UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

Introductory Econometrics 771

Appendix B - Fundamentals of Probability

Department of Economics

DEPARTMENT OF ECONOMICS

Motivation

- We spoke about randomness earlier
- Nothing is "certain" in Econometrics, but we would like to attach some *probability* to events
  - And the results we obtain from analyses

An example used throughout

- Airline with 100 seats available
  - Should the airline book 100 seats?
  - Or overbook?
    - Because there is a probability that some people will not show up
    - Resulting in an underutilised plane with lower profits
    - But running the risk of having to compensate passengers who are bumped off a flight that *could probably* be full

- Some definitions
  - Experiment
    - *Procedure that can be **infinitely be repeated** and has a **well-defined set of outcomes***
    - Example: tossing coin; outcome can only be H or T
  - Random Variable
    - Takes on numerical values that are the outcome of an experiment
    - Example: Number of H when flipping a coin ten times
      - Random? – We do not know ahead how often this will happen
      - If we conduct 10 trials and then another 10 trials, we will obtain different outcomes for the two experiment sets
    - Example: Number of individuals showing up for a flight

B1. Random variables and their probability distributions

- Random variable denoted with UPPERCASE letters
  - *W, X, Y, Z*
  - *X = #Times H is thrown – before the trials have been conducted*
    - Possible result is part of set of predefined outcomes: {0,1,2,…,10}
- *OUTCOMES* denoted with LOWERCASE letters
  - *w, x, y, z*
  - *x = 4 – an outcome of one of the trials after it is conducted*
  - *x = 6 – another outcome of the trial*
- Collection of random variables X_1, X_2,… X_{20}
  - Each represents (say) the household income of 20 different households
  - With possible outcomes : x_1=10000, x_2=15000… x_{20}=*2000*
  - Why a random variable for *each* household
    - Because before the experiment we only know a possible set of outcomes which could realise for *each* household
    - 0 to infinity

- "Qualitative" random variables?
  - H or T are not numbers?
  - Define random variable as *X = 1 if H is thrown, 0 if T is thrown*
  - Binary variable: "success/failure"
    - Terminology is borrowed from statistical literature, but does not necessarily have any value attached to it
    - *Bernoulli* random variable

Discrete Random Variables

- Takes on only finite or countably infinite number of values
- Example: Bernoulli
  - Unbiased coin: $P(X = 1) = 0.5$ and $P(X = 0) = 0.5$
- Example: Number of people showing up for flight
  - Define for random customer
    - X = 1 if customer shows up *with* $P(X = 1) = \theta$
    - X = 0 if customer does not show up *with* $P(X = 0) = 1 - \theta$
    - *"Theta"* (θ) is crucial to airline's decision
      - Estimate using historical data!
- Generally, for *k* possible outcomes $\{x_1, x_2, ...x_k\}$

$$p_j = P(X = x_j), j = 1, 2, ...k$$

$$0 \leq p_j \leq 1$$

$$\sum_{j=1}^{k} p_j = 1$$

Discrete Random Variables

- Probability Density Function (pdf) of X summarises all info concerning outcomes of X and their probability of occuring

$$f(x_j) = p_j, j = 1, 2, ..., k$$

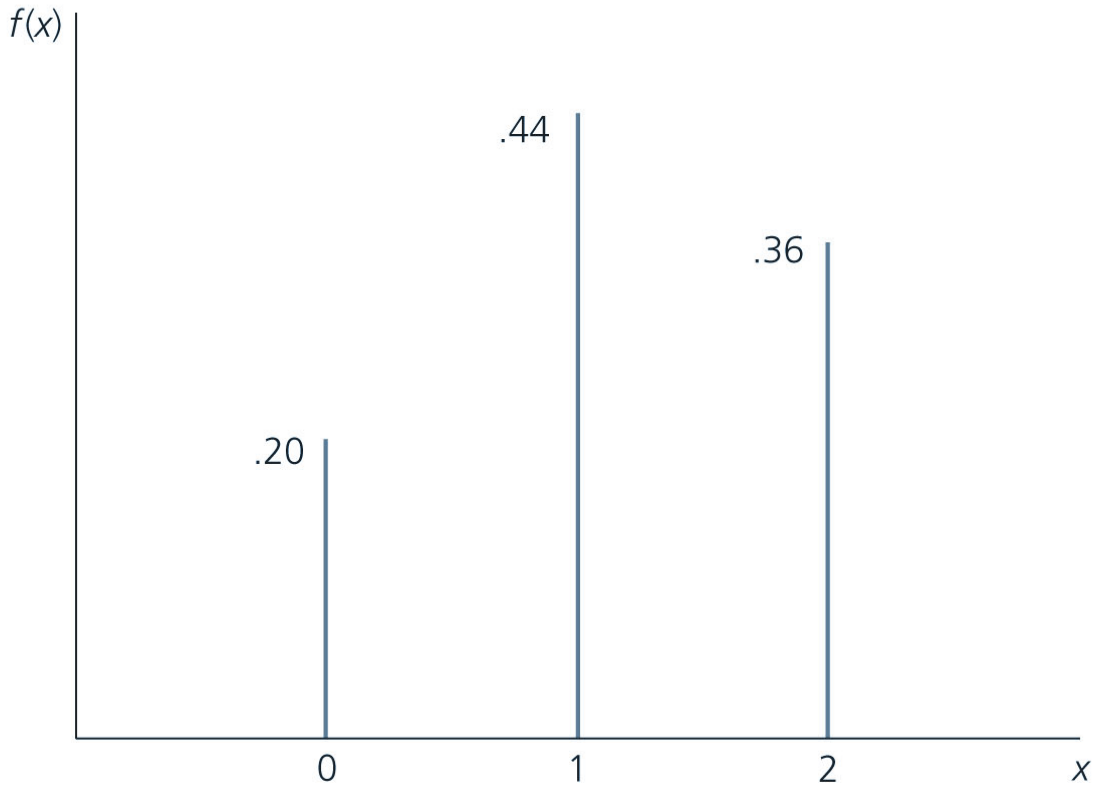
$$f(x) = 0, j \neq 1, 2, ..., k$$

- Example

PDF

FIGURE B.1

The pdf of the number of free throws made out of two attempts.



Continuous Random Variables

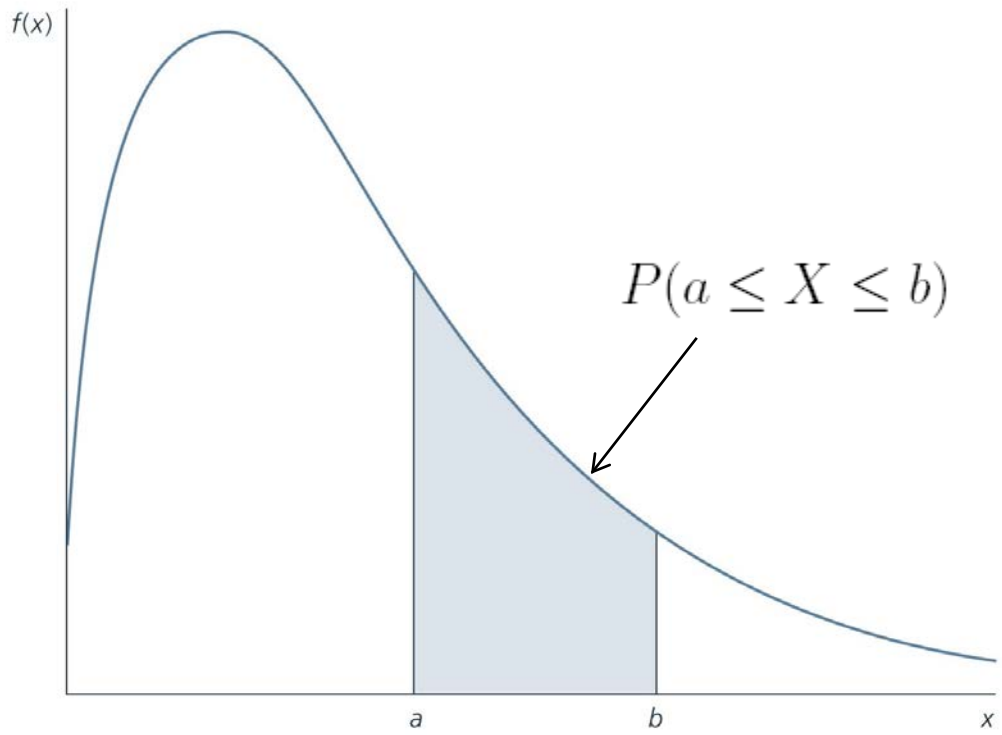
- Can take on *any* real value with *zero* probability
  - Even though once we see the outcome, it is actually realised!
  - But a value very-very-very close to the realised value was equally possible, so that it may as well have been that value
    - Cannot match the probability with the infinitely possible outcomes
- Example: monetary amounts measured in *cents*
  - We know the outcome can only be 100 values (for the first Rand, at least)
  - So technically it is discrete
  - But the "distance" between 10c and 11c is so small to us that we treat it as a continuous variable
  - Also so many possible outcomes that we do not want to find the info on each of these

Continuous Random Variables

- Work with ranges – area under curve (integral)
  - Area under whole pdf is 1

FIGURE B.2

The probability that X lies between the points a and b.



Continuous Random Variables

- Easier to work with Cumulative Distribution Function (cdf):

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

  - Discrete: sum the probabilities for all the values below *x*
  - Continuous: area under pdf to the left of *x*
    - integration
    - Useful things to know in using the cdf:

$$P(X > c) = 1 - P(X \leq c) = 1 - F(c)$$

$$P(a \leq X \leq b) = F(b) - F(a)$$

$$P(X > c) = P(X \geq c)$$

B2. Joint & Conditional Distributions/ Independence

- Not only interested in one variable at a time
  - What about considering the probability of arriving for a flight *and* being a business traveller? (JOINT)
    - Still using all travellers to calculate probability
  - Or arriving for flight *conditional* on being a business traveller?
    - Only using business travellers to calculate probability
  - Or independence – business travel independent of flight arrival (ie no relationship).

Joint Distributions and Independence

- *X* and *Y* discrete, with (*X*,*Y*) having a joint distribution with joint pdf:

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

- Can also be defined for continuous variables, but not NB for our purposes
- X and Y are independent if and only if

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

Joint Distributions and Independence

- Example
  - Two throws into basket
  - X = get first throw in ~Bernoulli
  - Y = get second throw in ~Bernoulli
  - Suppose P(X=1) = P(Y=1) = 0.8
    - IF X and Y are independent (ie, if the second throw is not influenced by the outcome of the first throw)

$$P(X = 1, Y = 1)$$
$$= P(X = 1)P(Y = 1) = 0.8 * 0.8 = 0.64$$

    - Meaning, the probability of getting both throws in is 0.64!
    - NOT valid if variables are dependent on each other
- Def can be extended to more than 2 random variables

Using independence

- Back to airline example
  - Suppose arriving is denoted by Y_i for each passenger (i)
  - θ is again *P("success") = P(use reservation)*
  - Therefore every $Y_i \sim Bernoulli(\theta)$
  - If each person's decision to use the ticket is independent and Bernoulli distributed
    - (not necessarily true, because of group bookings)
    - Let $X = Y_1 + Y_2 + Y_3 + \ldots + Y_n$ = # *arrivals*
    - Then $X \sim$ *binomial*$(n ; \theta)$ *with pdf*

$$f(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, x = 0, 1, 2, ..., n$$

- Therefore, say the flight can take 100 people, but 120 have booked, what is the probability that more than 100 people will arrive?

Using independence

$$P(X > 100) = P(X = 101) + P(X = 102) + ... + P(X = 120)$$

- With *n=120* and θ =0.85 and using the formula for the binomial pdf, we get 0.659
  - which is a high risk for airline
- If *n=110 and* θ =0.85, we get 0.024, which is low risk

Conditional Distributions

- Suppose we would only want to know the properties of a distribution for a certain sub-population
  - Use conditional distribution of *Y* **given** *X*
  - Tells us something about *Y* given that we are at value *X*
    - *This in essence is the basis of the rest of this course!*

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

    - If X and Y are independent, then the conditional distribution of Y is simply the same as the pdf of "unconditional" Y
      - Intuitively correct!

Conditional Distributions

- X = first throw a success?
- Y = second throw a success?
  - Does the second throw depend on the first throw?
    - Conditional distribution:

$$\left.\begin{aligned} f_{Y|X}(1|1) &= 0.85 \\ f_{Y|X}(0|1) &= 0.15 \end{aligned}\right\} = 1$$

$$\left.\begin{aligned} f_{Y|X}(1|0) &= 0.70 \\ f_{Y|X}(0|0) &= 0.30 \end{aligned}\right\} = 1$$

    - At first glance, yes…
      - Getting the first throw in, increases the chance of getting the second in, while it decreases the chance of missing the second
  - We will see later in the course how this translates to regression analysis

B3. Features of Probability Distributions

- Measure of Central Tendency
  - *Expected value* of a distribution $\mathrm{E}(\mathrm{X})=\mu$
    - Weight each outcome of the random variable by its probability of occuring
    - If we know the pdf exactly, this will give us the *population mean*, or the average we would have if we knew everything about the variable
      - For a discrete variable:

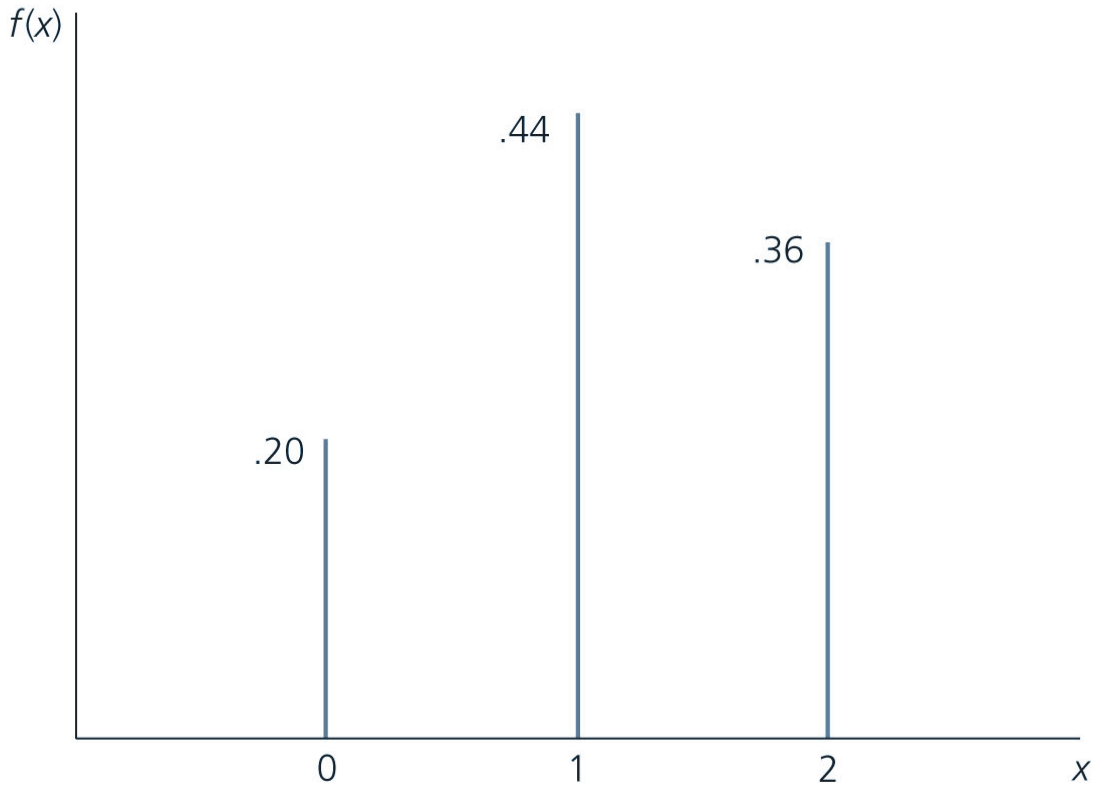
$$\mathrm{E}(\mathrm{X})=\mathrm{x}_1*f(x_1)+x_2*f(x_2)+\ldots+x_k*f(x_k)$$
$$=\sum_{j=1}^{k} x_j f(x_j)$$

    - Expected value of number of free throws from two attempts?

E(X)=…

FIGURE B.1

The pdf of the number of free throws made out of two attempts.



Expected Values

- Need not be one of the outcomes of the variables
  - So it is a bit clumsy for a discrete variable
  - Makes more sense for a continuous variable
    - Now, because we are summing over all the values between the discrete points, we need to take the integral (same as measuring the area under the weighted *pdf*)

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

    - Still a weighted average
  - The expected value of a function of a variable?

$$\mathrm{E}[\mathrm{g}(\mathrm{X})]=\int_{-\infty}^{\infty} g(x) f_X(x) dx$$

  - And

$$\mathrm{E}[\mathrm{g}(\mathrm{X})] \neq g[E(X)] \qquad \mathrm{E}[\mathrm{X}^2] \neq [E(X)]^2$$

Properties of Expected Values

$$\mathrm{E(c)=c}$$
$$\mathrm{E(aX+b)=aE(X)+b}$$
$$\mathrm{E}(\textstyle\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i E(X_i)$$

- Last line
  - for $a_i = 1$ we get that the expected value of a sum is the sum of expected values

Airline example

- Suppose *X*~*Binomial(n,* θ *)*
  - Or the sum of *n Bernoulli(*θ*)* variables (Y_i)
    - $X = Y_1 + Y_2 + Y_3 + \ldots + Y_n$

$$\mathrm{E}(\mathrm{X}) = \mathrm{E}\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} E(Y_i) = \sum_{i=1}^{n} (\theta) = n\theta$$

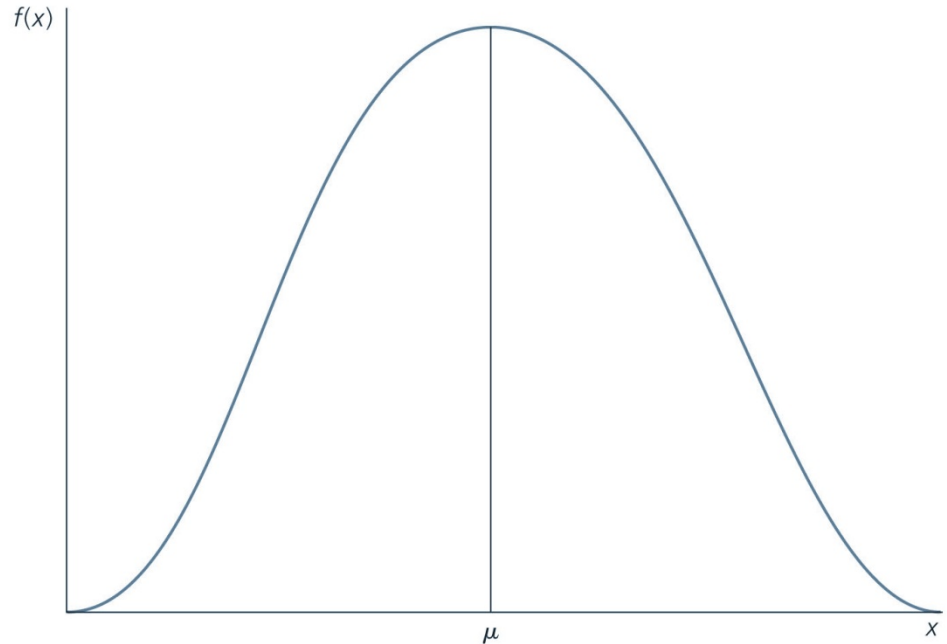
- *n*=120 and θ =0.85
  - *E(X)* = 120*0.85 = 102 >100 !!!
- *n*=110 and θ =0.85
  - *E(X)* = 110*0.85 = 93.5 <100

More measures of central tendency

- Median
  - Divides the area under pdf in half
  - Not the same as the expected value, *except if the distribution is symmetric*
  - Sometimes used if our summary measure is sensitive to outliers

FIGURE B.3

A symmetric probability distribution.



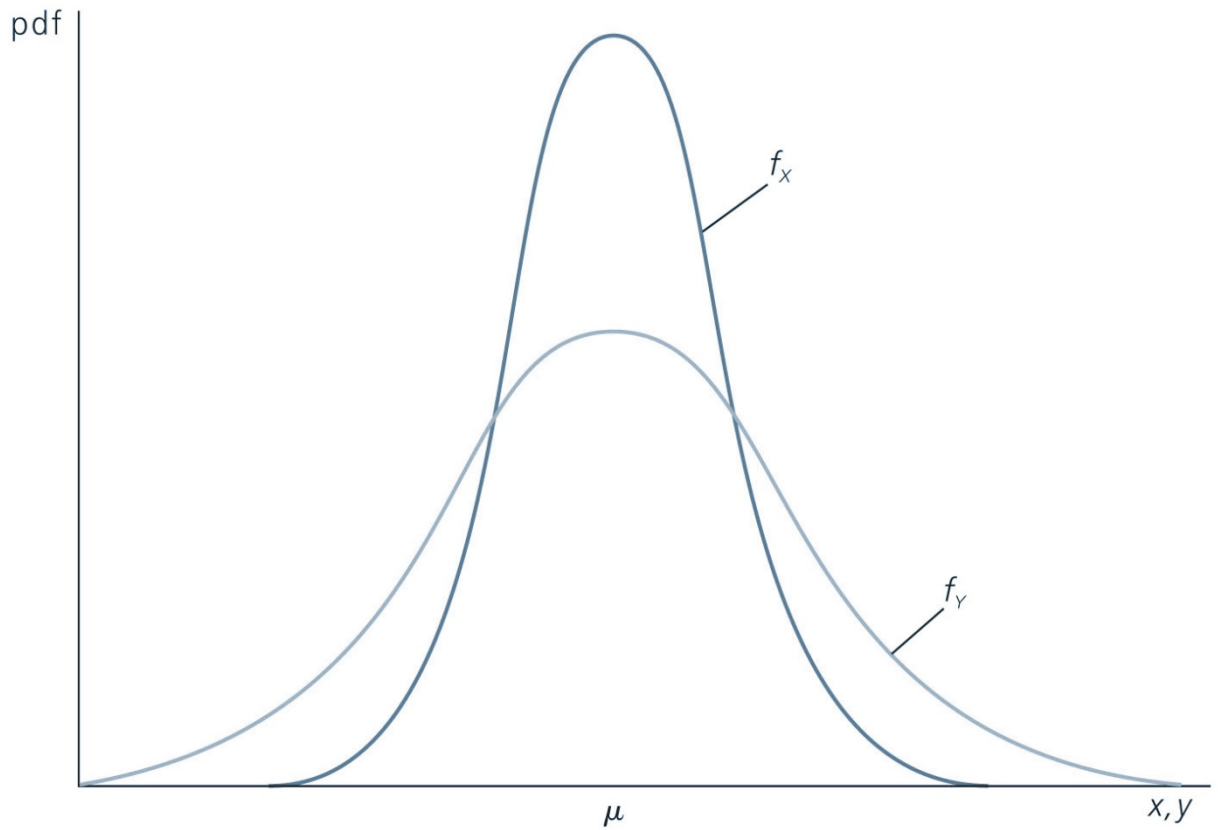
Measures of Variability

- Economists are often accused of only looking at the "average" case
  - Perhaps focusing on measures of central tendency is what brings forth this accusation
    - But we are actually interested in the entire distribution!
  - Variability helps us to see how much of the distribution is actually concentrated around this "central" area

Same central tendency, different variation

FIGURE B.4

Random variables with the same mean but different distributions.



Variance

- Measures the "squared distance" of each point from the mean
  - This distance is also random, as it can change with every outcome of *X*, so we take the expected value of this
    - How far is each *X on average* from the mean?
      - So we are perhaps too obsessed with averages ☺

$$\mathrm{Var(X)}=\mathrm{E}[(\mathrm{X}\text{-}\mu)^2] = \sigma_X^2 \succ 0$$

      - Sometimes for computational ease, we note that:

$$\mathrm{Var(X)} = \mathrm{E}(\mathrm{X}^2) - \mu^2$$

Variance

- Properties
  - *Var(X)* = 0 if the random variable is a constant
    - Ie the distance between *X* and *E(X)* is zero in all instances
  - Similar property as we had for *E(X)*, but note the difference

$$\mathrm{Var}(\mathrm{aX}+\mathrm{b})=\mathrm{a}^2 Var(X)$$

Standard deviation

- Simply the positive square root of the variance
  - Similar properties to variance

Standardizing random variables

$$\text{Z}=(\text{X}-\mu)/\sigma = (1/\sigma) * X - (1/\sigma) * \mu$$

- Useful in many applications to center the variable around zero

$$\begin{aligned}\text{E}(\text{Z}) &= (1/\sigma) * E(X) - (1/\sigma) * \mu \\ &= (1/\sigma) * \mu - (1/\sigma) * \mu = 0\end{aligned}$$

- And to give it a variance of 1

$$\text{Var}(\text{Z}) = (1/\sigma)^2 * Var(X) = \sigma^2/\sigma^2 = 1$$

- We will work with zero mean variables in regression models later in the course
- NOTE: this is not to be confused with a variable being "normally distributed"
  - *Even if we get standardized normal distributions (more later)*

Skewness and Kurtosis

- Skewness
  - In contrast to a symmetric distribution, a skew distribution has long tails
  - Use 3rd order moments to establish this
- Kurtosis
  - Is the distribution "sharp" or "flat" at *E(X)*
  - Usually compared to normal distribution (symmetric)
    - Later
  - Use 4th order moments to establish this

- Measures of association
  - Summary measures for the joint distribution of two variables
- Covariance
  - Two random variables *X* and *Y*

$$\mu_X = E(X)$$
$$\mu_Y = E(Y)$$
$$(\text{X-}\mu_X)(Y - \mu_Y)$$

    - If *X* is above its mean and the same for *Y* then $(\text{X-}\mu_X)(Y - \mu_Y)$ >0
      - Similar if both are below their mean
    - If one is above and the other below its mean, then $(\text{X-}\mu_X)(Y - \mu_Y)$ <0
  - $(\text{X-}\mu_X)(Y - \mu_Y)$ therefore indicates if two variables are in the "same position" relative to their mean

- Covariance
  - Therefore average this for the whole distribution of *X* and *Y* to obtain a measure of whether the variables are positively or negatively associated
    - Magnitude is difficult to interpret

$$\mathrm{Cov(X,Y)=E[(X-}\mu_X)(Y-\mu_Y)] = \sigma_{XY}$$

    - >0: positive **linear** relationship
    - <0: negative **linear** relationship
  - Alternative formulation

$$\mathrm{Cov(X,Y)=E[XY]-E[X]E[Y]}$$
$$\mathrm{=E[XY]\ if\ E[X]=0\ or\ E[Y]=0}$$

    - The latter is often true, so in many contexts you will see the final expression instead of references to Covariance

Properties of Covariance

- If *X* and *Y* are independent, *Cov(X,Y)* = 0
  - Note that the converse is not true in most cases!
- Changing the scale of a variable changes the covariance:

$$\mathrm{Cov}(a_1X + b_1; a_2Y + b_2) = a_1a_2Cov(X;Y)$$

  - Implication: changing unit can increase covariance dramatically
    - Therefore we do not attach value to magnitude of covariance
    - This is why we use the Correlation Coefficient

Correlation

- The "unitless" version of covariance

$$\rho_{XY} = Corr(X;Y) = \frac{Cov(X;Y)}{sd(X)sd(Y)} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

  - Sign is the same as covariance
    - Because standard deviations are always positive
  - Other properties (such as independence) also carry over

  - However: $-1 \leq \rho \leq 1$
    - Where 1 means a perfect positive linear relationship
    - Where -1 means a perfect negative linear relationship
    - 0 means no relationship at all

Properties of correlation

- Changing both variables by scale factors
  - If both are positive, then correlation does not change
  - If one is negative, then correlation becomes negative

Back to variance

- Covariance of a variable with itself is simply the Variance: $\mathrm{Cov}(X;X)=\mathrm{Var}(X)$
  - Can you show this?
  - Hence the reference to Variance-Covariance Matrix in some instances
- For constants *a* and *b*

$$\mathrm{Var}(aX+bY)=a^2Var(X)+b^2Var(Y)+2abCov(X;Y)$$

If $\mathrm{Cov}(X;Y)=0$:

$$\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$$

$$\mathrm{Var}(X-Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$$

  - *Can be extended to sums of more random variables*

Conditional Expectation

- In Econometrics we are interested in explaining one variable, say *Y (eg Wage)*, in terms of another, say *X (eg Education)*
  - Conditional distribution: how do wages change with respect to education? OR: how does the expected value of wages differ *given* different levels of education
  - Summary measure for this?
    - Conditional expectation
      - Suppose *X* has taken on a specific value, *x*
      - Compute expected value of *Y* given this outcome
        - Discrete

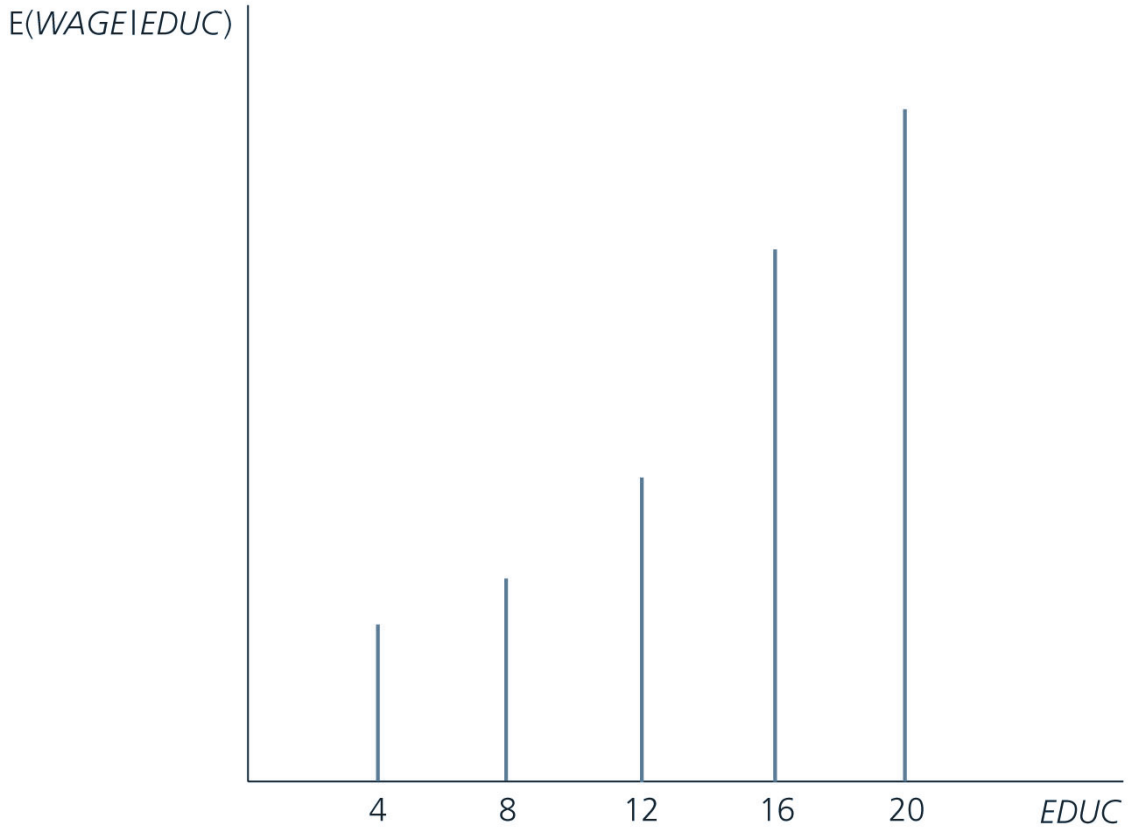
$$\mathrm{E}(\mathrm{Y}|X=x)=\sum_{j=1}^{m} y_j f_{Y|X}(y_j|X=x)$$

        - Integrate if continuous
      - THE BASIS of REGRESSION ANALYSIS and the REST OF THIS COURSE
        - We want to estimate these conditional relationships, for example $\mathrm{E}(\mathrm{WAGE}|EDUC)=1.05+0.45EDUC$

Illustration

FIGURE B.5

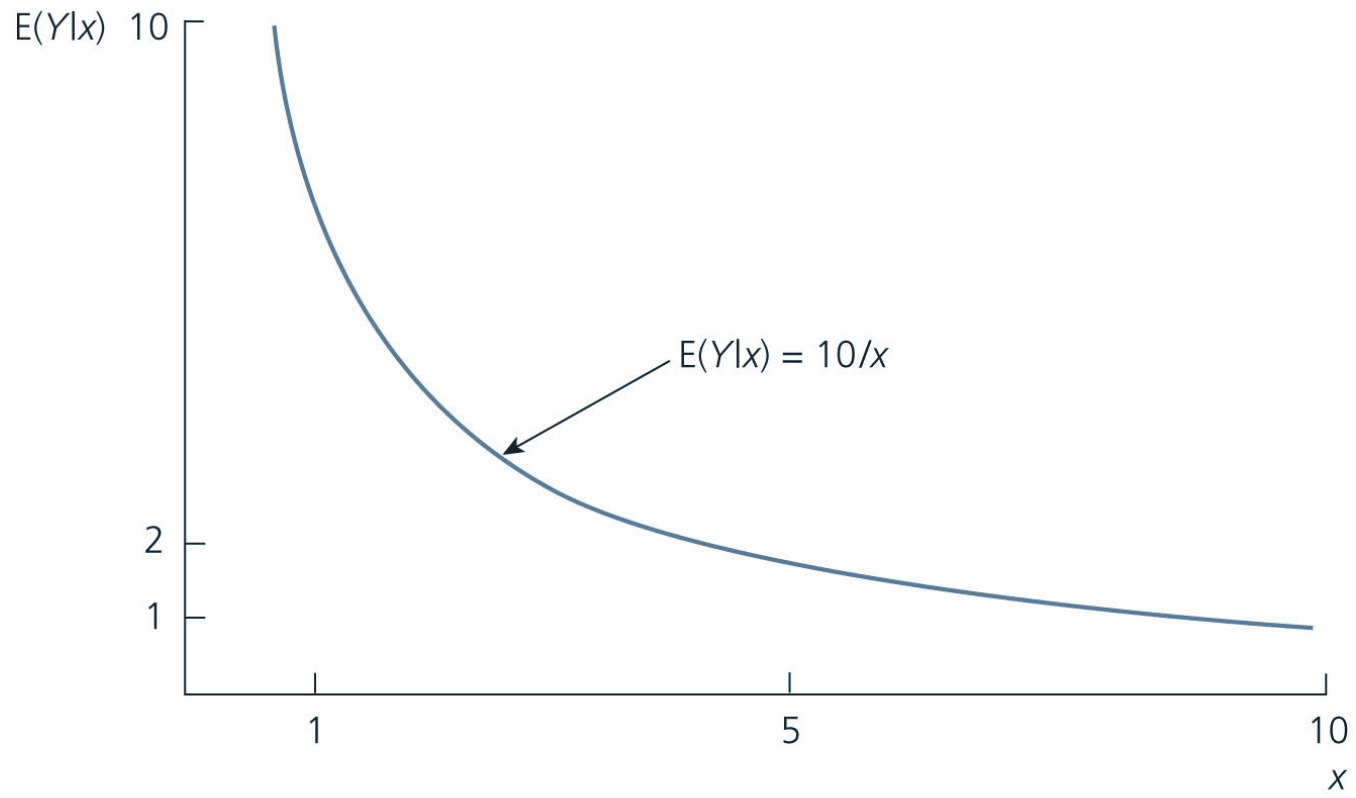
The expected value of hourly wage given various levels of education.



Or non-linear…

FIGURE B.6

Graph of E(*Y*|*x*) = 10/x.

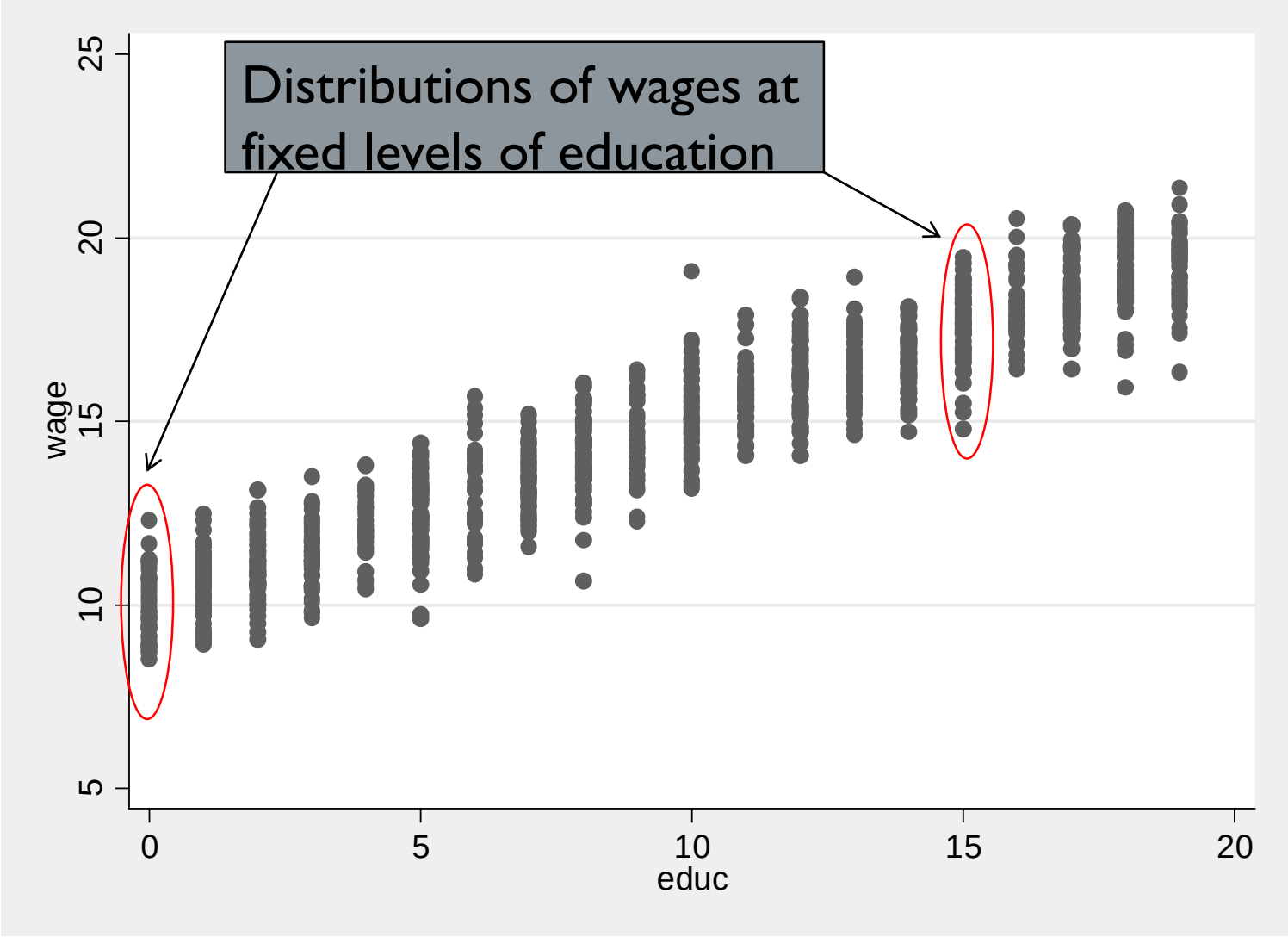


Illustration

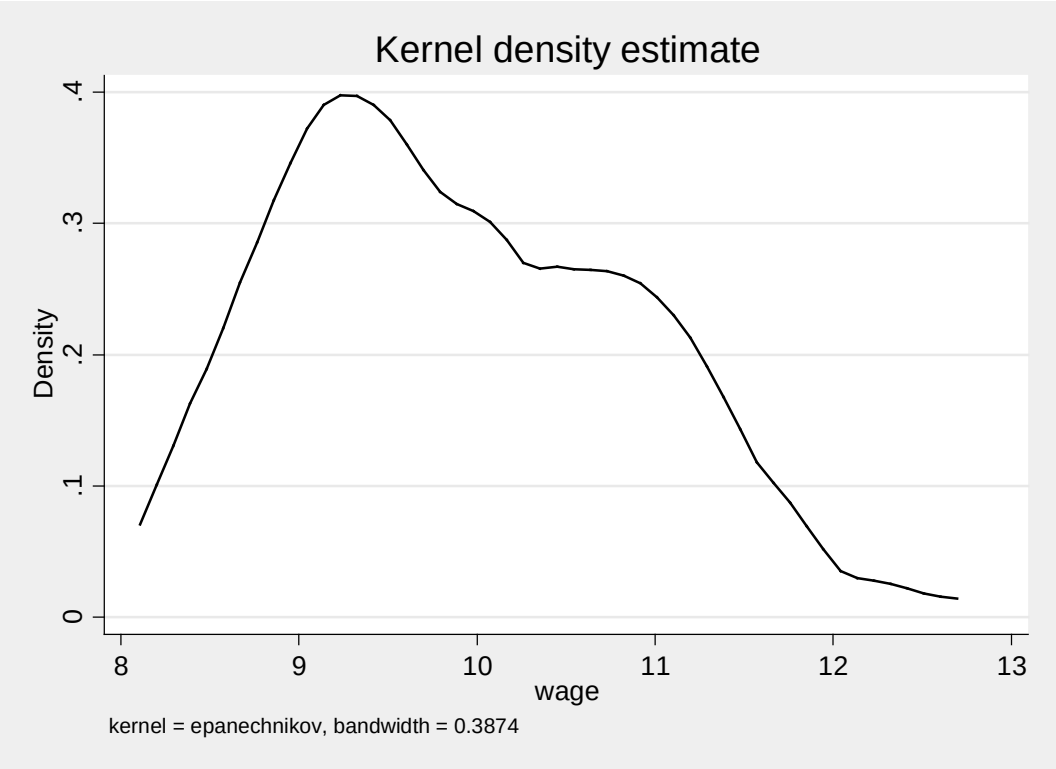
Appendix B - Monte Carlo - Conditi... | Untitled.do

```
clear all
set obs 1000
*Set up a generated process
gen educ = int(uniform()*20)
gen u = invnorm(uniform())
gen wage = 10 + 0.5*educ + u
*Run the regression
reg wage educ
*Plot the data to see the bivariate relationship
twoway(scatter wage educ)
*Now plot conditional distributions
kdensity wage if educ==0
kdensity wage if educ==15
*Illustrate E(wage|educ) as the regression line
tab educ, summ(wage)
egen E_wage = mean(wage), by(educ)
twoway(scatter wage educ) (lfit wage educ)
twoway(scatter wage educ) (lfit wage educ) (line E_wage educ, sort)
```

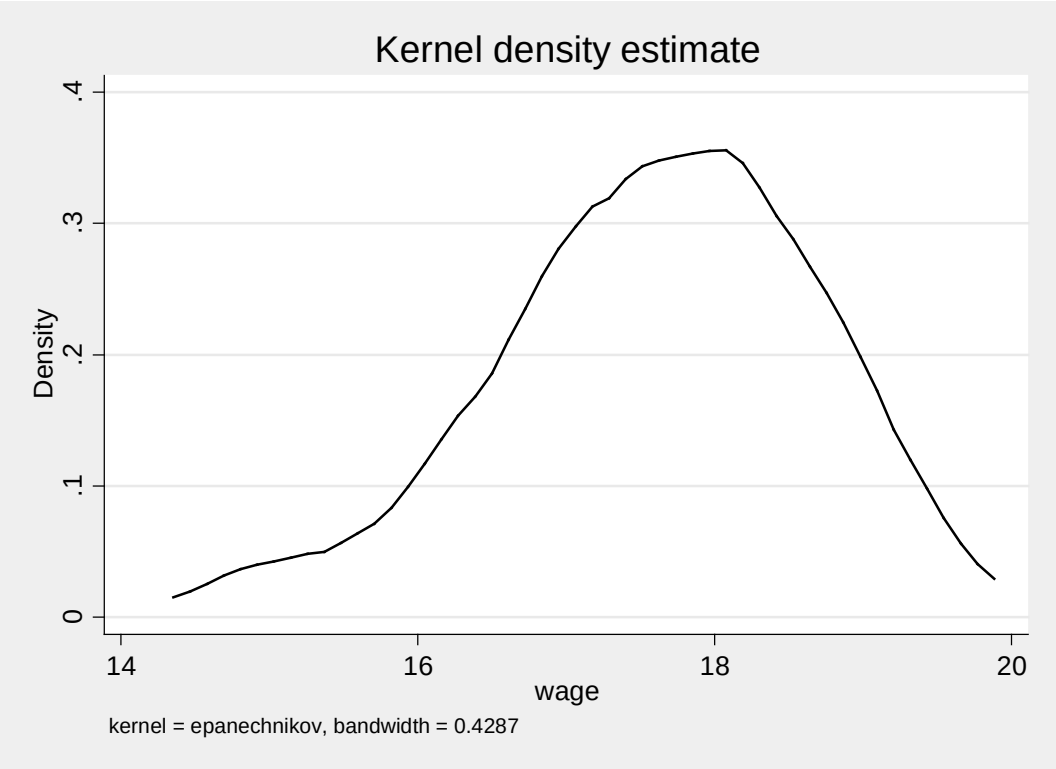
Illustration



Conditional Density of Wage at Educ=0



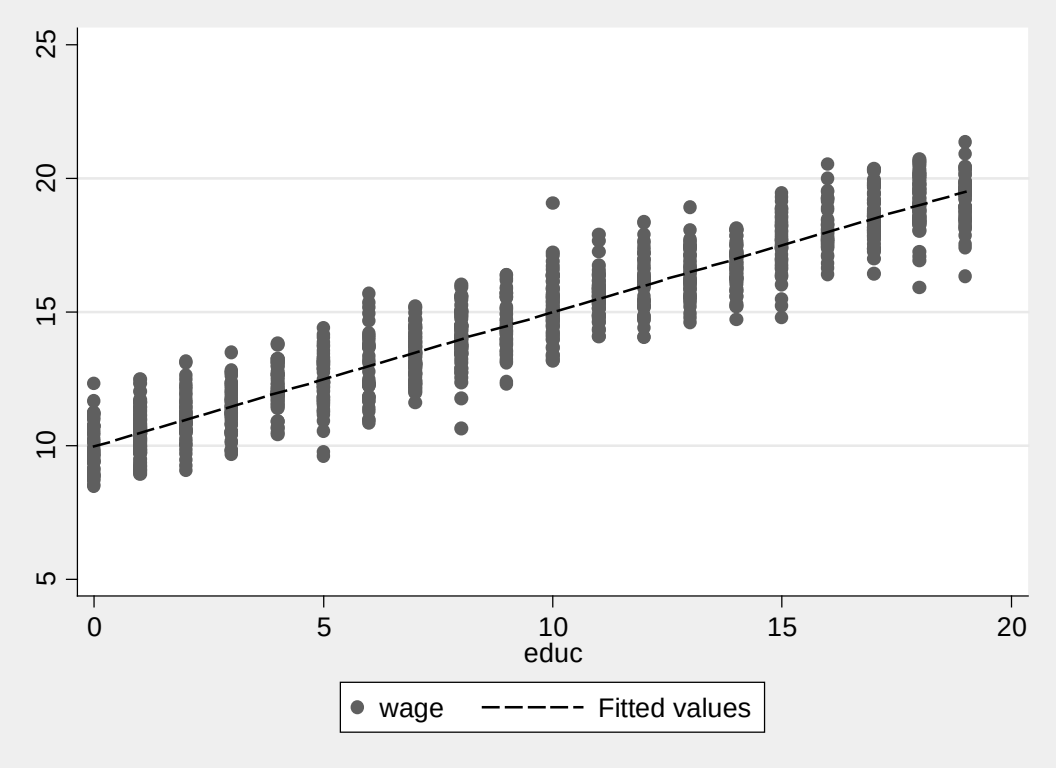
Conditional Density of Wage at Educ=15



Summary Statistics

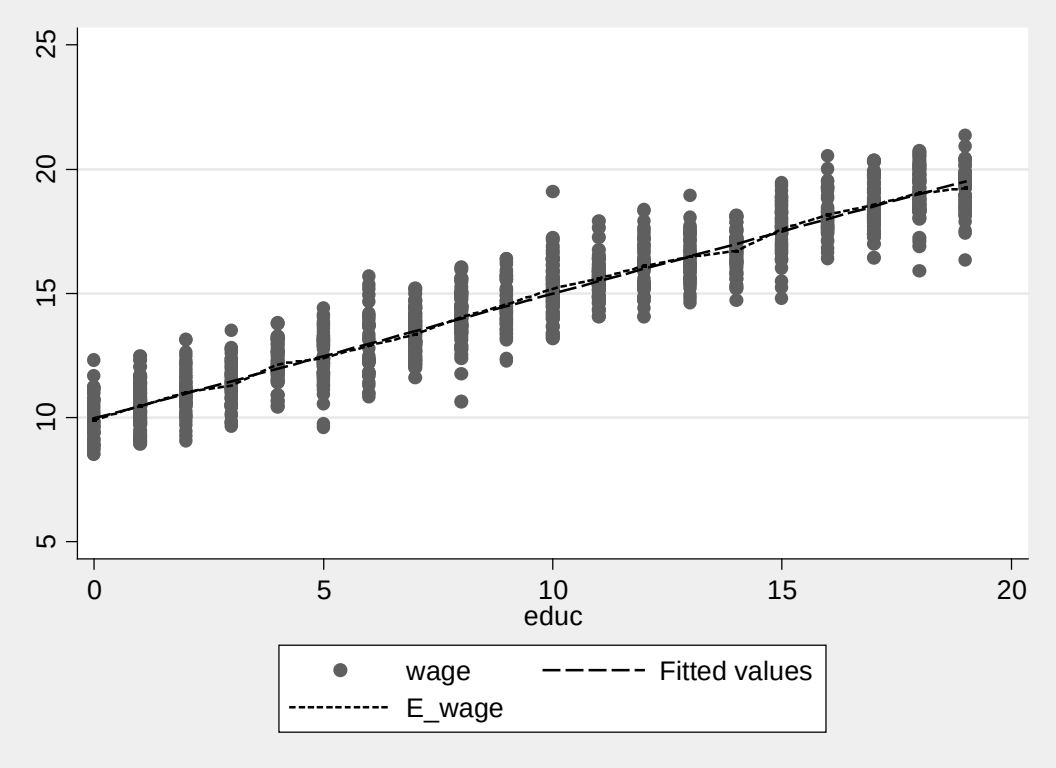
educ	Mean Wage	Std. Dev.	Freq.
0	9.905729	0.937564	49
1	10.47171	0.842344	56
2	11.0114	0.926732	47
3	11.30202	0.90709	47
4	12.14758	0.864056	42
5	12.40393	1.005311	56
6	12.89849	1.282484	40
7	13.34684	0.863737	65
8	14.04535	1.015865	58
9	14.55835	1.03795	50
10	15.1849	1.184434	54
11	15.59895	0.81929	48
12	16.10701	1.012083	55
13	16.47923	0.950206	45
14	16.70475	0.855622	45
15	17.59274	1.046137	49
16	18.16677	0.923558	40
17	18.55803	0.923854	48
18	19.03701	0.993022	53
19	19.2324	0.925188	53
Total	14.70705	3.043034	1000

What regression does



Essentially the same as a conditional mean…

- But we will extend this to multiple dimensions in regression estimates



Properties of Conditional Expectations

- Most important property

$$E[Y|X] = E[Y] if independent$$

- Also: if *E[Y|X] = E[Y]*, then *Cov(X,Y)=0*
- Other properties not as important for our purposes, but remember them if you encounter them

- Conditional Variance
  - Similar concept to conditional expectation
  - Will be important when we study heteroskedasticity
    - In other words does the variance of *Y* differ, conditional on different outcomes of *X*?

Conditional Standard Deviations

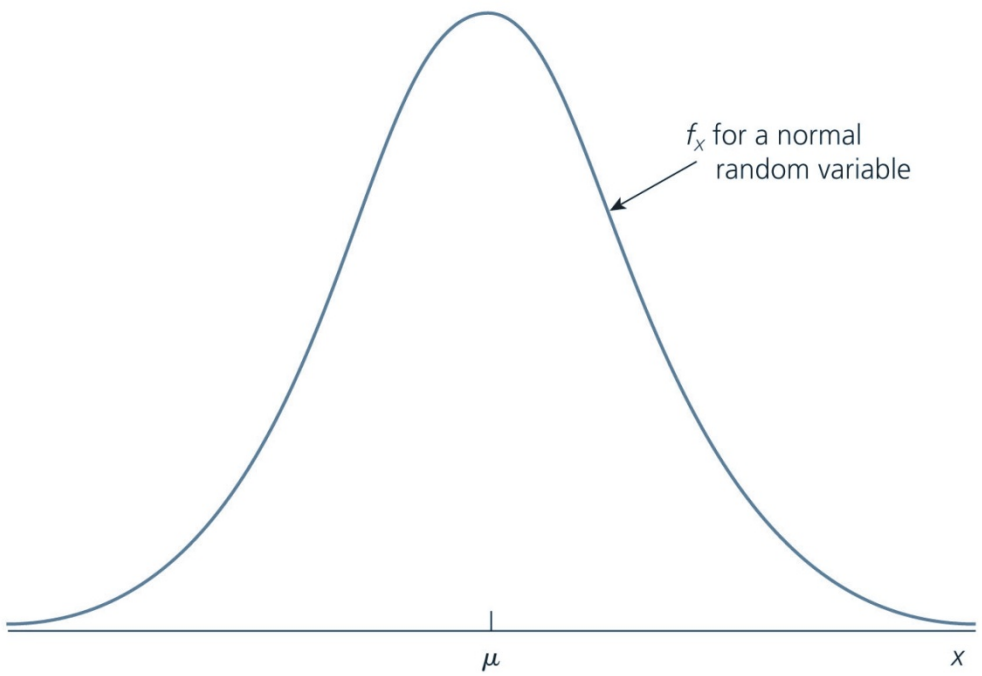
educ	Mean Wage	Std. Dev.	Freq.
0	9.905729	0.937564	49
1	10.47171	0.842344	56
2	11.0114	0.926732	47
3	11.30202	0.90709	47
4	12.14758	0.864056	42
5	12.40393	1.005311	56
6	12.89849	1.282484	40
7	13.34684	0.863737	65
8	14.04535	1.015865	58
9	14.55835	1.03795	50
10	15.1849	1.184434	54
11	15.59895	0.81929	48
12	16.10701	1.012083	55
13	16.47923	0.950206	45
14	16.70475	0.855622	45
15	17.59274	1.046137	49
16	18.16677	0.923558	40
17	18.55803	0.923854	48
18	19.03701	0.993022	53
19	19.2324	0.925188	53
Total	14.70705	3.043034	1000

Normal and Related Distributions

- Normal distribution (Gaussian distribution)
  - Basis for many other distributions
  - Used for inference in econometrics

FIGURE B.7

The general shape of the normal probability density function.



Normal Distribution

- Pdf formula is very complex
  - We fortunately have tables to deal with this!
  - Dependent *only* on the mean and variance of distribution
  - Symmetric $X \sim \mathrm{N}(\mu; \sigma^2)$
    - Therefore mean = median = μ
- Examples of variables that typically have this distribution
  - Height, weight, test scores
- Lognormal – non-symmetric distribution
  - If *log(X)* is normally distributed, then *X* is lognormally distributed
    - Income, Price, Wealth

Standard Normal Distribution

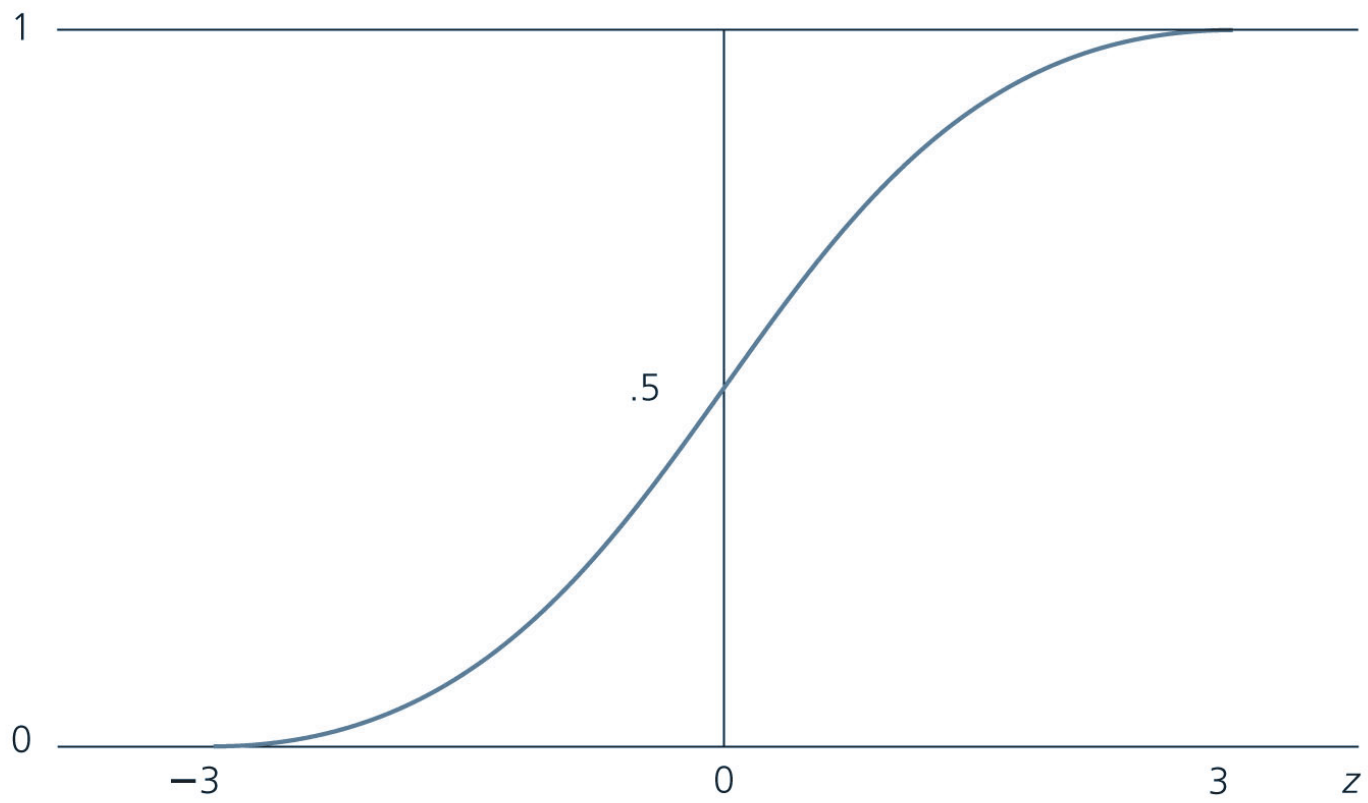
- *Z~N(0;1)*
  - *Remember standardisation?*
- Standard CDF often used, and is tabulated

$$\phi(z) = pdf$$
$$\Phi(z) = P(Z \prec z)$$
$$\mathrm{P}(\mathrm{Z} \succ z) = 1 - \Phi(z)$$
By symmetry:
$$\mathrm{P}(\mathrm{Z} \prec -z) = P(Z \succ z) = 1 - \Phi(z)$$
$$\mathrm{P}(\mathrm{a} \leq Z \leq b) = \Phi(b) - \Phi(a)$$

CDF

FIGURE B.8

The standard normal cumulative distribution function.



Properties of Standard Normal Distribution

- Use the following to convert any normal variable into a standard variable
  - Then you can use standardised tables to compute probabilities

$$X \sim N(\mu; \sigma^2) \Rightarrow \frac{X-\mu}{\sigma} \sim N(0;1)$$

  - Example

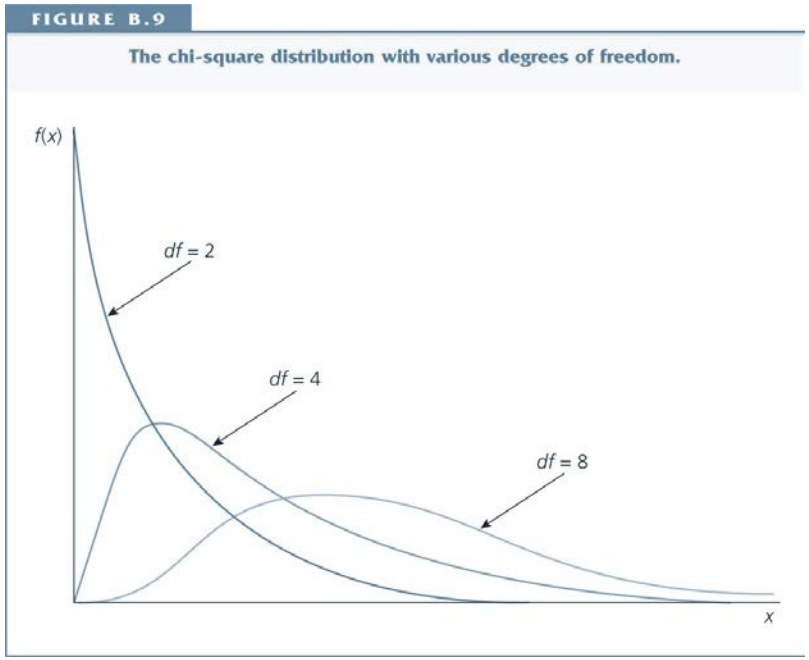
$$X \sim N(4;9)$$
$$P(2 \leq X \leq 6)$$
$$= P((2\text{-}4)/3 \leq (X-4)/3 \leq (6-4)/3))$$
$$= P(\text{-}2/3 \leq Z \leq 2/3)$$
$$= \Phi(0.67) - \Phi(-0.67)$$
$$= 0.749\text{-}0.251 = 0.498$$

Properties of the Normal Distribution

- Expectations and Variance Properties carry over
- Any linear combination of normally distributed variables is normally distributed
  - By implication the mean of normally distributed random variables is normally distributed
- Skewness = 0
- Kurtosis = 3

Chi-square Distribution

- If you square *n* standard normally distributed variables and sum then, then it has a Chi-Square distribution with *n* degrees of freedom
  - The higher the degrees of freedom, the more symmetric it becomes
    - Central limit theorem (later)



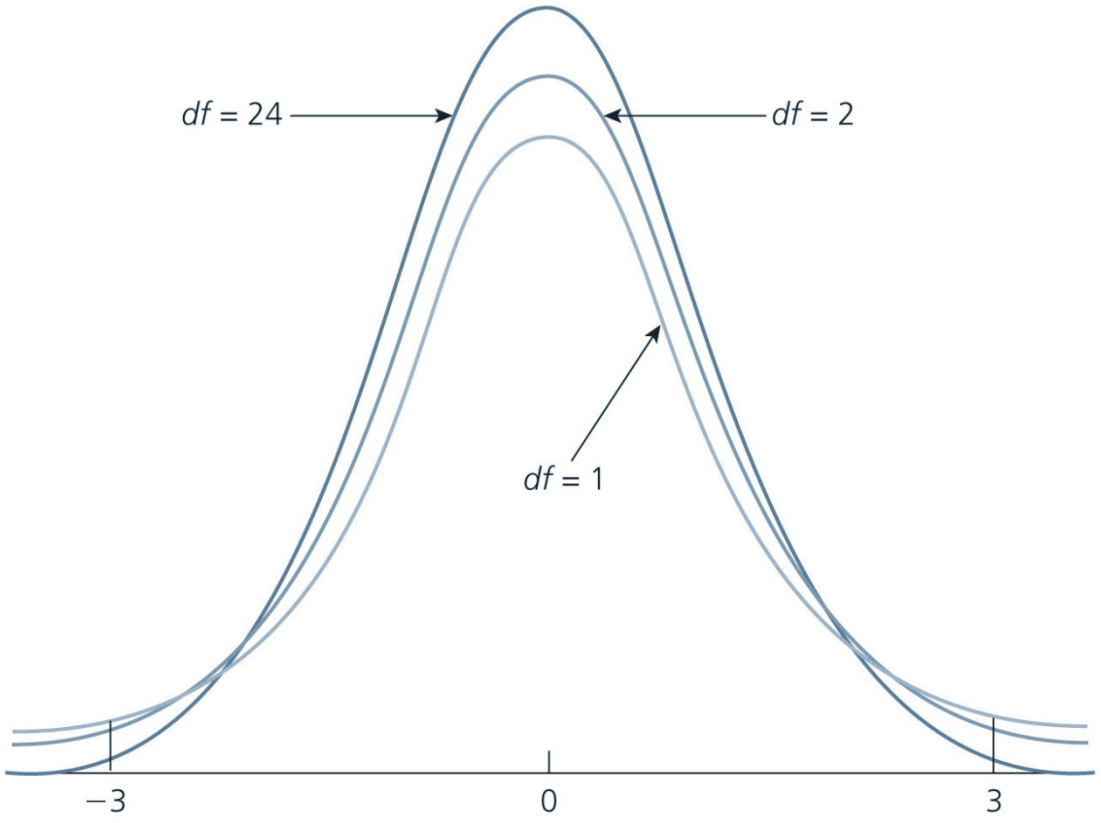
T-distribution

- Very common in econometrics
  - Derived from a normal and Chi-squared distribution
  - Similar to a normal distribution (symmetric)
    - Just more spread out (greater Variance)
    - As degrees of freedom increases, Variance becomes smaller and the distribution approaches a standard normal distribution

T-distribution

FIGURE B.10

The *t* distribution with various degrees of freedom.



Illustration

Appendix B - Monte Carlo - Distrib... | Untitled.do

```
1   clear all
2   set obs 1000
3   *Generate a random set of probabilities between 0 and 1
4   gen p = uniform()
5
6   *Simulate a normal Distribution
7   gen normal = invnorm(p)
8       *Plot the normal distribution to show it has a bell shape
9   twoway (hist normal) (kdensity normal)
10      *Obtain summary statistics to show that skewness = 0 and kurtosis = 3 (approximately)
11  summ normal, detail
12
13
14  *Simulate Chi2 distributions with various d.f.
15  gen chi2_1 = invchi2(1,p)
16  gen chi2_50 = invchi2(50,p)
17  gen chi2_900 = invchi2(900,p)
18      *Show that as d.f. increases, it becomes more symmetric
19  twoway (kdensity chi2_1) (kdensity chi2_50)
20      *Show that as d.f. increases, the skewness and kurtosis approach 0 and 3 respectively
21  summ chi2*, detail
22
23  *Simulate T distributions with various d.f.
24  gen t_1 = invttail(1,p)
25  gen t_50 = invttail(50,p)
26  gen t_900 = invttail(900,p)
27      *Show that as d.f. increases, it becomes more symmetric and that it approaches normality as d.f. -> infinity
28  twoway (kdensity t_1) (kdensity t_50)
29  twoway (kdensity t_900) (kdensity normal)
30      *Show that as d.f. increases, the skewness and kurtosis approach 0 and 3 respectively
31  summ t*, detail
```

F Distribution

- Important for simultaneously testing multiple hypotheses
  - Quotient of scaled Chi-squared Variables
  - Numerator AND Denominator degrees of freedom, associated with the respective Chi-Squared variables

F Distribution

FIGURE B.11

The F_{k_1,k_2} distribution for various degrees of freedom, k_1 and k_2.