Econometrics in Economics / Introduction to Econometrics

Refresher on stats and probability

based on Wooldridge (2019), App A-C; Stock and Watson (2020), Ch. 2–3

Patrick Schmidt (based on slides by **Simon Heß** and **Daniel Gutknecht**)

Winter 23

Road map

last part:

- what is econometrics?
- example: student test performance and class size
- causality and experimental vs observational data

this part:

- refresher on statistics and probability

The landscape of statistics? Where are we?

Informal definition: Statistics

Statistics encompasses the collection, analysis, and interpretation of data

The landscape of statistics? Where are we?

Informal definition: Statistics

Statistics encompasses the collection, analysis, and interpretation of data

descriptive statistics:

- condensing and communicating information from data (visually or numerically)

The landscape of statistics? Where are we?

Informal definition: Statistics

Statistics encompasses the collection, analysis, and interpretation of data

descriptive statistics:

- condensing and communicating information from data (visually or numerically)

inferential statistics / statistical inference:

- use data analyses to learn about (to "infer") properties of the underlying probability distribution

The landscape of statistics? Where are we?

Informal definition: Statistics

Statistics encompasses the collection, analysis, and interpretation of data

descriptive statistics:

- condensing and communicating information from data (visually or numerically)

inferential statistics / statistical inference:

- use data analyses to learn about (to “infer”) properties of the underlying probability distribution
  - key aspect: Uncertainty quantification.

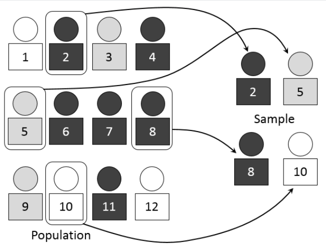
mathematical statistics:

- the tools (probability theory) to carry out statistical analyses

This course's focus: inferential statistics using samples of data

- start from an **economic model**: **unobserved** parameters of interest $\mapsto$ **observables** (data)
  - solving the model for the parameters gives us a way to estimate it

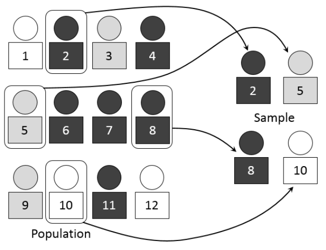
1. can we estimate the **mean wage** in the **population**?



This course's focus: inferential statistics using samples of data

- start from an **economic model**: **unobserved** parameters of interest $\mapsto$ **observables** (data)
  - solving the model for the parameters gives us a way to estimate it
- data is rarely complete: have to rely on **samples**

1. can we estimate the **mean wage** in the **population**?

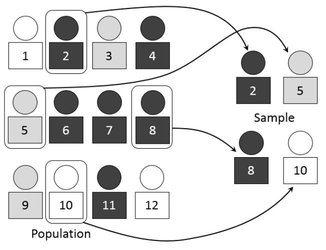


This course's focus: inferential statistics using samples of data

- start from an **economic model**: **unobserved** parameters of interest $\mapsto$ **observables** (data)
  - solving the model for the parameters gives us a way to estimate it
- data is rarely complete: have to rely on **samples**

Ex.: We observe wages of 1000 individuals living in Germany. What can we infer about the population (= all inidividuals living in Germany)?

1. can we estimate the **mean wage** in the **population**?

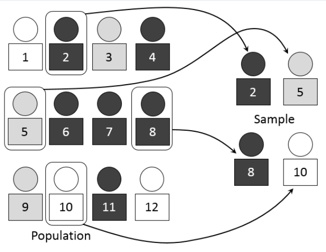


This course's focus: inferential statistics using samples of data

- start from an **economic model**: **unobserved** parameters of interest $\mapsto$ **observables** (data)
  - solving the model for the parameters gives us a way to estimate it
- data is rarely complete: have to rely on **samples**

Ex.: We observe wages of 1000 individuals living in Germany. What can we infer about the population (= all inidividuals living in Germany)?

1. can we estimate the **mean wage** in the **population**?
   - yes, if the sample is **representative**

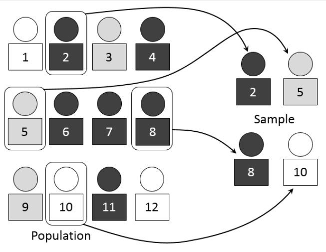


This course's focus: inferential statistics using samples of data

- start from an **economic model**: **unobserved** parameters of interest $\mapsto$ **observables** (data)
  - solving the model for the parameters gives us a way to estimate it
- data is rarely complete: have to rely on **samples**

Ex.: We observe wages of 1000 individuals living in Germany. What can we infer about the population (= all inidividuals living in Germany)?

1. can we estimate the **mean wage** in the **population**?
   - yes, if the sample is **representative**
2. can we study **wage diff's** (skills, gender)?

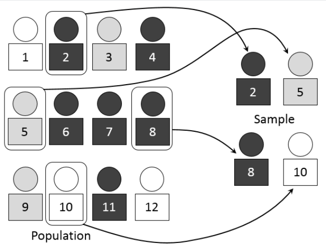


This course's focus: inferential statistics using samples of data

- start from an **economic model**: **unobserved** parameters of interest $\mapsto$ **observables** (data)
  - solving the model for the parameters gives us a way to estimate it
- data is rarely complete: have to rely on **samples**

Ex.: We observe wages of 1000 individuals living in Germany. What can we infer about the population (= all inidividuals living in Germany)?

1. can we estimate the **mean wage** in the **population**?
   - yes, if the sample is **representative**
2. can we study **wage diff's** (skills, gender)?
   - yes, if the sample is representative for each subgroup

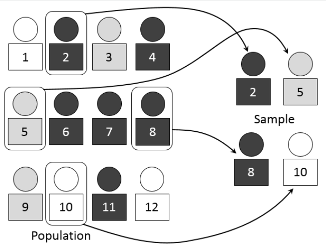


This course's focus: inferential statistics using samples of data

- start from an **economic model**: **unobserved** parameters of interest $\mapsto$ **observables** (data)
  - solving the model for the parameters gives us a way to estimate it
- data is rarely complete: have to rely on **samples**

Ex.: We observe wages of 1000 individuals living in Germany. What can we infer about the population (= all inidividuals living in Germany)?

1. can we estimate the **mean wage** in the **population**?
   - yes, if the sample is **representative**
2. can we study **wage diff's** (skills, gender)?
   - yes, if the sample is representative for each subgroup
3. **how large is the difference** between our sample and the population?

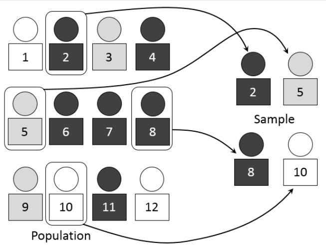


This course's focus: inferential statistics using samples of data

- start from an **economic model**: **unobserved** parameters of interest $\mapsto$ **observables** (data)
  - solving the model for the parameters gives us a way to estimate it
- data is rarely complete: have to rely on **samples**

Ex.: We observe wages of 1000 individuals living in Germany. What can we infer about the population (= all inidividuals living in Germany)?

1. can we estimate the **mean wage** in the **population**?
   - yes, if the sample is **representative**
2. can we study **wage diff's** (skills, gender)?
   - yes, if the sample is representative for each subgroup
3. **how large is the difference** between our sample and the population?
   - quantifying uncertainty is possible and important



Overview: Sampling, estimation, inference, identification

Overview: Sampling, estimation, inference, identification

sample/data

observed

population

observable
but usually unobserved

underlying structure

unobservable

Overview: Sampling, estimation, inference, identification

sample/data

observed

population

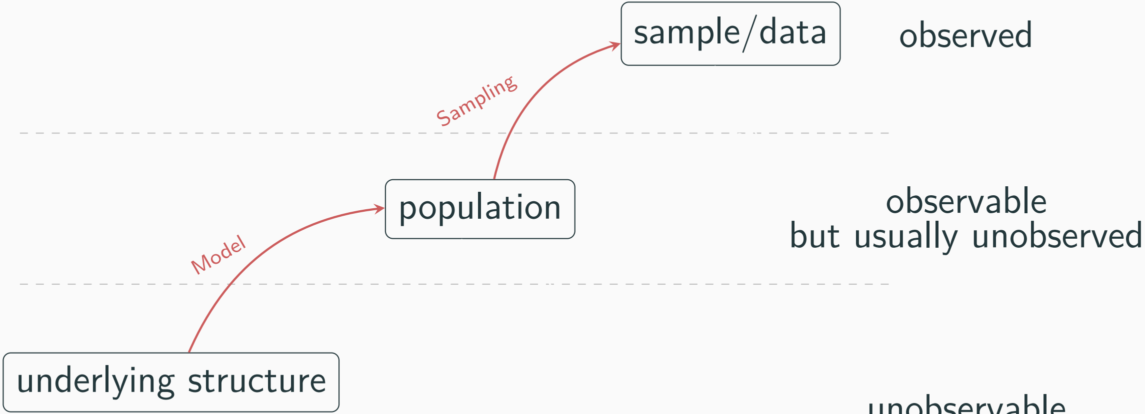
observable
but usually unobserved

Model

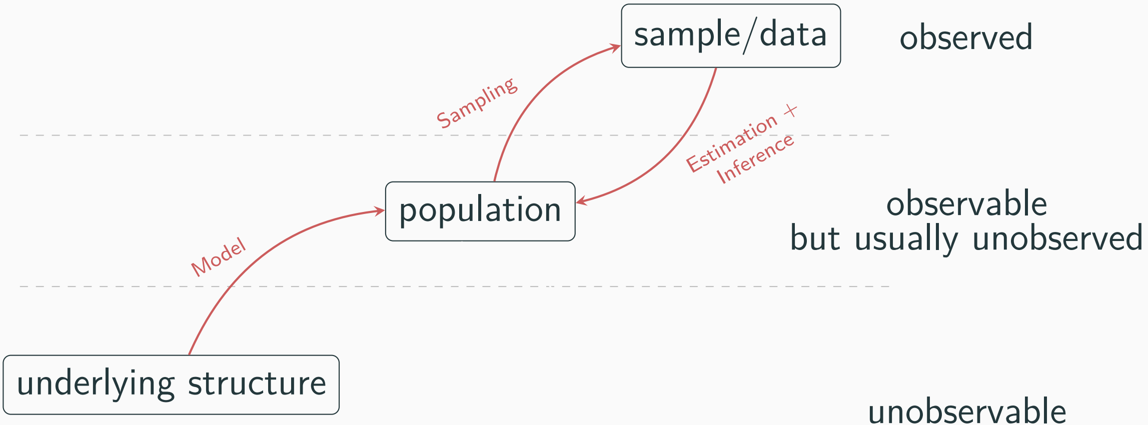
underlying structure

unobservable

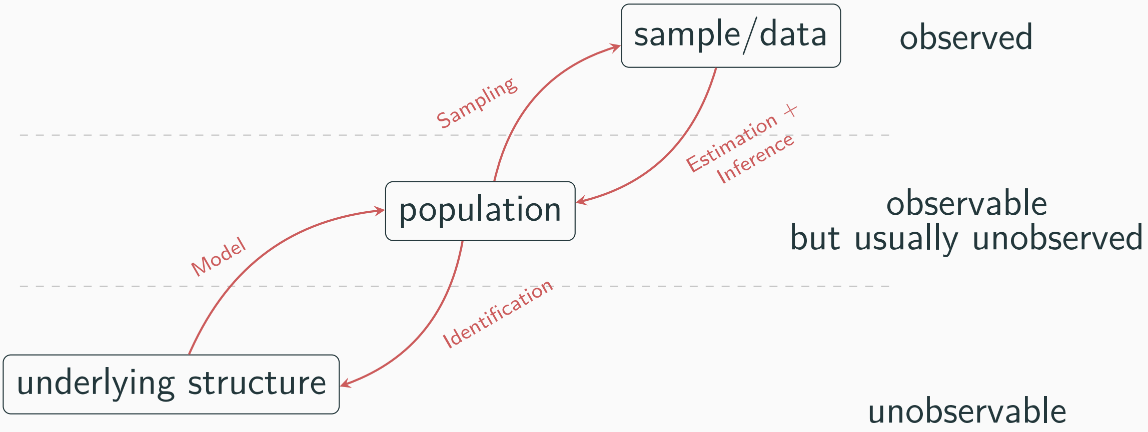
Overview: Sampling, estimation, inference, identification



Overview: Sampling, estimation, inference, identification



Overview: Sampling, estimation, inference, identification



Terminology

Informal definition: Estimator/Estimate/Estimand

- The parameter to be estimated in a population is called the **estimand**

- **Estimand**: $\mathbb{E}[x]$:
  - e.g., mean salary of Goethe University alumna

  Proposed **estimator**: the average defined as $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

  Given three observed values with 10, 20, and 30, the **estimate** is $\frac{1}{3}(10 + 20 + 30) = 20$

Terminology

Informal definition: Estimator/Estimate/Estimand

- The parameter to be estimated in a population is called the **estimand**
- The **estimator** is a rule/formula that strives to estimate the estimand

- **Estimand**: $\mathbb{E}[x]$:
  - e.g., mean salary of Goethe University alumna

  Proposed **estimator**: the average defined as $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

  Given three observed values with 10, 20, and 30, the **estimate** is $\frac{1}{3}(10 + 20 + 30) = 20$

Terminology

Informal definition: Estimator/Estimate/Estimand

- The parameter to be estimated in a population is called the **estimand**
- The **estimator** is a rule/formula that strives to estimate the estimand
- Given a data sample, one can compute the estimator. The resulting value is called the **estimate**.

- **Estimand**: $\mathbb{E}[x]$:
  - e.g., mean salary of Goethe University alumna

  Proposed **estimator**: the average defined as $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

  Given three observed values with 10, 20, and 30, the **estimate** is $\frac{1}{3}(10 + 20 + 30) = 20$

Terminology

Informal definition: Estimator/Estimate/Estimand

- The parameter to be estimated in a population is called the **estimand**
- The **estimator** is a rule/formula that strives to estimate the estimand
- Given a data sample, one can compute the estimator. The resulting value is called the **estimate**.

- **Estimand**: $\mathbb{E}[x]$:
  - e.g., mean salary of Goethe University alumna

Proposed **estimator**: the average defined as $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Given three observed values with 10, 20, and 30, the **estimate** is $\frac{1}{3}(10+20+30) = 20$

Terminology

Informal definition: Estimator/Estimate/Estimand

- The parameter to be estimated in a population is called the **estimand**
- The **estimator** is a rule/formula that strives to estimate the estimand
- Given a data sample, one can compute the estimator. The resulting value is called the **estimate**.

Ex.: Estimate the mean of a variable x.

- **Estimand**: $\mathbb{E}[x]$:
  - e.g., mean salary of Goethe University alumna

Terminology

Informal definition: Estimator/Estimate/Estimand

- The parameter to be estimated in a population is called the **estimand**
- The **estimator** is a rule/formula that strives to estimate the estimand
- Given a data sample, one can compute the estimator. The resulting value is called the **estimate**.

Ex.: Estimate the mean of a variable x.

- **Estimand**: $\mathbb{E}[x]$:
  - e.g., mean salary of Goethe University alumna
- Proposed **estimator**: the average defined as $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
  - Note: The estimator is a formula

Terminology

Informal definition: Estimator/Estimate/Estimand

- The parameter to be estimated in a population is called the **estimand**
- The **estimator** is a rule/formula that strives to estimate the estimand
- Given a data sample, one can compute the estimator. The resulting value is called the **estimate**.

Ex.: Estimate the mean of a variable x.

- **Estimand**: $\mathbb{E}[x]$:
  - e.g., mean salary of Goethe University alumna
- Proposed **estimator**: the average defined as $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
  - Note: The estimator is a formula
- Given three observed values with 10, 20, and 30, the **estimate** is $\frac{1}{3}(10 + 20 + 30) = 20$
  - The estimate is a number: 20

Terminology

Informal definition: Estimator/Estimate/Estimand

- The parameter to be estimated in a population is called the **estimand**
- The **estimator** is a rule/formula that strives to estimate the estimand
- Given a data sample, one can compute the estimator. The resulting value is called the **estimate**.

Ex.: Estimate the mean of a variable x.

- **Estimand**: $\mathbb{E}[x]$:
  - e.g., mean salary of Goethe University alumna
- Proposed **estimator**: the average defined as $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
  - Note: The estimator is a formula
- Given three observed values with 10, 20, and 30, the **estimate** is $\frac{1}{3}(10 + 20 + 30) = 20$
  - The estimate is a number: 20

What are these concepts for?

- The economic questions from the motivation are estimands:
    - The gender pay gap, the effect of class size on test scores, etc.

What are these concepts for?

- The economic questions from the motivation are estimands:
  - The gender pay gap, the effect of class size on test scores, etc.
- In this lecture we are going to learn about different estimators for those estimands
  - And about their properties: How good are they for estimating the estimand?
- And given data, we are computing estimates to inform us

What are these concepts for?

- The economic questions from the motivation are estimands:
  - The gender pay gap, the effect of class size on test scores, etc.
- In this lecture we are going to learn about different estimators for those estimands
  - And about their properties: How good are they for estimating the estimand?
- And given data, we are computing estimates to inform us
- But first: Estimators are random variables. So, let's recap.

What are these concepts for?

- The economic questions from the motivation are estimands:
  - The gender pay gap, the effect of class size on test scores, etc.
- In this lecture we are going to learn about different estimators for those estimands
  - And about their properties: How good are they for estimating the estimand?
- And given data, we are computing estimates to inform us
- But first: Estimators are random variables. So, let's recap.

Random variables

Definitions

Sample space: The set of all possible outcomes, typically denoted by Ω

Definitions

Sample space: The set of all possible outcomes, typically denoted by Ω

Ex.: Age of a random person

$\Omega^{\text{age}} = \{0, 1, 2 \ldots\} = \mathbb{N}$

Event: A subset of the sample space

Definitions

Sample space: The set of all possible outcomes, typically denoted by Ω

Ex.: Age of a random person

$\Omega^{\text{age}} = \{0, 1, 2 \dots\} = \mathbb{N}$

Event: A subset of the sample space

Ex.: Age of a random person

$\omega_1 = \{18\}$, $\omega_2 = \{1,2,3,4\}$

Random variable (RV): A function that assigns a value to each element of the sample space

Definitions

Sample space: The set of all possible outcomes, typically denoted by Ω

Ex.: Age of a random person

$\Omega^{\text{age}} = \{0, 1, 2 \ldots\} = \mathbb{N}$

Event: A subset of the sample space

Ex.: Age of a random person

$\omega_1 = \{18\}$, $\omega_2 = \{1,2,3,4\}$

Random variable (RV): A function that assigns a value to each element of the sample space

Ex.: Age of a random person

Random variable: The age in years of the first person you see after leaving this class

Remarks

- sampled data points are RVs

Remarks

- sampled data points are RVs
  - e.g., the income of a randomly sampled person in Frankfurt

Remarks

- sampled data points are RVs
  - e.g., the income of a randomly sampled person in Frankfurt
- any aggregation of RVs is also a RV: E.g. the average age of the next 10 people you see is a RV

Remarks

- sampled data points are RVs
  - e.g., the income of a randomly sampled person in Frankfurt
- any aggregation of RVs is also a RV: E.g. the average age of the next 10 people you see is a RV
  - estimators we discuss (e.g. $\frac{1}{n}\sum_i y_i$) are aggregations of data, which can be treated as RVs

Remarks

- sampled data points are RVs
  - e.g., the income of a randomly sampled person in Frankfurt
- any aggregation of RVs is also a RV: E.g. the average age of the next 10 people you see is a RV
  - estimators we discuss (e.g. $\frac{1}{n}\sum_i y_i$) are aggregations of data, which can be treated as RVs

The distribution of random variables

Discrete RV: y can take on finitely or countably many values of $\mathbb{R}$.

The distribution of random variables

Discrete RV: y can take on finitely or countably many values of $\mathbb{R}$.

- probabilities of y taking any specific value is written as

  $$\Pr(y = \text{value})$$

  where all probabilities are between 0 and 1 and they sum up to one

The distribution of random variables

Discrete RV: y can take on finitely or countably many values of $\mathbb{R}$.

- probabilities of y taking any specific value is written as

$$\Pr(y = \text{value})$$

  where all probabilities are between 0 and 1 and they sum up to one

Ex.: y represent the outcome of a fair die, 🎲

$\Pr(y = j) = 1/6 \quad$ for $\quad j \in \{1, \dots, 6\}$
$\sum_{i=j}^{6} \Pr(y = j) = 1.$

The distribution of random variables

Discrete RV: y can take on finitely or countably many values of $\mathbb{R}$.

- probabilities of y taking any specific value is written as

$$\Pr(y = \text{value})$$

where all probabilities are between 0 and 1 and they sum up to one

Ex.: y represent the outcome of a fair die, ⚂

$\Pr(y = j) = 1/6 \quad \text{for} \quad j \in \{1, \dots, 6\}$
$\sum_{i=j}^{6} \Pr(y = j) = 1.$

Continuous RV: y can take on all values from a subset of the real line $\mathbb{R}$

The distribution of random variables

Discrete RV: y can take on finitely or countably many values of $\mathbb{R}$.

- probabilities of y taking any specific value is written as
$$\Pr(y = \text{value})$$
where all probabilities are between 0 and 1 and they sum up to one

Ex.: y represent the outcome of a fair die,

$\Pr(y = j) = 1/6 \quad \text{for} \quad j \in \{1, \ldots, 6\}$
$\sum_{i=j}^{6} \Pr(y = j) = 1.$

Continuous RV: y can take on all values from a subset of the real line $\mathbb{R}$

- probability of it taking any specific value
$$\Pr(y = \text{value})$$
is always zero and not a useful description

The distribution of random variables

Discrete RV: y can take on finitely or countably many values of $\mathbb{R}$.

- probabilities of y taking any specific value is written as

$$\Pr(y = \text{value})$$

  where all probabilities are between 0 and 1 and they sum up to one

Ex.: y represent the outcome of a fair die,

$\Pr(y = j) = 1/6 \quad \text{for} \quad j \in \{1, \ldots, 6\}$
$\sum_{i=j}^{6} \Pr(y = j) = 1.$

Continuous RV: y can take on all values from a subset of the real line $\mathbb{R}$

- probability of it taking any specific value

$$\Pr(y = \text{value})$$

  is always zero and not a useful description

Ex.: y represents the time it takes to run the marathon distance

The probability $\Pr(y = t) = 0$ for $t = 2\text{h}$ and for any $t \in \mathbb{R}$.

The distribution of random variables

Discrete RV: y can take on finitely or countably many values of $\mathbb{R}$.

- probabilities of y taking any specific value is written as

$$\Pr(y = \text{value})$$

  where all probabilities are between 0 and 1 and they sum up to one

Ex.: y represent the outcome of a fair die,

$\Pr(y = j) = 1/6 \quad \text{for} \quad j \in \{1, \ldots, 6\}$
$\sum_{i=j}^{6} \Pr(y = j) = 1.$

Continuous RV: y can take on all values from a subset of the real line $\mathbb{R}$

- probability of it taking any specific value

$$\Pr(y = \text{value})$$

  is always zero and not a useful description

Ex.: y represents the time it takes to run the marathon distance

The probability $\Pr(y = t) = 0$ for $t = 2\text{h}$ and for any $t \in \mathbb{R}$.

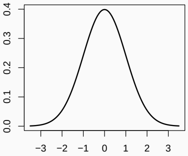
- Instead describe it by **density** or **cumulative probabilities**

▸ Marathon finishing times

Distribution functions

Probability density function (pdf):

$$f_y(\cdot)$$

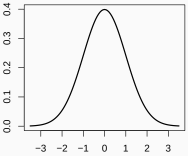


The area under $f_y(\cdot)$ between two points is the probability that the RV falls between these points

Distribution functions

Probability density function (pdf):

$$f_y(\cdot)$$



The area under $f_y(\cdot)$ between two points is the probability that the RV falls between these points

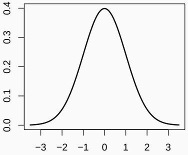
Ex.: Probability of finishing a marathon within 4 to 6 hrs

$\Pr(4 \leq y \leq 6) = \int_4^6 f_y(u)\mathrm{d}u = \Pr(4 < y < 6)$

Distribution functions

Probability density function (pdf):

$$f_y(\cdot)$$



The area under $f_y(\cdot)$ between two points is the probability that the RV falls between these points

Ex.: Probability of finishing a marathon within 4 to 6 hrs

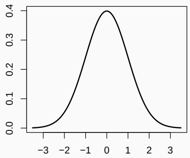
$\Pr(4 \leq y \leq 6) = \int_4^6 f_y(u)\mathrm{d}u = \Pr(4 < y < 6)$

- the density function $f_y(\cdot)$ always integrates to one
- continuous RVs are **fully characterized** by their density function

Distribution functions

Probability density function (pdf):

$$f_y(\cdot)$$



The area under $f_y(\cdot)$ between two points is the probability that the RV falls between these points

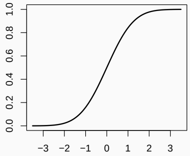
Ex.: Probability of finishing a marathon within 4 to 6 hrs

$\Pr(4 \leq y \leq 6) = \int_4^6 f_y(u)\mathrm{d}u = \Pr(4 < y < 6)$

- the density function $f_y(\cdot)$ always integrates to one
- continuous RVs are **fully characterized** by their density function
- the joint distribution of two variables can be described by a joint density function, e.g.: $f_{x,y}$

Cumulative distribution function (cdf)

$$F_y(\cdot)$$



Sometimes it is more convenient to visualize distributions as:

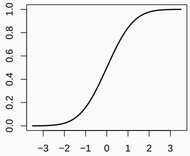
$$F_y(a) = \Pr(y \leq a) = \int_{-\infty}^{a} f_y(u)\mathrm{d}u.$$

- $F_y(6)$ is the probability of finishing within 6 hrs
- $F_y(4)$ is the probability of finishing within 4 hrs

- $f_y(\cdot)$ and $F_y(\cdot)$ carry the same information about y's distribution

Cumulative distribution function (cdf)

$$F_y(\cdot)$$



Sometimes it is more convenient to visualize distributions as:

$$F_y(a) = \Pr(y \leq a) = \int_{-\infty}^{a} f_y(u)\mathrm{d}u.$$

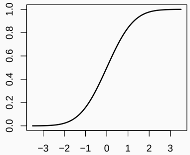
Ex.: Probability of finishing a marathon within 4 to 6 hrs

- $F_y(6)$ is the probability of finishing within 6 hrs
- $F_y(4)$ is the probability of finishing within 4 hrs

- $f_y(\cdot)$ and $F_y(\cdot)$ carry the same information about y's distribution

Cumulative distribution function (cdf)

$$F_y(\cdot)$$



Sometimes it is more convenient to visualize distributions as:

$$F_y(a) = \Pr(y \leq a) = \int_{-\infty}^{a} f_y(u)\mathrm{d}u.$$

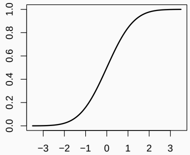
Ex.: Probability of finishing a marathon within 4 to 6 hrs

- $F_y(6)$ is the probability of finishing within 6 hrs
- $F_y(4)$ is the probability of finishing within 4 hrs
- $F_y(6) - F_y(4)$ is the prob of finishing within 4-6 hrs

- $f_y(\cdot)$ and $F_y(\cdot)$ carry the same information about y's distribution

Cumulative distribution function (cdf)

$$F_y(\cdot)$$



Sometimes it is more convenient to visualize distributions as:

$$F_y(a) = \Pr(y \leq a) = \int_{-\infty}^{a} f_y(u)\mathrm{d}u.$$

Ex.: Probability of finishing a marathon within 4 to 6 hrs

- $F_y(6)$ is the probability of finishing within 6 hrs
- $F_y(4)$ is the probability of finishing within 4 hrs
- $F_y(6) - F_y(4)$ is the prob of finishing within 4-6 hrs

- $f_y(\cdot)$ and $F_y(\cdot)$ carry the same information about y's distribution

Moments (1)

We often need summaries of distributions and don't want to describe the whole $f_y(\cdot)$ or $\Pr(\cdot)$.

Mean:

- describes the **location** of a distribution

Moments (1)

We often need summaries of distributions and don't want to describe the whole $f_y(\cdot)$ or $\Pr(\cdot)$.

Mean:

- describes the **location** of a distribution
- defined as the **expected value** of a RV y, $\mathbb{E}[y]$ (often short μ_y)
    - Discrete:

$$\mathbb{E}[y] = \sum_{k=0}^{K} k \Pr(y = k)$$

Moments (1)

We often need summaries of distributions and don't want to describe the whole $f_y(\cdot)$ or $\Pr(\cdot)$.

Mean:

- describes the **location** of a distribution
- defined as the **expected value** of a RV y, $\mathbb{E}[y]$ (often short μ_y)
  - Discrete:

$$\mathbb{E}[y] = \sum_{k=0}^{K} k \Pr(y = k)$$

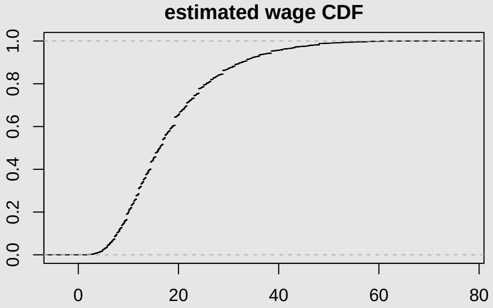
  - Continuous:

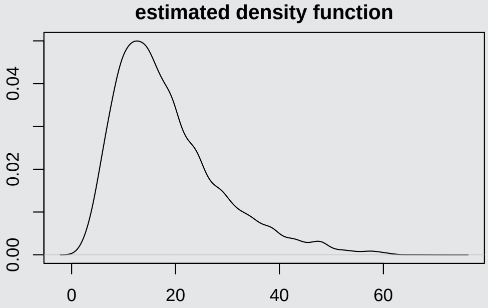
$$\mathbb{E}[y] = \int_{-\infty}^{+\infty} t f_y(t) \mathrm{d}t$$

- often but not *always* a great summary measure

Example of a skewed distribution

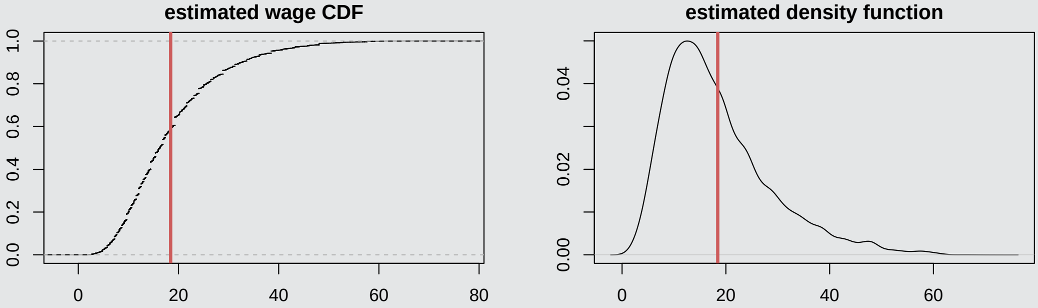
Ex.: Hourly wages of 61k people (2004 Current Population Survey)





Example of a skewed distribution

Ex.: Hourly wages of 61k people (2004 Current Population Survey)



$\mathbb{E}(\text{wage}) = 18.44\$/\text{hr}$

Moments (2)

Variance:

- describes the **dispersion** of a distribution

Moments (2)

Variance:

- describes the **dispersion** of a distribution
- defined as mean squared distance to expected value

$$\operatorname{var}(y) = \mathbb{E}[(y - \mu_y)^2] = \mathbb{E}[y^2] - \mu_y^2$$

Moments (2)

Variance:

- describes the **dispersion** of a distribution
- defined as mean squared distance to expected value

$$\mathrm{var}(y) = \mathbb{E}[(y - \mu_y)^2] = \mathbb{E}[y^2] - \mu_y^2$$

- alternatively we often use the standard deviation:

$$\sigma_y = \sqrt{\mathrm{var}(y)}$$

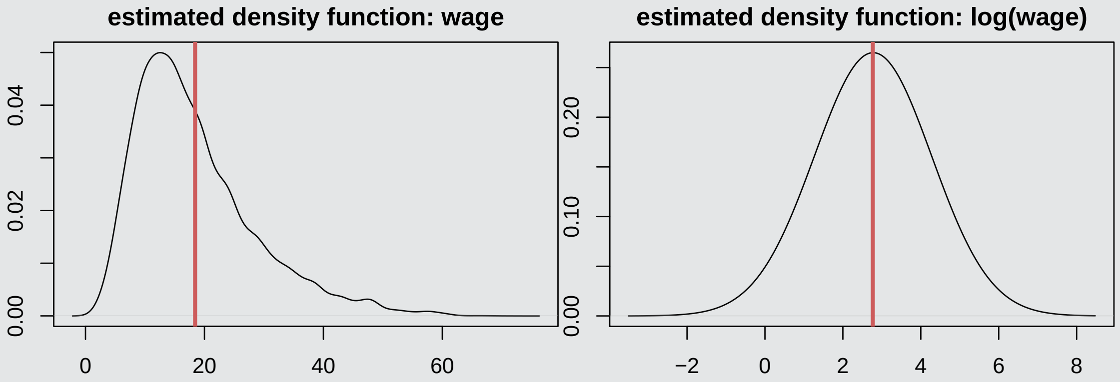
Example of a skewed distribution (i.e. with outliers)

Outliers are a reason why economists often use logarithms for wage data.

Example of a skewed distribution (i.e. with outliers)

Outliers are a reason why economists often use logarithms for wage data.

Ex.: Hourly wages of 61k people (2004 Current Population Survey)



log(wage) is precisely described by mean and variance. The distribution of wage is asymmetric and more

Joint distributions and covariance

Covariance: The covariance between x and y is

$$\mathrm{cov}(x, y) = \mathbb{E}[(x - \mu_x)(y - \mu_y)] = \sigma_{xy}.$$

- It is easily shown that: $\mathrm{cov}(x, y) = \mathbb{E}[xy] - \mu_x \mu_y$

Joint distributions and covariance

Covariance: The covariance between x and y is

$$\mathrm{cov}(x, y) = \mathbb{E}[(x - \mu_x)(y - \mu_y)] = \sigma_{xy}.$$

- It is easily shown that: $\mathrm{cov}(x, y) = \mathbb{E}[xy] - \mu_x \mu_y$

Properties:

- measures of the **linear association** between x and y
- $\mathrm{cov}(x, y) > 0$ means a positive relation between x and y

Joint distributions and covariance

Covariance: The covariance between x and y is

$$\mathrm{cov}(x, y) = \mathbb{E}[(x - \mu_x)(y - \mu_y)] = \sigma_{xy}.$$

- It is easily shown that: $\mathrm{cov}(x, y) = \mathbb{E}[xy] - \mu_x \mu_y$

Properties:

- measures of the **linear association** between x and y
- $\mathrm{cov}(x, y) > 0$ means a positive relation between x and y
- if x and y are independent ($x \perp y$), then $\mathrm{cov}(x, y) = 0$ (the converse is not true!)
- $\mathrm{cov}(x, x) = \mathrm{var}(x)$

Joint distributions and covariance

Covariance: The covariance between x and y is

$$\mathrm{cov}(x, y) = \mathbb{E}[(x - \mu_x)(y - \mu_y)] = \sigma_{xy}.$$

- It is easily shown that: $\mathrm{cov}(x, y) = \mathbb{E}[xy] - \mu_x \mu_y$

Properties:

- measures of the **linear association** between x and y
- $\mathrm{cov}(x, y) > 0$ means a positive relation between x and y
- if x and y are independent ($x \perp y$), then $\mathrm{cov}(x, y) = 0$ (the converse is not true!)
- $\mathrm{cov}(x, x) = \mathrm{var}(x)$
- its units are (units of x) $\times$ (units of y)

Correlation coefficient:

- **unitless** measure of **linear** association defined in terms of covariance and variances:

$$\operatorname{corr}(x, y) = \frac{\operatorname{cov}(x, y)}{\sqrt{\operatorname{var}(x)\operatorname{var}(y)}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}.$$

Correlation coefficient:

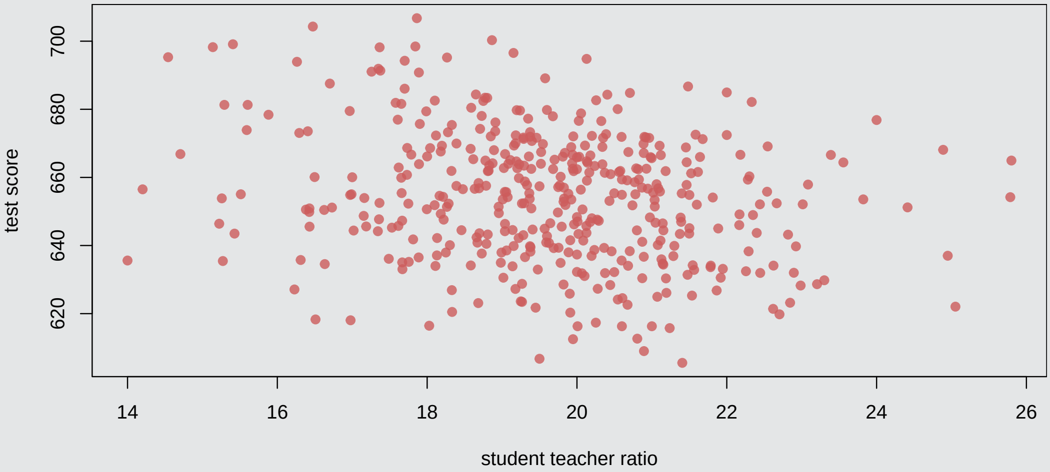
- **unitless** measure of **linear** association defined in terms of covariance and variances:

$$\operatorname{corr}(x, y) = \frac{\operatorname{cov}(x, y)}{\sqrt{\operatorname{var}(x)\operatorname{var}(y)}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}.$$

Properties:

- $-1 \leq \operatorname{corr}(x, y) \leq 1$.
- $\operatorname{corr}(x, y) = 1$ means perfect positive linear association
- $\operatorname{corr}(x, y) = -1$ means perfect negative linear association
- $\operatorname{corr}(x, y) = 0$ means no linear association

Ex.: Districts' avg. test score and class size data (CASchools)



$\operatorname{corr}(\text{test score}, \text{student teacher ratio}) = -0.23$ and $\operatorname{cov}(\text{test score}, \text{student teacher ratio}) = -8.16$

Recap: Random variables

- Tool to mathematically describe things that are random/uncertain

Recap: Random variables

- Tool to mathematically describe things that are random/uncertain
- RVs are fully characterized by:
  - probabilities for each outcome (discrete case)
  - pdf/cdf (continuous case)

Recap: Random variables

- Tool to mathematically describe things that are random/uncertain
- RVs are fully characterized by:
  - probabilities for each outcome (discrete case)
  - pdf/cdf (continuous case)
- Sometimes more conveniently characterized by moments:
  - mean, variance, . . .

Recap: Random variables

- Tool to mathematically describe things that are random/uncertain
- RVs are fully characterized by:
  - probabilities for each outcome (discrete case)
  - pdf/cdf (continuous case)
- Sometimes more conveniently characterized by moments:
  - mean, variance, ...
- Sometimes easier to work with transformed variables, e.g. $\log(\cdot)$
- Multiple random variables can be described by a joint distribution:
  - correlation can be described by the correlation coefficient or covariance

Recap: Random variables

- Tool to mathematically describe things that are random/uncertain
- RVs are fully characterized by:
  - probabilities for each outcome (discrete case)
  - pdf/cdf (continuous case)
- Sometimes more conveniently characterized by moments:
  - mean, variance, ...
- Sometimes easier to work with transformed variables, e.g. $\log(\cdot)$
- Multiple random variables can be described by a joint distribution:
  - correlation can be described by the correlation coefficient or covariance

Note: Sample moments vs population moments

- note difference between the **mean** (a property of the RV, also called "**population moment**") and the **average** (also called **sample moment**), which refers to the estimator and is itself a random variable.

Conditional distributions

Conditional distributions

Typically we don't study one RV **in isolation**, but **relate** it to **other variables**.

Probability of completing a marathon given that one is older than 50

Conditional distributions

Typically we don't study one RV **in isolation**, but **relate** it to **other variables**.

Conditional distribution:

- distribution of a RV y conditional on another RV x taking on specific values

Probability of completing a marathon given that one is older than 50

Conditional distributions

Typically we don't study one RV **in isolation**, but **relate** it to **other variables**.

Conditional distribution:

- distribution of a RV y conditional on another RV x taking on specific values

Ex.: Test scores (continuous)

The distribution of test scores given that class size is smaller than 20

Probability of completing a marathon given that one is older than 50

Conditional distributions

Typically we don't study one RV **in isolation**, but **relate** it to **other variables**.

Conditional distribution:

- distribution of a RV y conditional on another RV x taking on specific values

Ex.: Test scores (continuous)

The distribution of test scores given that class size is smaller than 20

Ex.: Completing a marathon (discrete)

Probability of completing a marathon given that one is older than 50

Conditional distributions

Typically we don't study one RV **in isolation**, but **relate** it to **other variables**.

Conditional distribution:

- distribution of a RV y conditional on another RV x taking on specific values

Ex.: Test scores (continuous)

The distribution of test scores given that class size is smaller than 20

Ex.: Completing a marathon (discrete)

Probability of completing a marathon given that one is older than 50

(Discrete, as we're not looking at finishing times ($\Omega = \mathbb{R}^+$) but at finishing: $\Omega = \{\text{no}, \text{yes}\}$)

Bayes rule yields conditional distributions

Bayes rule

For two events A and B with $\Pr(B) > 0$, we have that:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Bayes rule yields conditional distributions

Bayes rule

For two events A and B with $\Pr(B) > 0$, we have that:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

In the **discrete case**:

$$\Pr(y = a|x = b) = \frac{\Pr(y = a, x = b)}{\Pr(x = b)}$$

Bayes rule yields conditional distributions

Bayes rule

For two events A and B with $\Pr(B) > 0$, we have that:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

In the **discrete case**:

$$\Pr(y = a|x = b) = \frac{\Pr(y = a, x = b)}{\Pr(x = b)}$$

Ex.: Completing a marathon (discrete)

$$\Pr(\text{completing}|\text{age} > 50) = \frac{\Pr(\text{completing,age}>50)}{\Pr(\text{age}>50)}$$

Bayes rule yields conditional distributions

Bayes rule

For two events A and B with $\Pr(B) > 0$, we have that:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

In the **discrete case**:

$$\Pr(y = a|x = b) = \frac{\Pr(y = a, x = b)}{\Pr(x = b)}$$

Ex.: Completing a marathon (discrete)

$$\Pr(\text{completing}|\text{age} > 50) = \frac{\Pr(\text{completing}, \text{age}>50)}{\Pr(\text{age}>50)}$$

In the **continuous case**:

$$f_{y|x}(a|b) = \frac{f_{y,x}(a,b)}{f_x(b)},$$

- $f_x(\cdot)$ is the density of x
- $f_{x,y}(\cdot)$ is the joint density of x and y
- $f_{y|x}(\cdot)$ is the conditional density of y given x

The conditional expectation/mean of y given x ...

The conditional expectation/mean of y given x ...

... is the **mean** of the **conditional distribution** of y given x

Ex.: Test scores (continuous)
The expected test scores given that class size is smaller than 20.

The conditional expectation/mean of y given x ...

... is the **mean** of the **conditional distribution** of y given x

Ex.: Test scores (continuous)
The expected test scores given that class size is smaller than 20.

If y is continuous, we write:

$$\mathbb{E}[y|x=b] = \int_{-\infty}^{+\infty} u f_{y|x}(u|x=b)\mathrm{d}u.$$

The conditional expectation/mean of y given x ...

... is the **mean** of the **conditional distribution** of y given x

Ex.: Test scores (continuous)
The expected test scores given that class size is smaller than 20.

If y is continuous, we write:

$$\mathbb{E}[y|x = b] = \int_{-\infty}^{+\infty} u f_{y|x}(u|x = b)\mathrm{d}u.$$

Ex.: Marathon (discrete)
The probability of finishing a marathon given that one is older than 50.

If y is discrete with mass points at $\{y_1, y_2, \ldots\}$:

$$\mathbb{E}[y|x = b] = \sum_{j=1}^{\infty} y_j \Pr(y = y_j|x = b).$$

Ex.: Hourly wages of 61k people (2004 Current Population Survey)

For simplicity, treat this data as if it's the population, not sample. Ignore that means are estimated sample means.

Mean:

- $\mathbb{E}[\log(\text{wage})] = 2.77$

Ex.: Hourly wages of 61k people (2004 Current Population Survey)

For simplicity, treat this data as if it's the population, not sample. Ignore that means are estimated sample means.

Mean:

- $\mathbb{E}[\log(\text{wage})] = 2.77$

Conditional means:

- $\mathbb{E}[\log(\text{wage})|\text{gender} = \text{male}] = 2.86$

Ex.: Hourly wages of 61k people (2004 Current Population Survey)

For simplicity, treat this data as if it's the population, not sample. Ignore that means are estimated sample means.

Mean:

- $\mathbb{E}[\log(\text{wage})] = 2.77$

Conditional means:

- $\mathbb{E}[\log(\text{wage})|\text{gender} = \text{male}] = 2.86$
- $\mathbb{E}[\log(\text{wage})|\text{gender} = \text{female}] = 2.65$

Law of iterated expectations

For two random variables x and y:

$$\mathbb{E}[y] = \mathbb{E}[\mathbb{E}[y|x]]$$

Law of iterated expectations

For two random variables x and y:

$$\mathbb{E}[y] = \mathbb{E}[\mathbb{E}[y|x]]$$

sometimes written as:

$$\mathbb{E}_y[y] = \mathbb{E}_x[\mathbb{E}_y[y|x]]$$

Law of iterated expectations

For two random variables x and y:

$$\mathbb{E}[y] = \mathbb{E}[\mathbb{E}[y|x]]$$

sometimes written as:

$$\mathbb{E}_y[y] = \mathbb{E}_x[\mathbb{E}_y[y|x]]$$

I.e, the expectation of y can be computed from the conditional expectations of y given x.

Law of iterated expectations

For two random variables x and y:

$$\mathbb{E}[y] = \mathbb{E}[\mathbb{E}[y|x]]$$

sometimes written as:

$$\mathbb{E}_y[y] = \mathbb{E}_x[\mathbb{E}_y[y|x]]$$

I.e, the expectation of y can be computed from the conditional expectations of y given x.

Implication for discrete x

we can compute $\mathbb{E}[y]$ from $\mathbb{E}[y|x]$ through casework:

$$\mathbb{E}[y] = p_1\,\mathbb{E}[y|x = v_1] + p_2\,\mathbb{E}[y|x = v_2] + p_3\,\mathbb{E}[y|x = v_3] + \ldots$$

Law of iterated expectations

Ex.: Hourly wages of 61k people (2004 Current Population Survey)

Consider again hourly wages

- $\mathbb{E}[\log(\text{wage})] = 2.77$
- $\mathbb{E}[\log(\text{wage})|\text{gender} = \text{male}] = 2.86$
- $\mathbb{E}[\log(\text{wage})|\text{gender} = \text{female}] = 2.65$

Law of iterated expectations

Ex.: Hourly wages of 61k people (2004 Current Population Survey)

Consider again hourly wages

- $\mathbb{E}[\log(\text{wage})] = 2.77$
- $\mathbb{E}[\log(\text{wage})|\text{gender} = \text{male}] = 2.86$
- $\mathbb{E}[\log(\text{wage})|\text{gender} = \text{female}] = 2.65$

Observe that:

- $\Pr[\text{gender} = \text{male}] = 0.56$ and $\Pr[\text{gender} = \text{female}] = 0.44$

Law of iterated expectations

Ex.: Hourly wages of 61k people (2004 Current Population Survey)

Consider again hourly wages

- $\mathbb{E}[\log(\text{wage})] = 2.77$
- $\mathbb{E}[\log(\text{wage})|\text{gender} = \text{male}] = 2.86$
- $\mathbb{E}[\log(\text{wage})|\text{gender} = \text{female}] = 2.65$

Observe that:

- $\Pr[\text{gender} = \text{male}] = 0.56$ and $\Pr[\text{gender} = \text{female}] = 0.44$

L.I.E. implies:

$$\begin{aligned}\mathbb{E}[\log(\text{wage})] = \; & \mathbb{E}\left[\log(\text{wage})\,|\text{gender}{=}\text{male}\right] \times \Pr(\text{gender}{=}\text{male}) \\ & + \mathbb{E}\left[\log(\text{wage})\,|\text{gender}{=}\text{female}\right] \times \Pr(\text{gender}{=}\text{female}) \\ = & 2.86 \times 0.56 + 2.65 \times 0.44 \approx 2.77.\end{aligned}$$

Law of iterated expectations

Ex.: LIE in a proof

To be proven: $\mathbb{E}[e|x] = 0 \Rightarrow \mathbb{E}[ex] = 0$

Law of iterated expectations

Ex.: LIE in a proof

To be proven: $\mathbb{E}[e|x] = 0 \Rightarrow \mathbb{E}[ex] = 0$

$$\mathbb{E}[ex] \overset{L.I.E.}{=} \mathbb{E}[\mathbb{E}[ex|x]]$$

Law of iterated expectations

Ex.: LIE in a proof

To be proven: $\mathbb{E}[e|x] = 0 \Rightarrow \mathbb{E}[ex] = 0$

$$\begin{aligned}\mathbb{E}[ex] &\overset{L.I.E.}{=} \mathbb{E}[\mathbb{E}[ex|x]] \\ &= \mathbb{E}[x\mathbb{E}[e|x]]\end{aligned}$$

Law of iterated expectations

Ex.: LIE in a proof

To be proven: $\mathbb{E}[e|x] = 0 \Rightarrow \mathbb{E}[ex] = 0$

$$\begin{aligned}
\mathbb{E}[ex] &\overset{L.I.E.}{=} \mathbb{E}[\mathbb{E}[ex|x]] \\
&= \mathbb{E}[x\mathbb{E}[e|x]] \\
&= \mathbb{E}[x \underbrace{\mathbb{E}[e|x]]}_{=0}
\end{aligned}$$

Law of iterated expectations

Ex.: LIE in a proof

To be proven: $\mathbb{E}[e|x] = 0 \Rightarrow \mathbb{E}[ex] = 0$

$$\begin{aligned}
\mathbb{E}[ex] &\overset{L.I.E.}{=} \mathbb{E}[\mathbb{E}[ex|x]] \\
&= \mathbb{E}[x\mathbb{E}[e|x]] \\
&= \mathbb{E}[x\,\underbrace{\mathbb{E}[e|x]]}_{=0} \\
&= \mathbb{E}[x \cdot 0]
\end{aligned}$$

Law of iterated expectations

Ex.: LIE in a proof

To be proven: $\mathbb{E}[e|x] = 0 \Rightarrow \mathbb{E}[ex] = 0$

$$
\begin{aligned}
\mathbb{E}[ex] &\overset{L.I.E.}{=} \mathbb{E}[\mathbb{E}[ex|x]] \\
&= \mathbb{E}[x\mathbb{E}[e|x]] \\
&= \mathbb{E}[x \underbrace{\mathbb{E}[e|x]}_{=0}] \\
&= \mathbb{E}[x \cdot 0] \\
&= \mathbb{E}[0]
\end{aligned}
$$

Law of iterated expectations

Ex.: LIE in a proof

To be proven: $\mathbb{E}[e|x] = 0 \Rightarrow \mathbb{E}[ex] = 0$

$$
\begin{aligned}
\mathbb{E}[ex] &\overset{L.I.E.}{=} \mathbb{E}[\mathbb{E}[ex|x]] \\
&= \mathbb{E}[x\mathbb{E}[e|x]] \\
&= \mathbb{E}[x \underbrace{\mathbb{E}[e|x]}_{=0}] \\
&= \mathbb{E}[x \cdot 0] \\
&= \mathbb{E}[0] \\
&= 0
\end{aligned}
$$

Conditional variance

Just ike the mean, other moments have conditional counterparts

Conditional variance

$$\mathrm{var}(y|x=b) \equiv \sigma_y^2(x=b) = \mathbb{E}[(y-\mathbb{E}[y|x=b])^2|x=b]$$

Recap: Conditional distributions

We are often interested in describing the relationship between 2+ RVs

- Conditional distributions describe how y behaves for given values of x
- Conditional distributions can be obtained via Bayes rule.
- Conditional distributions can be described by conditional moments (e.g., conditional mean, variance)

Regression lines (OLS) are econometric models for conditional means

A larger part of what we study in this class are how to estimate conditional moments (means).

References

References

Allen, Eric, Patricia Dechow, Devin Pope, and George Wu. 2016. "Reference-Dependent Preferences: Evidence from Marathon Runners." *Management Science* 63 (April). https://doi.org/10.1287/mnsc.2015.2417.

Stock, James H, and Mark W Watson. 2020. *Introduction to Econometrics*. Pearson.

Wooldridge, Jeffrey M. 2019. *Introductory Econometrics: A Modern Approach*. Cengage learning.

Appendix

Marathon finishing times