Statistics
Tutorial 01

Philipp Scherer & Jens Wiederspohn

29.04.2020

Disclaimer!

- **The content of the slides partly relies on material of Philipp Prinz, a former Statistics tutor. Like us, he's just a student. Therefore we provide no guarantee for the content of the slides or other data/information of the tutorial.**
- **Please note that the slides will not cover the entire lecture content. To pass the exam, it is still absolutely necessary to deal with the Wooldridge in detail!**

Econometrics in general

What is econometrics all about?

- Econometrics = statistics for economists (and politicians)
- Mathematical tools (e.g. regression) for economic problems
  - Different focus and interpretation
- Use of statistical methods and empirical data to...
  - estimate (economic) relationships
  - test theories
  - evaluate policies
  - forecast variables

Econometrics in general

What is an econometric model in general?

- Econometric models are formalized economic models
  - How does demand change if the prices are increased?
    $demand_i = \beta_0 + \beta_1 * price_i + u_i$
  - What is the effect of election fraud on re-election perspectives?
    $votes_i = \beta_0 + \beta_1 * ability_i + \beta_2 * integrity_i + \beta_3 * fraud_i + u_i$
- Relationship between dependent and independent variables
- The behaviour of variables is described with equations
- Parameters (β) are estimates of the effect size of a variable
- Error term (u) contains all unobserved factors
  - Specification of error term affects models massively!

Econometrics in general

Important data structures and their characteristics

1. Cross sectional data
   - Sample of units at one point in time (ignore small time diff.)
2. Time Series Data
   - Observations on a variable over time (annually, monthly,...)
   - Often, we need to account for time dependency of data
3. Pooled Cross Sections
   - Combination of cross sections to increase sample size
   - Random samples → "same population"
4. Panel or Longitudinal Data
   - Time series for each member in cross-sectional data set
   - Data on the same units across time
   - Allows estimation of effect over time + new methods (e.g. FE)

Random variables

- Random variable are usually denoted "X"
- Values of X are numerical outcomes of random phenomena
  - X = **set of possible values** from a random experiment
  - x = **particular outcome** of a random experiment
    - If we roll a dice 10 times, we have $X_i = \{1, 2, 3, 4, 5, 6\}$
    - The realized throws might be: $x_1 = 2, x_2 = 4, ..., x_6 = 6$
- Each element of X can have a different probability
  - $P(X = value)$ = probability of a certain value
    - Dice: $P(X = x_i) = \frac{1}{6}$, with $x_i = 1, 2, ..., 6$
  - Probability that X takes on a range of values
    - X is lower than 3 $\rightarrow P(1 \leq X < 3)$
    - $P(1 \leq X < 3) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$
- "Events" = set of outcomes (e.g. all outcomes with $X > 2$)

Discrete vs. continuous

- Two types of random variables
  1. Discrete random variables $\rightarrow$ **Probability mass function**
     - Takes a countable number of distinct values (0,1,2,3,...)
     - Most often, discrete random variables are counts
     - Number of attendants in a tutorial, inhabitants,...
  2. Continuous random variables $\rightarrow$ **Probability density function**
     - Takes an infinite number of possible values
       $\rightarrow$ Probability of a single value is 0!
       $\rightarrow$ Not defined at specific values $\rightarrow$ interval of values!
     - Represented by area under curve (integral)
     - Height, weight, time,...

Probability mass function

- Probability distribution of a discrete random variable
- List of probabilities associated with the possible values of X
  - E.g. $P(X = x_i) = \frac{1}{6}, i = 1, 2, ..., 6$
  - Or:

Outcome	Probability
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

Table 1: pmf of a dice throw

Probability density function

- Probability distribution of a continuous random variable
- Density curve represents pdf
  - Curve must not have negative values
  - Total area under the curve (= probability) is equal to 1
- Most important continuous pdf is normal distribution
  - $P(X{=}x) \rightarrow P(Z{=}z)$
  - E.g. what is the probability that Z is larger than 1.96?

Cumulative distribution function

- CDF gives the probability that a value is equal to x or smaller
  - $F(x) = P(X \leq x)$
- Probability is given by the sum of all probabilities (discrete var.) or the integral under the curve for all values x_j that are smaller or equal to x
  - Probability that a value is larger than x: $1 - F(x)$
- CDF is bound between 0 and 1
- PDF is the derivative of the CDF: $F'(x) = f(x)$

Binomial Distribution

Binomial Experiment

- Experiment consists of n repeated trials
- Each trial can result in just two possible outcomes ($\rightarrow$ "bi")
- Probability of success is the same on every trial
- Outcome on one trial does not affect the outcome on other trials $\rightarrow$ independence
- Binomial experiment = repeated Bernoulli experiment
  - Recall: Bernoulli random variable is binary

Binomial distribution

- Probability distribution of binomial random variables
  - Binomial random variables = number of successes x in n repeated trials of a binomial experiment
- Suppose a coin is flipped 2 times, $X = \{0, 1, 2\}$
  - Probability of 0 heads:
    $P(X = 0) = b(0, 2, 0.5) = \binom{2}{0} * 0.5^0 * 0.5^{2-0} = 0.25$

Successes	Probability
0	0.25
1	0.50
2	0.25

Table 2: Binomial distribution of 2 coin flips

Binomial distribution

How likely are 60 heads after 100 coin tosses?

$\rightarrow P(X = 60) = b(60, 100, 0.5) = \binom{100}{60} * 0.5^{60} * 0.5^{40} = 0.012$

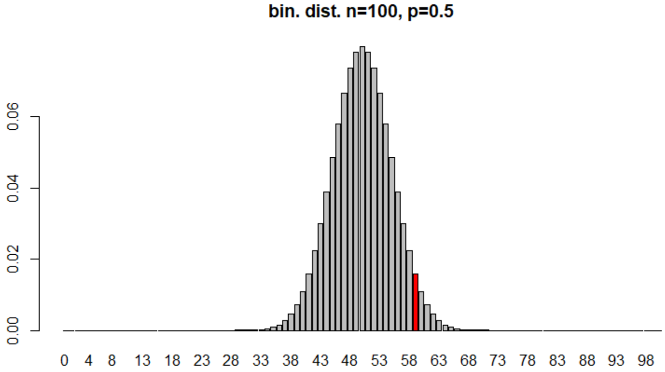


Figure 1: Binomial distribution with n=100 and p=0.5

Joint distributions

- We no longer look at the distribution of only one variable
  - Number of facebook friends and age
  - Level of air pollution and rate of respiratory illness in cities
  - Cities and housing prices
- With 2+ random variables, we can study joint distributions
  - Discrete X and Y: joint probability mass function
  - Continuous X: joint probability density function
  - Discrete X and continuous Y: more complicated

Joint probability mass function

- Probability of the joint outcome X=x and Y=y
- $f_{X,Y}(x,y) = P(X = x, Y = y)$
- Joint pmf of two coin tosses is $p(i,j) = 0.25$

X_i/Y_j	0	1
0	P(0,0) = 0.25	P(0,1) = 0.25
1	P(1,0) = 0.25	P(1,1) = 0.25

- Criteria for joint pmf:
  1. $(0 \leq p(x_i, y_j) \leq 1)$
  2. Total probability $= 1 \rightarrow \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) = 1$

$\rightarrow$ Wooldrige does not distinguish between pmf and pdf

Joint probability density function

- Replace discrete set of values, e.g. $X = \{0, 1\}$, by continuous intervals and use integrals instead of sums
- Suppose X is in an interval [a,b] and Y is in an interval [c,d]
  - Pair (X, Y) takes values in the rectangle [a,b] * [c,d]
  - Joint pdf $f(x, y)$ gives probability density at (x,y)
  - $f(x, y)dxdy =$ probability that (X, Y) is in a rectangle of width dx and height dy around (x, y)
- Criteria for joint pdf:
  1. $0 \leq f(x, y)$ (density not smaller than 0)
  2. Total probability is 1 $\rightarrow \int_c^d \int_a^b f(x, y)dxdy = 1$

Independence

- Joint pmf/pdf is easily computed if X and Y are independent:
  - $f_{X,Y} = f_X(x) * f_Y(y)$ **iff** $X \perp\!\!\!\perp Y$ (X and Y are independent)
    $\rightarrow$ Holds for discrete + continuous random variables
  - f_X and f_Y are called "marginal probability densities"
- If $X \perp\!\!\!\perp Y$, knowing the outcome of X does not change the probabilities of Y
- Functions of independent X and Y are also independent
  - If $X \perp\!\!\!\perp Y$, $(3X) \perp\!\!\!\perp (log(Y) + 2)$
- What is the probability that it's beergarden wheather and that there is no statistics lesson? (They are independent)
  - Joint probability = product of probabilities
  - $P(W = 1, S = 0) = P(W = 1) * P(S = 0)$

Conditional distributions

- Conditional probability density function summarizes information how X affects Y
  - $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$
  - Conditional probability = joint probability/marginal probability
  - "$Y|X$" = "Y given X" or "Y conditional on X"
- Probability of right answer after right/wrong answer?
  - $f_{X|Y}(1|1) = 0.70$
  - $f_{X|Y}(1|0) = 0.45$
- If $X \perp\!\!\!\perp Y$, $f_{Y|X}(y|x) = f_Y(y)$ & $f_{X|Y}(x|y) = f_X(x)$
  - X tells us nothing about the probability of Y's outcomes
  - The fact that we got "heads" in the first toss (X) tells us nothing about the probability of heads in a second toss (Y)

In a Nutshell

1. Marginal probability
   - Probability of $X = x$
   - Probability that person is female
     $P(X = 1)$
2. Joint probability
   - Probability of $X = x$ and $Y = y$
   - Probability that one is female and the answer is right
     $P(X = 1, Y = 1) = P(X = 1) * P(Y = 1)$
3. Conditional probability
   - Probability of $Y = y$ given $X = x$
   - Probability that answer is right if a person is female
     $P(Y = 1|X = 1) = \frac{P(X=1, Y=1)}{P(X=1)}$

Whats E(X)?

- E(X)= weighted average of all possible values of X
  - Weights are determined by pdf
  - $E(X) =$ sum of probabilities of outcomes * outcomes
  - $E(X) = \sum_{j=1}^{k} x_j * f(x_j)$
  - If X is continuous, we use the integral again
    - $E(X) = \int_{-\infty}^{\infty} x * f(x)dx$
- For functions of X, this works analogously (Wool. Ex. B4)
- Why is a random variable X even random?
  - Recall: $X = \mu + \epsilon$
  - μ is fixed, but the error ϵ is random and varies!
  - Don't confuse notation and the specific cases!
    - In regressions, Y is a random variable (depends on X and ϵ)
    - X comes from data and is often assumed to be deterministic

Fun with E(X)

- Constants do not vary, therefore $E(c) = c$
  $\rightarrow \frac{1}{5} * (5 + 5 + 5 + 5 + 5) = 5$
- With $\times$ and $\div$ we can take out constants as the expected value of the constant is the constant itself
  - $E(aX) = a * E(X)$
- $\pm$ allow us to expand and factorise the expected value
  - $E(aX + b) = E(aX) + E(b) = a * E(X) + b$
  - $E(X) + E(Y) = E(X + Y)$
- This also holds for sums:
  - Expected value of a sum is sum of the expected value
  - $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$
  - $E\left(\sum_{i=1}^{n} a * X_i\right) = n * a * \sum_{i=1}^{n} E(X_i) \rightarrow a + a + ... + a = n * a$

What is Variance?

- Information about the spread of distributions
  - How far is X away from its expected value?
  - Expected distrance from X to its mean: $\sigma_X^2 = Var(X) = E\left[(X-\mu)^2\right]$
  - Solving $Var(X)$ yields: $Var(X) = E(X^2) - \mu^2$
- $Var(X) = 0$ **iff** $P(X = c) = 1$ and $E(X) = c$
  $\rightarrow$ If X is a constant, $E(X) = c$ and $Var(X) = c^2 - c^2 = 0$
  $\rightarrow$ Variance of a constant is 0 since constants do not vary
- $Var(aX + b) = a^2 Var(X)$
  - Adding a constant does not change variance of X
  - We can take a constant out of the variance by squaring it
    $\rightarrow Var(2X) = 2^2 * Var(X) = 4 * Var(X)$

Fun with $V(X)$

- $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
- If there is no relationship between X and Y ($\rightarrow X \perp\!\!\!\perp Y$):
  $Var(X + Y) = Var(X) + Var(Y)$, since $Cov(X, Y) = 0$
- Even if $Var(X - Y)$, we have $Var(X) + Var(Y)$
  - $Var(X - bY) = Var(X) + (-b)^2 Var(Y) = Var(X) + b^2 Var(Y)$

Fun with $V(X)$ II

- Formula $Var(X) = E(X^2) - \mu^2$ can simplify computation
  - Remember $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
  - For binary dependent variables ($X = \{0, 1\}$), $E(X^2) = E(X)$!
    $\rightarrow 0^2 = 0$ and $1^2 = 1$
  - Most often, this applies for:
    - Bernoulli and binomial
    - Logit and probit

Standard deviation

- Positive square root of variance
- $sd(X) = \sqrt{Var(X)}$
- Same properties as Var(X):
  1. $sd(c) = 0$
     $\rightarrow$ Standard deviation of constant $= 0$
  2. $sd(aX + b) = |a| * sd(X)$
     - $sd(aX) + sd(b) = sd(aX) = |a| * sd(X)$
       $\rightarrow sd(2X) = 2 * sd(X) = sd(-2X)$
       $\rightarrow |a|$ since sd(X) is positive square root of Var(x)

Covariance

- Summary measure of joint distribution of X and Y
  $\rightarrow$ How do X and Y vary with on another, on average?
  - Do variables move in the same or opposite directions?
  - Amount of linear dependence between two random variables
    - Lin. dependent if Y can be written as linear combination of X
      $\rightarrow$ If $Y = (1, 2, 3)$ and $X = (2, 4, 6)$ then $Y = 0.5X$
      $\rightarrow$ $Y = b_1 + b_2^2 X$ is no linear combination
- $Cov(X, Y) = E(X - \mu_x)(Y - \mu_Y) = E(XY) - \mu_X \mu_Y$
  - $Cov(X, Y) = Cov(Y, X)$
- Direction easy to interpret, magnitude depends on scale
  - Covariance changes if we use Kilometers instead of meters
    $\rightarrow$ Advantage of correlation coefficient!

Properties of covariance

1. If $X \perp\!\!\!\perp Y, Cov(X, Y) = 0$
   - If X and Y are independent, the covariance is 0
   - For constants a_i and b_i, $Cov(a_1X + b1, a_2Y + b_2) = a_1a_2Cov(X, Y)$
     $\rightarrow$ Adding a constant to one/both r.v. does not change Cov
     $\rightarrow$ If we use $X^* = 1000X$, $Cov(X^*, Y) = 1000 * Cov(X, Y)$
   - $|Cov(X, Y)| \leq sd(X) * sd(Y)$
     $\rightarrow$ Absolute covariance cannot be larger than product of SDs

Correlation

- Describes relationship between two variables X and Y
  - $\rho_{X,Y} = Corr(X,Y) = \frac{Cov(X,Y)}{sd(X)*sd(Y)}$
- Properties:
  1. $-1 \leq Corr(X,Y) \leq 1$
     - No dependence on measurement of X and Y
     - Corr is only negative if Cov is negative (sd always positive)
  2. $Corr(a_1X + b_1, a_2Y + b_2) = +Corr(X,Y)$, if $a_1a_2 > 0$
     $Corr(a_1X + b_1, a_2Y + b_2) = -Corr(X,Y)$, if $a_1a_2 < 0$
     - Correlation coefficient of X and Y is not changed by constants

Exercise 1a

Is X a continuous random variable?

- Does the result consist of countable numbers?
- Yes! Therefore it is a discrete random variable and not a continuous variable!

Exercise 1b and 1c

Calculate P(X=3)

- “What is the probability, that we throw with one shot a 3?”
- $P(X{=}3) = 1/6 = 0.167$

Calculate P(X=7)

- “What is the probability, that we throw with one shot a 7?
- Not possible, because our dice only includes only the numbers 1-6.
- $P(X{=}7) = 0$

Exercise 1d

Calculate E(X)

- $E(X) =$ sum of probabilities of outcomes * outcomes
- $E(X) = \sum_{j=1}^{k} x_j * f(x_j)$
- E(X)= 1/6 * 1 + 1/6 * 2 + ... 1/6 * 6 = 3.5

Exercise 1e and 1f

Suppose that a friend doubles the number of your dice and gives the corresponding amount of money in Euro. How much Euro do you expect in average to receive your friend?

- $E(X) = 2 * E(X) = 2 * 3.5 = 7$

You throw another dice with 8 sides, which is not manipulated ,as well. Denote the result by Y. Calculate $P(X = 3, Y = 7)$

- Are X and Y independent from each other? Yes? Then we can easily compute $P(X = 3, Y = 7)$
- $P(X = 3, Y = 7) = P(X = 3) * P(Y = 7) = 1/6 * 1/8 = 0.02083$

Exercise 1g and 1h

Calculate $P(X = 7, Y = 7)$

- Not possible, because our dice only includes only the numbers 1-6.
- $P(X = 7, Y = 7) = 0$

Calculate $E(X + Y)$

- Note that $E(X + Y) = E(X) + E(Y)$
- $E(X+Y) = (1/6*(1+2+\ldots.6))+(1/8*(1+2+\ldots.8)) = 3.5+4.5 = 8$

Exercise 1i and 1j

Calculate $E(2X + Y)$

- Note that $E(2X + Y) = 2 * E(X) + E(Y)$
- $E(2X + Y) = 2 * 3.5 + 4.5 = 11.5$

Are X and Y independent ? Answer with yes or no

- Has X any influence on the result of Y or other way round?
- In other words: “Does it matter for our eight-sided dice, what we throw with our six-sided dice?
- No it does not! Therefore, X and Y are independent from each other

Exercise 1k and 1m (I switched the order of m and l)

You throw the above dice with 8 sides two more times. That is, you throw the dice 3 times in total. Denote the frequency that you obtain the number 7 by Z. Give all the possible values of Z

- $Z = \{0, 1, 2, 3\}$
- In words: "If we throw our dice 3 times in total, we could possibly obtain the number 7 none, one, two or three times."

What distribution does Z have?

- Two possible results: "7" or "not 7"
- Therefore: Binomial distribution

Exercise 1l and 1n (I switched the order of m and l)

Calculate $P(Z = 3)$

- Use the Binomial function: $\binom{n}{k} * x^k * (1 - x)^{n-k}$
- $P(Z = 3) = \binom{3}{3} * 0.125^3 * (1 - 0.125)^{3-3} = 1 * 0.125^3 = 0.001953125$

Calculate $P_{Z|Y}(3|7)$

- The first throw of the dice Y is already given and it was a 7. Therefore we can calculate the probability of obtaining a 7 with all three throws with the conditional probability density function.
- $P_{Z|Y}(3|7) = \frac{f_{Z,Y}(z,y)}{f_X(y)}$
- $f_{Z,Y}(z, y) = \frac{1}{512} = 0.001953125$
- $f_X(y) = 0.125$
- $P_{Z|Y}(3|7) = \frac{\frac{1}{512}}{0.125} = 0.015625$

Exercise 2a and 2b

Calculate $P(X = 2000)$

- $P(X = 2000) = 0$
- Continuous random variables take on an infinite number of possible values $\rightarrow$ Probability of a single value is 0!

Calculate $P(X > 2000)$

- $P(X > 2000) = 1 - F(2000) = 1 - 0.7 = 0.3$
- where $F()$ is the cdf of X
- the cdf gives the probability that a value is equal to x or smaller

Exercise 2c

What does cdf stand for?

- cumulative distribution function (not density!)
- cdf gives the probability that a value is equal to x or smaller
  - $F(x) = P(X \leq x)$
- Probability is given by the sum of all probabilities (discrete var.) or the integral under the curve for all values x_j that are smaller or equal to x
  - Probability that a value is larger than x: $1 - F(x)$
- cdf is bound between 0 and 1

Exercise 2d and 2e

Denote the derivative of $F()$ by $f()$. What is $f()$?

- probability density function
- pdf is the derivative of the cdf: $F'(x) = f(x)$

Calculate $\int_{-\infty}^{\infty} f(x)dx$

- $\int_{-\infty}^{\infty} f(x)dx = 1$
- The pdf summarizes the information about possible outcomes of random variable X and their corresponding probabilities.
  - Sum of all probabilities must be 1
  - Normal distribution is a pdf (for continuous random variables!)

Exercise 2f and 2g

Calculate $\int_{-\infty}^{\infty} xf(x)dx$

- $\int_{-\infty}^{\infty} xf(x)dx = 1000$
- X is a continuous random variable, therefore $E(X)$ is defined as $\int_{-\infty}^{\infty} xf(x)dx$
  - $E(X) = \int_{-\infty}^{\infty} xf(x)dx$
- From the exercise we have $E(X) = 1000$

In this country, all citizens have to pay a tax of 30% of income. How much tax did an average citizen pay in the last year?

- $E(tax) = 12 * 0.3 * E(X)$
- $E(tax) = 12 * 0.3 * 1000 = 3600$

Exercise 2h and 2i

Calculate $E(X+Y)$

- $E(X+Y) = E(X) + E(Y) = 1000 + 800 = 1800$
- the expected value of the sum is the sum of expected values

Calculate $Var(X+Y)$

- $Var(X+Y) = Var(X) + Var(Y) + 2 * Cov(X,Y)$
- $Var(X+Y) = 500000 + 400000 + 2 * 100000 = 1100000$

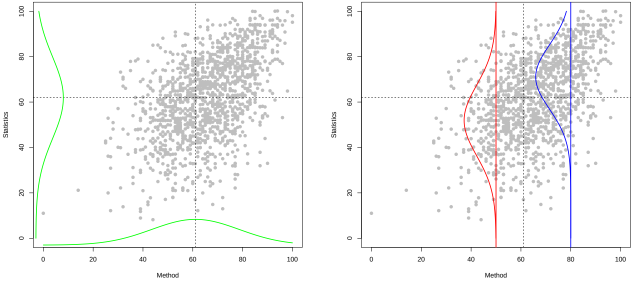
Exercise 2j

Calculate $Corr(X+Y)$

- $Corr(X+Y) = \frac{Cov(X,Y)}{sd(X)*sd(Y)}$
- $sd(X) = \sqrt{Var(X)} = \sqrt{500000}$
- $sd(Y) = \sqrt{Var(Y)} = \sqrt{500000}$
- $Cov(X,Y) = 100000$
- $Corr(X+Y) = \frac{100000}{\sqrt{500000}*\sqrt{500000}} = 0.2236068$

Exercise 3

Where do you find the three types of distributions in the figures?



- Marginal: "The distribution of single variables" $\rightarrow$ Green
- Conditional: "The distribution of one variable, given an value of another" $\rightarrow$ Red & Blue
- Joint: "The probability distribution for two (or more) variables, in order to look for a relationship between them." $\rightarrow$ Grey dots