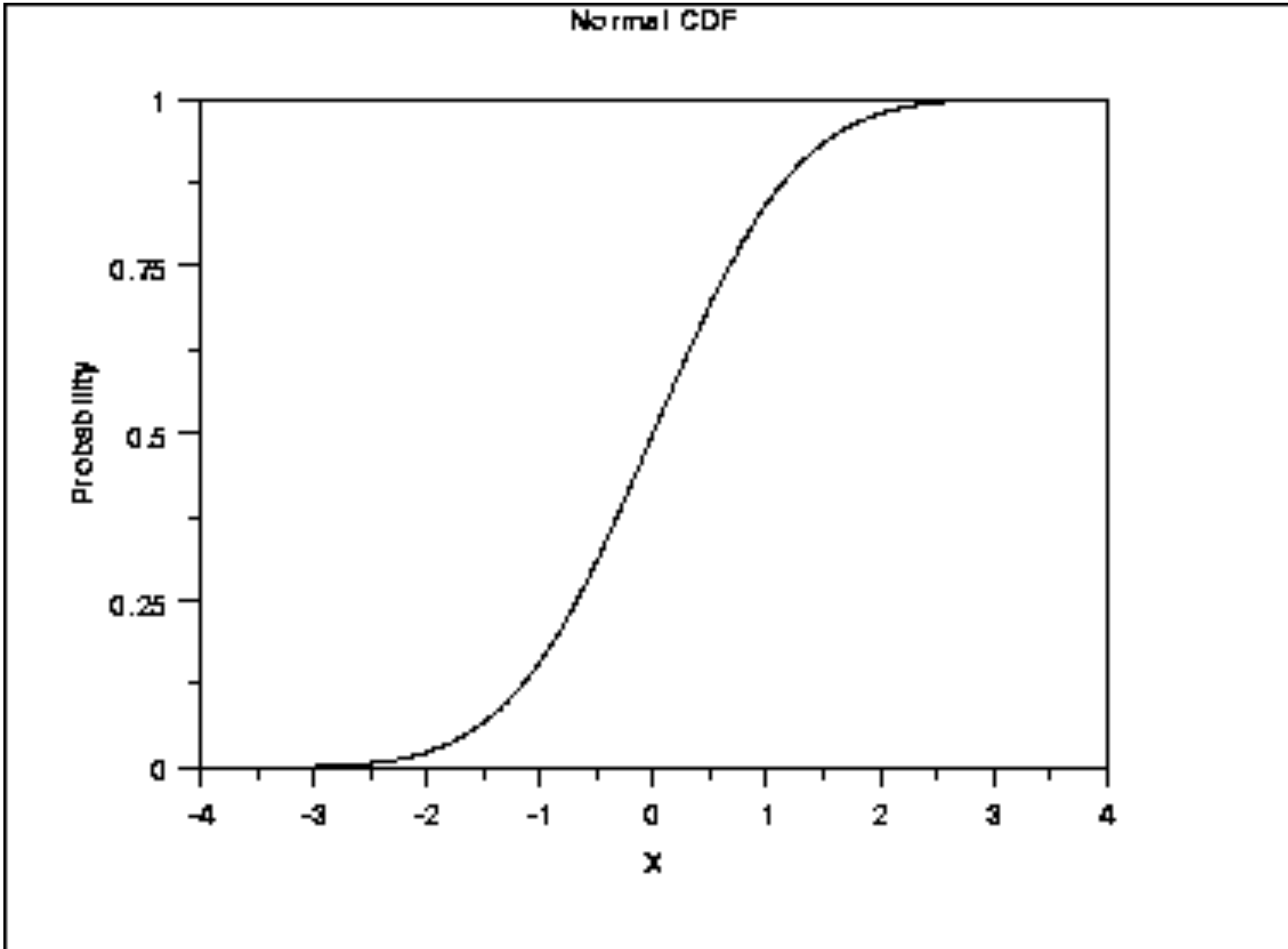
The Normal Distribution



Cumulative Density Function definition

For a random variable, Y, the probability Y is less than or equal to some particular value, y, in the range of Y defines the cumulative density function.

Cumulative Density Function



So the PDF represents the probability at a point, while the CDF represents the cumulative probability ***up to*** a point.

Examples

Coin flip example

	Bias of coin towards heads				
Heads	.1	.3	.5	.7	.9
0	.59	.17	.03	.00	.00
1	.33	.36	.16	.03	.00
2	.07	.31	.31	.13	.01
3	.01	.13	.31	.31	.07
4	.00	.03	.16	.36	.33
5	.00	.00	.03	.17	.59

- The Normal distribution is the most widely used distribution in social sciences.
  - It seems to fit a lot of what we try to measure.
  - Estimation techniques we use often assume that the coefficients and disturbance terms are normally distributed.

Linking DVs, distributions, and models

- Since least squares assumes normality, non-normal data pose a major problem for LS.

- ML produces estimates that are normally distributed of the underlying phenomenon we are interested in.

- To do this, we need to transform the estimates into the underlying distribution.

Several main points to think about

1. Always think about the distribution of our dependent variables.

2. Always recognize the possibility of censoring, truncation, skewedness, and other quirks of the dependent variable.

3. Once we choose a distribution to describe the DV, we need some way to link that distribution to some independent variables.

4. These choices have a substantial effect on what models we estimate.

Several main points to think about

1. Always think about the distribution of our dependent variables.

2. Always recognize the possibility of censoring, truncation, skewedness, and other quirks of the dependent variable.

3. Once we choose a distribution to describe the DV, we need some way to link that distribution to some independent variables.

4. These choices have a substantial effect on what models we estimate.

Several main points to think about

1. Always think about the distribution of our dependent variables.

2. Always recognize the possibility of censoring, truncation, skewedness, and other quirks of the dependent variable.

3. Once we choose a distribution to describe the DV, we need some way to link that distribution to some independent variables.

4. These choices have a substantial effect on what models we estimate.

Several main points to think about

1. Always think about the distribution of our dependent variables.

2. Always recognize the possibility of censoring, truncation, skewedness, and other quirks of the dependent variable.

3. Once we choose a distribution to describe the DV, we need some way to link that distribution to some independent variables.

4. These choices have a substantial effect on what models we estimate.

Several main points to think about

1. Always think about the distribution of our dependent variables.

2. Always recognize the possibility of censoring, truncation, skewedness, and other quirks of the dependent variable.

3. Once we choose a distribution to describe the DV, we need some way to link that distribution to some independent variables.

4. These choices have a substantial effect on what models we estimate.

Several main points to think about

1. Always think about the distribution of our dependent variables.

2. Always recognize the possibility of censoring, truncation, skewedness, and other quirks of the dependent variable.

3. Once we choose a distribution to describe the DV, we need some way to link that distribution to some independent variables.

4. These choices have a substantial effect on what models we estimate.

- Other books online free from UNO library:
  - Franses, Philip Hans. 2003. *Concise Introduction to Econometrics : An Intuitive Guide*. Cambridge University Press
  - Kennedy, Paul. 2003. *A Guide to Econometrics*. Fifth Edition. MIT Press.

- Questions?