Quantitative economics (L11QUE-ECON1045) Quantitative methods (L11QUM–ECON1047) Mathematical Economics and Statistical Methods (ECON1046)

Lecture Notes

Alessandro Ruggieri*

April 7, 2020

Abstract

These lecture notes are meant for undergraduate students attending an introductory course in Statistics and Probability. It is not meant to substitute standard textbooks but to complement them instead. This text is devided into seven sections, covering 1) descriptive statistics, 2) set and probability theory, 3) discrete random variables, 4) continuous random variables, 5) estimators and sampling distributions, 6) hypothesis testing and 7) linear regression. For each of these chapters, several examples and exercises are provided.

*Contact: School of Economics, Sir Clive Granger building, University of Nottingham, University Park, Nottingham, NG7 2RD, aruggierimail@gmail.com. The usual disclaimers apply.

Contents

Chapter 1

Descriptive Statistics

1.1. Representation of data

Consider a generic sample of n observations for a given variable x, i.e. $x_1, x_2, ...x_n$. There are at least three possible representation of these observations

1. *Raw representation*: You write down the data in the order you observe them

2. *Ordered representation*: You put the observations on the real line and renumber them in an ascending order

3. *Frequency representation*: For each observation, you construct absolute frequency

$$n_i = \#\text{observation}_i$$

and relative frequency

$$f_i = \frac{n_i}{\sum_{i=1}^{n} n_i}$$

Notice that

$$\sum_{i=1}^{n} n_i = n \implies \sum_{i=1}^{n} f_i = 1$$

1.2. Measures of centrality

Consider a generic sample of n observations for a given variable x, $x_1, x_2, ...x_n$.

Mean The sample mean of x is the usual arithmetic mean, defined as follows

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Median The sample median of x is the observation having the property that at least 50% percent of the data are less than or equal to it and at least 50% percent of the data values are greater than or equal to it. If two data values satisfy this condition, then the sample median percentile is the arithmetic average of these values.

Mode The mode of x is the most frequent observation. Note that it is possible for a frequency distribution to have more than one most frequent observation.

Geometric Mean The geometric mean is used to measure the rate of change of a variable over time. For a sample of n observations, it is equal to the n^{th} root of a product of n values:

$$\bar{x}_g = (x_1 x_2 ... x_n)^{\frac{1}{n}}$$

The geometric mean is used to calculate the average rate of return/growth over a series of time periods. In a sample of n of one-period return/growth $x_i = 1 + r_i$, the average rate of return/growth over n period is equal to

$$\bar{r}_g = \bar{x}_g - 1$$

1.3. Measures of dispersion

Range Denoted by x_{min} the smallest observation in the sample, and by x_{max} the largest observation. Then the range is defined as the difference of these two values, i.e.

$$\text{range} = x_{max} - x_{min}$$

Percentiles The idea behind percentiles is the same as with the median: certain intervals should contain certain fractions of the total number of observations. To determine the sample $100p$ percentile in a sample of size n, we must determine the data value such that

1. At least np of the data values are less than or equal to it.
2. At least $n(1-p)$ of the data values are greater than or equal to it.

Therefore, to find the sample $100p$ percentile of a data set

1. Arrange the data in increasing order.
2. If np is not an integer, determine the smallest integer greater than np. The data value in that position is the sample $100p$ percentile.
3. If np is an integer, then the average of the values in positions np and is the sample $100p$ percentile.

Interquartile range The sample $25th$ percentile is called the first quartile. The sample $50th$ percentile is called the median or the second quartile. The sample $75th$ percentile is called the third quartile. The interquartile range is difference between the third and the first quartile of a sample.

Variance The sample variance of x is defined by

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

where $x_i - \bar{x}$ are called deviations from the mean. The quantity

$$s_x = \sqrt[2]{s_x^2}$$

is called sample standard deviation.

1.4. Measures of relationship between variables

Consider a generic sample of n observations for two given variables x and y, i.e. $x_1, x_2, ...x_n$, and $y_1, y_2, ...y_n$.

Covariance The sample covariance between x and y is defined by

$$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

where (x_i, y_i) are observations on a pair of variables

Correlation coefficient A sample correlation coefficient between x and y is defined by

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

where s_{xy} is the sample covariance and s_x and s_y are sample standard deviations. We say that the variables x and y are positively correlated if $r_{xy} > 0$ and negatively correlated if $r_{xy} < 0$. In case $r_{xy} = 0$ the variables are called uncorrelated. In the extreme cases where $r_{xy} = 1$ (-1) we talk about perfect positive (negative) correlation.

1.5. Exercises

Exercise 1 The time (in seconds) that a sample of $n = 10$ employees took to complete a task is: $\{14, 28, 40, 13, 25, 27, 20, 29, 49, 66\}$. Find the following for this data:

- Sample mean
- Sample median

- Sample variance
- Sample standard deviation
- Sample coefficient of variation

Exercise 2 Consider the following sample of prices of certain goods and quantities sold:

$$\text{Price} = \{10, 15, 20, 25, 30\}$$
$$\text{Quantity} = \{100, 90, 75, 50, 0\}$$

- Sketch a scatter plot of price (along the x-axis) and quantity (along the y-axis). Do you think these two variables will have a negative covariance, a positive covariance, or zero covariance?
- Compute and interpret the covariance and the correlation coefficient.

Exercise 3 Consider the following sample of observation for monthly income (in 1000GBP):

$$X = \{1.0, 1.0, 1.1, 1.1, 1.1, 1.2, 1.2, 1.4, 1.5, 1.6, 1.6, 2.2, 3.4, 7.6, 10.6, 26.4\}$$

Find the following for this data:

- Sample mean
- Sample median
- Sample mode
- Sample range
- Sample variance
- Sample standard deviation.

Plot the relative frequency distribution and the cumulative relative frequency distribution.

Chapter 2

Set and probability theory

2.1. Set operation

A set is a collection of objects, which are elements of the set. Sets are usually denoted using capital letters, while elements are denotes using lower-case letters. The elements of a set can be interpreted as the outcomes or the results of an experiment or a random trial.

Sample space The sample space is an exhaustive collection of objects/outcomes of an experiment.

Let S denotes a sample space. Let A and B be two sets of elements that belong to the sample space.

Union The union of two sets A and B, $A \cup B$ is the set of elements which belong to A, to B or both. If $A \cup B = S$, they we say that A and B are mutually exhaustive.

Intersection Consider two sets A and B. The intersection $A \cap B$ is the set of elements common to A and B . A and B are said to be disjoint if the intersection of A and B is an empty set, i.e. if $A \cap B = \emptyset$. We say also that A and B are mutually exclusive.

Subsets We say that A is a subset of B, i.e. $A \subset B$, if all elements of A belong to B. If A is not a subset of B, we write $A \not\subset B$

Complementarity The difference $A \setminus B$ is the set of elements of A which do not belong to S . A complement $\overline{A}$ is defined as

$$\overline{A} = S \setminus A$$

and denotes the set of elements that belong to S outside A.

2.2. Set identities

Below some set relations.

- A set B can be obtained as a union of two disjoint sets, i.e. $B = (A \cap B) \cup (\overline{A} \cap B)$
- $A \cup B = A \cup (\overline{A} \cap B)$
- Distributive law: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- First De Morgan's law: $\overline{A \cup B} = \overline{A} \cap \overline{B}$
- Second De Morgan's law: $\overline{A \cap B} = \overline{A} \cup \overline{B}$

2.3. Probability

Let S be a sample space. By probability on S we mean a numerical function $P(A)$ of events $A \subset S$ such that the following properties hold:

- *non-negativity*: for any $A \subset S$, i.e. for any event that is in the sample space, $0 \leq P(A) \leq 1$
- Let $A \subset S$, and let $O_i \in A$ for $i = 1, ...k$. Then
$$P(A) = \sum_{i=1}^{k} P(O_i)$$
- *completeness*: $P(S) = 1$

Notice that

- if A and B are mutually exclusive, i.e. $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$
- if A and B are not mutually exclusive, i.e. $A \cap B \neq \emptyset$, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- if A and B are collectively exhaustive, i.e. $A \cup B = S$, then $P(A \cup B) = 1$. Therefore, $P(\overline{A}) = P(S \setminus A) = P(S) - P(A) = 1 - P(A)$.

An impossible event can be defined by $P(A) = 0$. When $P(A) = 1$, we say that A is a sure event.

Joint probability The event $A \cap B$ is called a joint event and its probability $P(A \cap B)$ is called a joint probability.

Conditional probability Let $P(B) > 0$. The probability $P(A|B)$ is called probability of A conditional on B and is defined as follows

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

This obviously implies that $P(A \cap B) = P(A|B)P(B)$.

Independence Events A and B are called statistically independent if occurrence of one of them does not influence in any way the chances of occurring of the other, i.e. if $P(A \cup B) = P(A)P(B)$. When $P(B) > 0$, then this is equivalent to $P(A \cup B) = P(A)$

Bayes theorem: $P(A|B) = P(B|A)\frac{P(A)}{P(B)}$

Proof. Using the definition of conditional probability, we can write:

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

2.4. Exercises

Exercise 1 Out of a group of 20 people, 15 enjoy cycling and 8 enjoy reading. 2 of them enjoy neither cycling nor reading. How many people enjoy both cycling and reading?

Exercise 2 A committee of two members is to be chosen from a group of 4 male and 2 female candidates. If each member has the same probability of being chosen, what is the probability that both members are male?

Exercise 3 The manager of a music store finds that 30% of customers entering the store ask ban assistant for help, and 20% of customers entering the store make a purchase before leaving. Also, 15% of customers entering the store both ask an assistant for help and make a purchase before leaving.

- What is the probability that a customer entering the store will ask an assistant for help or make a purchase or both?
- What is the probability that a customer that asks an assistant for help will make a purchase before leaving?
- Consider the events “asks assistant for help" and “makes purchase"
  - Are these events mutually exclusive?
  - Are they collectively exhaustive?
  - Are they independent?

Exercise 4 Let the sample space be the collection of all possible outcomes of rolling one die. Let A be the event: "# rolled is even". Let B be the event: "# rolled is at least 4". Find the following:

- complement of A, $\bar{A}$
- complement of B, $\bar{B}$
- intersection of A and B
- intersection of $\bar{A}$ and B
- union of A and B
- union of A and $\bar{A}$

Are A and B mutually exclusive? Are A and B collectively exhaustive?

Chapter 3

Discrete random variables

In probability and statistics, a random variable, is described informally as a variable whose values depend on outcomes of a random phenomenon. A discrete random variable X is a pair of values x_i and probabilities $p_i \equiv P(X = x_i)$, for all $i = 1, 2, ...n$, where n is finite integer. We call *probability distribution* the list of probabilities $\{p_i\}_{i=1}^n$ associated to outcomes $\{x_i\}_{i=1}^n$

Linear combination Suppose that X and Y are two discrete random variables with the same probability distribution $\{p_i\}_{i=1}^n$. Let a and b be real numbers. The random variable $aX + bY$ is called a linear combination of X and Y with coefficients a and b.

3.1. Probability distributions

Joint probability function Let X and Y be two random variables. A joint probability function $P(x, y)$ is used to express the probability that X takes the specific value x_i and simultaneously Y takes the value y, i.e. $P(x, y) = P(X = x, Y = y)$. By definition of probability, it must be that $\sum_x \sum_y P(x, y) = 1$.

Cumulative probability function Given a random variable X with probability distribution $P(x)$, the cumulative distribution function of x, $F(x)$ is defined as follows as the probability that X is less than or equal to x, i.e.

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

By construction, $0 \leq F(x) \leq 1$.
Consider now a pair of random variable X and Y with joint probability function $P(x, y)$. Analogously, the cumulative distribution function for this pair of random variables is defined in terms of their joint probability distribution:

$$F(x, y) = P(X \leq x, Y \leq y)$$

Marginal probability function Consider a pair of random variable X and Y with joint probability function $P(x, y)$. Their marginal probabilities are defined as

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

and

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

The marginal probability of X (Y) gives the probability of x (y) irrespective of Y (X).

3.2. Moments

Expected value The expected value of a discrete random variable is equal to

$$E[X] = \sum_{i=1}^{n} P(X = x_i)x_i$$

Notice that for any random variables, X and Y, with the same probability distribution $\{p_i\}_{i=1}^{n}$, and any numbers, a and b one has

$$E[aX + bY] = aE[X] + bE[Y]$$

Proof $E[aX + bY] = \sum_{i=1}^{n}(ax_i + by_i)p_i = \sum_{i=1}^{n} ax_ip_i + \sum_{i=1}^{n} by_ip_i = a\sum_{i=1}^{n} x_ip_i + b\sum_{i=1}^{n} y_ip_i = aE[X] + bE[Y]$.

If a random variable X is such that there exit an event x_i with $p(X = x_i) = 1$ and $p(X = x_j) = 0$ for all $j \neq i$, then $E(X) = x_i$.

Notice finally that in general, if $g(X)$ is some function of the random variable X then $E[g(X)]$ and $g[E(x)]$ are not necessarily equal.

Variance The variance of a discrete random variable is given by:

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

The standard deviation is the positive square root of the variance.

Covariance Let X and Y be two discrete random variables, the covariance between X and Y is given by:

$$COV[X,Y] = \sum_{i=1}^{n}\sum_{j=1}^{n} P(X = x_i, Y = y_j)(x_i - E[X])(y_i - E[Y])$$

This expression is equivalent to:

$$COV[X,Y] = \sum_{i=1}^{n}\sum_{j=1}^{n} P(X = x_i, Y = y_j)(x_i y_i - x_i E[Y] - E[X]y_i + E[X]E[Y]) =$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} P(x_i, y_j)x_i y_i - \sum_{i=1}^{n}\sum_{j=1}^{n} P(x_i, y_j)x_i E[Y] - \sum_{i=1}^{n}\sum_{j=1}^{n} P(x_i, y_j)E[X]y_i + \sum_{i=1}^{n}\sum_{j=1}^{n} P(x_i, y_j)E[X]E[Y] =$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} P(x_i, y_j)x_i y_i - E[Y]\sum_{i=1}^{n}\sum_{j=1}^{n} P(x_i, y_j)x_i - E[X]\sum_{i=1}^{n}\sum_{j=1}^{n} P(x_i, y_j)y_i + E[X]E[Y]\sum_{i=1}^{n}\sum_{j=1}^{n} P(x_i, y_j) =$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} P(x_i, y_j)x_i y_i - E[Y]\sum_{i=1}^{n} P(x_i)x_i - E[X]\sum_{j=1}^{n} P(y_j)y_i + E[X]E[Y] =$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} P(x_i, y_j)x_i y_i - E[X]E[Y]$$

Notice that for any random variables, X and Y, and for a constant b, it is true that:

- $VAR[X + Y] = VAR[X] + VAR[Y] + 2COV[X, Y]$
- $VAR[bX] = b^2 VAR[X]$
- $VAR[b] = 0$

Independence We say that X and Y are independent if the events independence condition is satisfied for every pair of their values, i.e. if

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

for all i and for all j.

Proposition If X and Y are independent random variables, then $COV(X, Y) = 0$

Proof. Use the definition of covariance derived above.

$$COV[X,Y] = \sum_{i=1}^{n}\sum_{j=1}^{n} P(x_i, y_j)x_i y_i - E[X]E[Y]$$

If $P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$, then

$$COV[X,Y] = \sum_{i=1}^{n}\sum_{j=1}^{n} P(X = x_i)P(Y = y_j)x_i y_i - E[X]E[Y] =$$
$$\sum_{i=1}^{n} P(X = x_i)x_i \sum_{j=1}^{n} P(Y = y_j)y_i - E[X]E[Y] =$$
$$E[X]E[Y] - E[X]E[Y] = 0$$

3.3. Examples of Discrete Random Variables

3.3.1 Bernoulli distribution

A Bernoulli random variable X describes the outcome of an experiment with only two possible outcomes, i.e. success ($x = 1$) and failure ($x = 0$), and a probability of success $P(X = 1)$ equal to $0 \leq p \leq 1$.

Expected value The expected value of a Bernoulli random variable is equal to the probability of success, p, i.e.

$$E[X] = \sum_{x} p(x)x = (1-p)0 + p1 = p$$

Variance The variance of a Bernoulli random variable is equal to the probability of success times the probability of failure, $p(1-p)$

$$VAR[X] = \sum_{x} p(x)(x - E[X])^2 = (1-p)(0-p)^2 + p(1-p)^2 =$$
$$= (1-p)(p^2 - 2p) + p(1 + p^2 - 2p) = p^2 - p^3 + p + p^3 - 2p^2 = -p^2 + p =$$
$$p(1-p)$$

3.3.2 Binomial distribution

A Binomial random variable describes the outcome of a series of n independent Bernoulli trials. Two assumptions are needed to construct a Binomial variable as a combination of Binomial r.v.:

- the probability of success for each trial is constant
- trials are independent: the outcome of one observation does not affect the outcome of the other

The probability of x successes out of n trials is defined as follows

$$P(X = x) = C_x^n p^x (1-p)^{(n-x)}$$

where C^n_x is the number of combination of x elements out of n, i.e.

$$C^n_x = \frac{n!}{x!(n-x)!}$$

Expected value The expected value of a Binomial random variable with n trials is equal to n times the expected value of a Bernoulli r.v., i.e. $E[X] = np$.

Proof. Using the formula for the expected value, we get

$$E[X] = \sum_x x C^n_x p^x (1-p)^{(n-x)}$$

Notice that $xC^n_x = nC^{n-1}_{x-1}$. Therefore:

$$\begin{aligned} E[X] &= \sum_x nC^{n-1}_{x-1} p^x (1-p)^{(n-x)} = n \sum_x C^{n-1}_{x-1} p^x (1-p)^{(n-x)} \\ &= np \sum_x C^{n-1}_{x-1} p^{x-1} (1-p)^{(n-1)-(x-1))} = np \end{aligned}$$

since $\sum_x C^{n-1}_{x-1} p^{x-1} (1-p)^{(n-1)-(x-1))} = 1$.

Variance The variance of a Binomial random variable with n trials is equal to n times the variance of a Bernoulli r.v., i.e. $VAR[X] = np(1-p)$.

3.3.3 Poisson distribution

A Poisson random variable is a discrete r.v. that represents a number of successes occurring in a fixed interval of time or space if events occur with a known constant mean rate and independently of the time since the last event. The probability of x successes in a given length of time is defined as follows:

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

where $\lambda > 0$ denotes the average number of successes in a given subinterval of time.

Expected value The expected value of a Poisson r.v. is equal to $E[X] = \lambda$.

Variance The variance of a Poisson r.v. is equal to $VAR[X] = \lambda$.

3.4. Exercises

Exercise 1 X is the number of heads from two tosses of a coin, i.e.

$$
\begin{aligned}
X &= 0 \quad \text{w/ prob} = 0.25 \\
X &= 1 \quad \text{w/ prob} = 0.50 \\
X &= 2 \quad \text{w/ prob} = 0.25
\end{aligned}
$$

Calculate expected value and variance of X.

Exercise 2 Consider the joint probability distribution

$$
\begin{aligned}
P(X = 0, Y = 1) &= 0.30 \\
P(X = 1, Y = 1) &= 0.25 \\
P(X = 0, Y = 2) &= 0.20 \\
P(X = 1, Y = 2) &= 0.25
\end{aligned}
$$

where X represents the number of exams a student has in a day during final examinations and Y represents the number of snacks eaten by the student during the same day.

- Find the marginal probability distributions for X and Y .
- Calculate the expected value and variance of X and Y .
- Calculate the probability of having one exam conditional on eating two snacks.
- Calculate the covariance of X and Y.

Exercise 3 The probability that a student applying to a college will be admitted is 60%. Suppose a student applies to 6 colleges. What is the probability that exactly 2 colleges admit her?

Exercise 4 Smartphones are shipped from the manufacturer in packets of 12. The probability of a phone being faulty is 0.1 and such faults are independent.

- What is the probability that no more than 2 phones in a shipment are faulty?
- If a shop receives 6 shipments, what is the probability that at least one shipment will contain 3 or more faulty phones?
- Let Y denote the number of shipments containing 3 or more faulty phones. What is the probability that Y will exceed its mean by more than 2 standard deviations?

Chapter 4

Continuous random variables

Informally, a continuous random variables describe outcomes in probabilistic situations where the possible values some quantity can take form a continuum. A continuous random variable X is characterized by a sample space $S \subseteq \mathcal{R}$, which denotes the range of possible outcomes, and by a probability density function (pdf) $f(x)$, which gives the relative likelihood of any outcome in a continuum occurring. Similarly, any collection of continuous random variables, $X_1, ...X_k$ is characterized by a collection of sample spaces, $S_1, ..., S_k$ and by a joint probability density function, $f(x_1, ...x_k)$, which gives the relative likelihood of any collection of outcome in a continuum occurring jointly.

4.1. Distribution Functions

Cumulative Distribution Function Let X be a continuous random variable with pdf $f(x)$ defined over the real space $\mathcal{R}$. Then its cumulative distribution function (cdf), $F(x)$ is defined as follows:

$$F(x) = \int_{-\infty}^{x} f(x)dx \quad -\infty \leq x \leq \infty$$

Proposition Let X be a continuous random variable with pdf $f(x)$ defined over the real space $\mathcal{R}$. The probability that X takes value in the interval $[a, b] \subset \mathcal{R}$ is defines as follows:

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$$

Proof. Notice that:

$$P(a \leq X \leq b) = F(b) - F(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = \int_{a}^{b} f(x)dx$$

Proposition Let X be a continuous random variable with pdf $f(x)$ defined over the real space $\mathcal{R}$. The probability that X takes value $a \in \mathcal{R}$ is equal to 0.

Proof. Notice that:

$$P(X = a) = \int_a^a f(x)dx = 0$$

Proposition Any density $f(x)$ defined over a sample space S satisfies the completeness axiom, i.e.

$$\int_{x \in S} f(x)dx = 1$$

Joint Cumulative Distribution Function Let X and Y be two continuous random variables, each defined on the real space. Let their joint probability function be $f(x, y)$. Then their joint cumulative distribution function (cdf), $F(x, y)$ is defined as follows:

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(x, y)dxdy \quad -\infty \leq x \leq \infty \quad -\infty \leq y \leq \infty$$

Proposition Let X and Y be two continuous random variables, each defined on the real space. Let their joint probability function be $f(x, y)$. The probability that X takes value in the interval $[a, b] \subset \mathcal{R}$ and that Y takes a value in the interval $[c, d]$ is defines as follows:

$$P(a \leq X \leq b, c \leq X \leq d) = \int_c^d \int_a^b f(x, y)dxdy$$

Marginal Density Function Let X and Y be two continuous random variables, each defined on the real space. Let their joint probability function be $f(x, y)$. The marginal density function of X, $f(x)$ assigns probabilities to a range of values of X irrespective of the values of Y can take, and it is defined as follows

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

Similarly, the marginal density function of Y is equal to

$$f(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

4.2. Moments

Let X be a continuous random variable with pdf $f(x)$ defined over the real space $\mathcal{R}$.

Expected value The expected value of X is defined by

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

Variance The variance of X is defined by

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

Notice that

$$\begin{aligned}
VAR(X) = \int_{-\infty}^{\infty} (x - E[X])^2 f(x)dx = \\
\int_{-\infty}^{\infty} (x^2 - 2xE[X] + E[X]^2) f(x)dx = \\
\int_{-\infty}^{\infty} x^2 f(x)dx - \int_{-\infty}^{\infty} 2xE[X]f(x)dx + \int_{-\infty}^{\infty} E[X]^2 f(x)dx = \\
\int_{-\infty}^{\infty} x^2 f(x)dx - 2E[X] \int_{-\infty}^{\infty} xf(x)dx + E[X]^2 \int_{-\infty}^{\infty} f(x)dx = \\
\int_{-\infty}^{\infty} x^2 f(x)dx - E[X]^2 = E[X^2] - E[X]^2
\end{aligned}$$

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

Let X and Y be two continuous random variables, each defined on the real space. Let their joint probability function be $f(x, y)$.

Covariance The covariance of two continuous random variables X and Y is given by:

$$COV(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E[X])(y - E[Y]) f(x, y)dxdy$$

Notice that

$$\begin{aligned}
COV(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E[X])(y - E[Y]) f(x, y)dxdy = \\
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (xy - xE[Y] - yE[X] + E[X]E[Y]) f(x, y)dxdy = \\
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dxdy - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xE[Y]f(x, y)dxdy \\
- \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yE[X]f(x, y)dxdy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E[X]E[Y]f(x, y)dxdy = \\
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dxdy - \int_{-\infty}^{\infty} xE[Y] \left[\int_{-\infty}^{\infty} f(x, y)dy \right] dx - \int_{-\infty}^{\infty} yE[X] \left[\int_{-\infty}^{\infty} f(x, y)dx \right] dy + E[X]E[Y] = \\
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dxdy - E[Y] \int_{-\infty}^{\infty} xf(x)dx - E[X] \int_{-\infty}^{\infty} yf(y)dy + E[X]E[Y] = \\
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dxdy - 2E[X]E[Y] + E[X]E[Y] = \\
E[XY] - E[X]E[Y]
\end{aligned}$$

where $E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y)dxdy$.

Independence The random variables X and Y are statistically independent if the joint probability density function can be written as the product of the marginal density functions:

$$f(x,y) = f(x)f(y)$$

Proposition If X and Y are statistically independent, then $COV(X,Y) = 0$.

Proof. To see this, notice that

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x)f(y)dxdyz =$$
$$\int_{-\infty}^{\infty} xf(x)\left[\int_{-\infty}^{\infty} yf(y)dy\right]dx = \int_{-\infty}^{\infty} xf(x)E[Y]dx = E[Y]\int_{-\infty}^{\infty} xf(x)dx = E[X]E[Y]$$

which implies that

$$COV(X,Y) = E[XY] - E[X]E[Y] = E[X]E[Y] - E[X]E[Y] = 0$$

4.3. Examples of Continuous Random Variables

4.3.1 Uniform distribution

The uniform distribution is a continuous probability distribution that has equal probabilities for all possible outcomes of the random variable. Let X be a uniform random variable distributed over the interval $[a,b] \subset \mathcal{R}$. We write $X \sim U[a,b]$.

PDF The probability density function of X is equal to

$$f(x) = \frac{1}{b-a} \quad a \leq X \leq b$$

CDF The cumulative distribution function of X is equal to

$$F(z) = P(X \leq z) = \int_a^z \frac{1}{b-a}dx = \frac{1}{b-a}\int_a^z dx = \frac{1}{b-a}[x]_a^z = \frac{1}{b-a}(z-a) = \frac{z-a}{b-a}$$

Expected value The expected value of X is equal to

$$E(z) = \int_a^b xf(x)dx = \int_a^b x\frac{1}{b-a}dx = \frac{1}{b-a}\int_a^b xdx =$$
$$\frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{1}{b-a}(\frac{b^2}{2} - \frac{a^2}{2}) = \frac{(b^2-a^2)}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2}$$

Variance The expected value of X is equal to

$$VAR(z) = E[X^2] - E[X]^2 = \int_a^b x^2 \frac{1}{b-a} dx - \left(\int_a^b x f(x) dx \right)^2 =$$
$$\frac{1}{b-a} \int_a^b x^2 dx - \frac{b+a}{2} = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b - \frac{b+a}{2} =$$
$$\frac{1}{3(b-a)} \left[b^3 - a^3 \right] - \frac{b+a}{2} = \frac{(b-a)^2}{12}$$

4.3.2 Standard Normal distribution

A standard normal random variable, denoted Z, has a probability density function defined as follows

$$f(z) = \frac{1}{\sqrt[2]{2\pi}} \exp^{-\frac{z^2}{2}}$$

Proposition The density of a standard normal random variable is symmetric about zero, i.e. $f(z) = f(-z)$

Proof. To see this, notice that

$$f(-z) = \frac{1}{\sqrt[2]{2\pi}} \exp^{-\frac{(-z)^2}{2}} = \frac{1}{\sqrt[2]{2\pi}} \exp^{-\frac{z^2}{2}} = f(z)$$

Expected value The expected value of standard normal random variable is zero.

Proof.

$$E[Z] = \int_{-\infty}^{\infty} z f(z) dz = \int_{-\infty}^{0} z f(z) dz + \int_{0}^{\infty} z f(z) dz$$

By symmetry, $\int_{-\infty}^{0} z f(z) dz = -\int_{0}^{\infty} z f(z) dz$, which implies that $E[Z] = 0$.

Variance The variance of standard normal random variable is one.

Proof. $VAR[Z] = E[Z^2] - E[Z]^2 = E[Z^2]$. Since $E[Z^2] = 1$, then $VAR[Z] = 1$.

4.3.3 Normal distribution

A normal variable X is defined as a linear transformation of the standard normal:

$$X = \mu + \sigma Z$$

where $\sigma > 0$. We write $X \sim \mathcal{N}(\mu, \sigma^2)$.

PDF The probability density function of normal random variable is equal to

$$f(x) = \frac{1}{\sigma \sqrt[2]{2\pi}} \exp^{-\frac{\left(\frac{x-\mu}{\sigma}\right)^2}{2}}$$

Expected value The expected value of normal random variable is μ.

Proof. $E[X] = E[\mu + \sigma Z] = E[\mu] + E[\sigma Z] = \mu + \sigma E[Z] = \mu$

Variance The variance of normal random variable is σ^2.

Proof. $VAR[X] = VAR[\mu + \sigma Z] = VAR[\mu] + VAR[\sigma Z] = 0 + \sigma^2 VAR[Z] = \sigma^2$

Proposition The probability that a normal r.v. X falls into the interval $[a, b]$ is equal to

$$P(a \leq X \leq b) = P(a \leq \mu + \sigma Z \leq b) =$$
$$P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) = P\left(Z \leq \frac{b-\mu}{\sigma}\right) - P\left(Z \leq \frac{a-\mu}{\sigma}\right)$$

4.4. Exercises

Exercise 1 X is a continuous random variable with probability density function (PDF)

$$f(x) = \frac{1}{9}x^2, \quad 0 \leq x \leq 3$$

- Find the following probabilities:
  - $P(0 \leq X \leq 1)$
  - $P(0 \leq X \leq 2)$
  - $P(1 \leq X \leq 2)$
- Find the expected value and variance of X.

Exercise 2 X is a continuous random variable with the following PDF

$$f(x) = 3x^2, \quad 0 \leq x \leq 1$$

Compute E[X] and VAR[X].

Exercise 3 The random variable Z has a standard normal distribution.

- Find the following probabilities:
  - $P(0 < Z < 1.20)$

- $P(-1.33 < Z < 0)$
- $P(Z > 1.33)$
- $P(-0.77 < Z < 1.68)$

- Find x given that $P(x < Z < 1.68) = 0.2$

Exercise 4 The tread life of a particular brand of tyre has a normal distribution with mean 35000 miles and standard deviation 4000 miles.

- What is the probability that a tyre of this brand will have a tread life between 35000 and 38000 miles?
- What is the probability that a tyre of this brand will have a tread life of less than 32000 miles?

Exercise 5 A repair team is responsible for a stretch of oil pipe 2 miles long. The distance at which any fracture occurs can be represented by a uniformly distributed random variable, with the PDF, $f(x) = 0.5$.

- Find the CDF of X
- Find the probability that any given fracture occurs between 0.5 mile and 1.5 miles along the stretch pipeline.

Exercise 6 A client has an investment portfolio whose mean value is equal to 1000000GBP, with a standard deviation of 30000GBP. Assume that the value of the portfolio follows a Normal distribution. Determine the probability that the portfolio is between 970000GBP and 1060000GBP.

Chapter 5

Estimators

Definition. A random sample is a sample that satisfies two conditions:

- Every object has an equal probability of being selected
- The objects are selected independently.

Consider a random sample $x_1, ..., x_n$. Notice that - although any given sample observation x_i takes a specific numerical value - in each sample it is still a random variable, since if the sampling process were repeated it will then take a different numerical value. Therefore each of these observations are ex-ante identical and independent random variables.

5.1. Definition and properties

Let $x_1, ..., x_n$ be a random sample of X. Let τ be a population parameter that characterizes X and let T be a measure designed to estimate τ. Since each observations in a sample is ex-ante an iid random variable, also T is random variable.

Estimators. Let T be a statistics to estimate τ. Then T is an estimator of τ. It is a random variable and depends on the sample data. Its distribution is called *sampling distribution.*

Unbiasedness. An estimator T of τ is said to be unbiased if

$$E[T] = \tau$$

The bias of an estimator T is defined as the difference between its expected value and value of the population parameter, i.e.

$$\text{bias}[T] = E[T] - \tau$$

If $E[T] > \tau$, we say that T is biased upward. If $E[T] < \tau$, we say that T is biased downward.

Efficiency. Let T_1 and T_2 be two two unbiased estimators of τ. Let $Var[T_1] < VAR[T_2]$. Then T_1 is said to be more efficient than T_2.

Mean square error. A combined measure of bias and inefficiency is given by the mean squared error, defined as follows:

$$\begin{aligned}
\mathrm{MSE}[T] = \mathrm{E}[(T-\tau)^2] = \mathrm{E}[(T-\mathrm{E}[T]+\mathrm{E}[T]-\tau)^2] = \\
\mathrm{E}[(T-\mathrm{E}[T])^2 + (\mathrm{E}[T]-\tau)^2 + 2(T-\mathrm{E}[T])(\mathrm{E}[T]-\tau)] = \\
\mathrm{E}[(T-\mathrm{E}[T])^2] + \mathrm{E}[(\mathrm{E}[T]-\tau)^2] + 2\mathrm{E}[(T-\mathrm{E}[T])(\mathrm{E}[T]-\tau)]] = \\
\mathrm{E}[(T-\mathrm{E}[T])^2] + \mathrm{E}[(\mathrm{E}[T]-\tau)^2] + 2\mathrm{E}[(T\mathrm{E}[T]-T\tau-(\mathrm{E}[T])^2+\mathrm{E}[T]\tau)] = \\
\mathrm{E}[(T-\mathrm{E}[T])^2] + \mathrm{E}[(\mathrm{E}[T]-\tau)^2] = \\
\mathrm{VAR}(T) + \mathrm{E}[\mathrm{BIAS}(T)^2] = \mathrm{VAR}(T) + \mathrm{BIAS}(T)^2
\end{aligned}$$

Among the unbiased estimators, the one with the minimum MSE has the smallest variance, i.e. is the most efficient.

Asyptotical unbiasedness. Let $\{T_n\}$ be a sequence of the same estimator T of τ, constructed over samples with different sizes n. Then T is said to be asyptotically unbiased if

$$\lim_{n\to\infty} E[T_n] = \tau$$

i.e. if the bias tends to zero as the sample size n increases.

Consistency. Let $\{T_n\}$ be a sequence of the same estimator T of τ, constructed over samples with different sizes n. Then T is said to be a consistent estimator of τ if the probability of deviations of T_n from τ decreases as n increases, i.e.

$$\lim_{n\to\infty} P[|T_n - \tau| > \epsilon] = 0$$

for any $\epsilon > 0$. Notice that if an estimator is asymptotically unbiased, then it is consistent if

$$\lim_{n\to\infty} VAR[T_n] = 0$$

Notice that any asympotically unbiased estimator can still display positive variance as n increases if it inconsistent.

5.1.1 Example: The sample mean

Consider a series of continuous iid random variable X. Assume μ the expected value of X and σ^2 be the variance. Let $\bar{x}$ be the sample mean of X from a sample of size n.

The expected value of $\bar{x}$ is equal to:

$$\mathrm{E}[\bar{x}] = \mathrm{E}\left[\frac{\sum_{i=1}^{n} x_i}{n}\right] = \frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n} x_i\right] =$$
$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[x_i] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[x_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}n\mu = \mu$$

This implies that the $\bar{x}$ is an unbiased estimator of μ, i.e.

$$\mathrm{BIAS}[\bar{x}] = \mathrm{E}[\bar{x}] - \mu = 0$$

The variance of $\bar{x}$ is equal to:

$$\mathrm{VAR}[\bar{x}] = \mathrm{VAR}\left[\frac{\sum_{i=1}^{n} x_i}{n}\right] = \frac{1}{n^2}\mathrm{VAR}\left[\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{VAR}[x_i] = \frac{\sigma^2}{n}$$

Notice that if X where not iid, then:

$$\mathrm{VAR}[\bar{x}] = \mathrm{VAR}\left[\frac{\sum_{i=1}^{n} x_i}{n}\right] = \frac{1}{n^2}\mathrm{VAR}\left[\sum_{i=1}^{n} x_i\right] = \frac{1}{n^2}\left[\sum_{i=1}^{n}\mathrm{VAR}[x_i] + 2\sum_{i=1}^{n}\sum_{j=i} COV[x_i, x_j]\right]$$

Let s^2 be the sample variance of X from a sample of size n. The expected value of s^2 is equal to

$$\mathrm{E}[s^2] = \mathrm{E}\left[\frac{\sum_{i=1}^{n}[x_i - \mathrm{E}[\bar{x}]]^2}{n-1}\right] = \sigma^2$$

The variance of s^2 is equal to

$$\mathrm{VAR}[s^2] = \mathrm{VAR}\left[\frac{\sum_{i=1}^{n}[x_i - \mathrm{E}[\bar{x}]]^2}{n-1}\right] = \frac{2\sigma^2}{n-1}$$

5.1.2 Theorem: The central limit theorem

Informally, the central limit theorem states that the sample mean of a random sample of n observations drawn from a population with any probability distribution will be approximately normally distributed, if n is large.

Theorem: Let $X_1, X_2, ...$ be i.i.d. random variables obtained by sampling from an arbitrary population. Let $S_n = X_1 + X_2 + ...X_n$. Denote by $\mathrm{E}[S_n]$ and $\mathrm{VAR}[S_n]$

expected value and variance of S_n. Let

$$z_n = \frac{S_n - \mathrm{E}[S_n]}{\sqrt{\mathrm{VAR}[S_n]}}$$

Then, z_n converge in distribution to a standard normal distribution for n that increases.

Corollary The random variable $\frac{S_n}{n}$ converges to a normal distribution with expected value $\frac{1}{n}\mathrm{E}[S_n]$ and variance $\frac{1}{n^2}\mathrm{VAR}[S_n]$.

5.1.3 Theorem: The law of large number

Informally, the law of large number states that, for a given a random sample of size n taken from a population mean, the sample mean will *approach* the population mean as n increases, regardless of the underlying probability distribution of the data.

Theorem: Let $X_1, X_2, ...$ be i.i.d. random variables obtained by sampling from an arbitrary population. Let $\bar{x}_n$ be the sample mean for a sample with size n. Then, for any positive number $\epsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}\left(|\bar{x}_n - \mu| > \epsilon\right) = 0$$

5.2. Confidence Interval

Informally, a confidence interval (CI) indicates a range of values that's likely to encompass the population value. The probability that the confidence interval encompasses the true value is called the confidence level of the CI.
To construct a confidence interval for a population parameter τ at confidence level $\alpha\%$, we have to find the interval where τ will falls into with probability $1-\alpha$. Practically, we identify a sample statistic that we cab use to estimate a population parameter, say $T \sim N(\tau, \sigma_T^2)$.

Case 1 $VAR[T]$ is a known parameter.

Then we compute

$$P(a \le T \le b) = 1 - \alpha$$

To do so, we standardize T and obtain:

$$P\left(\frac{a-\tau}{\sigma_T} \le Z \le \frac{b-\tau}{\sigma_T}\right) = 1 - \alpha$$

or equivalently,

$$P(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}) = 1 - \alpha$$

where $z_{\frac{\alpha}{2}}$ is called critical value of z statistics corresponding to α for a two-tail confidence interval. $z_{\frac{\alpha}{2}}$ can be computed using the statistical tables for a standard normal random variable. Therefore, we get:

$$\begin{aligned}
P\left(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}\right) &= 1 - \alpha \\
P\left(-z_{\frac{\alpha}{2}} \leq \frac{T - \tau}{\sigma_T} \leq z_{\frac{\alpha}{2}}\right) &= 1 - \alpha \\
P\left(-z_{\frac{\alpha}{2}}\sigma_T \leq T - \tau \leq z_{\frac{\alpha}{2}}\sigma_T\right) &= 1 - \alpha \\
P\left(T - z_{\frac{\alpha}{2}}\sigma_T \leq \tau \leq T + z_{\frac{\alpha}{2}}\sigma_T\right) &= 1 - \alpha
\end{aligned}$$

The quantity $z_{\frac{\alpha}{2}}\sigma_T$ is called margin of error.

Case 2 σ_T^2 is a unknown parameter. In this case, we standardize T using an estimator of σ_T^2, i.e. the sample variance s_T^2 and obtain:

$$P\left(\frac{a - \tau}{s_T} \leq Z \leq \frac{b - \tau}{s_T}\right) = 1 - \alpha$$

or equivalently,

$$P(-t_{n-1,\frac{\alpha}{2}} \leq t_{n-1} \leq t_{n-1,\frac{\alpha}{2}}) = 1 - \alpha$$

where t_{n-1} is now a t-student random variable with $n-1$ degrees of freedom, and $t_{n-1,\frac{\alpha}{2}}$ are critical values of t associated to α for a two-tail confidence interval. $t_{n-1,\frac{\alpha}{2}}$ have to be computed using now the statistical tables for a t-student random variable. Therefore in this case, we get:

$$\begin{aligned}
P(-t_{n-1,\frac{\alpha}{2}} \leq t_{n-1} \leq t_{n-1,\frac{\alpha}{2}}) &= 1 - \alpha \\
P\left(T - t_{n-1,\frac{\alpha}{2}}s_T \leq \tau \leq T + t_{n-1,\frac{\alpha}{2}}s_T\right) &= 1 - \alpha
\end{aligned}$$

Property. As n increases, t_{n-1} converges to z. This means that for values $n > 30$, one can use values from the statistical table for z.

5.3. Exercises

Exercise 1 Suppose you are a drink producer and your drinks are sold in 250ml bottles. Due to some imperfections in the production process, the actual volume in each bottle varies. The production process has a mean of 250ml and a standard deviation of 20ml. A consumer watchdog takes a sample of 30 of your bottles and measures their content precisely. You will get bad press if the sample mean is below 245ml. What is the probability that you will get bad press?

Exercise 2 A random sample of 16 bags of a chemical were tested to estimate the mean impurity content. It is known that the impurity content is distributed normally. The sample mean impurity content was 20.4 grams, and the sample standard deviation was 6.4 grams. Find the 95% confidence interval for the population mean.

Exercise 3 An auditor takes a random sample of 400 invoices relating to the activities of a company in a particular year. The sample mean of the invoices is 250GBP and the sample standard deviation is 64GBP. Find a 95% confidence interval for the population mean of the company invoices in the same year.

Chapter 6

Hypothesis testing

In hypothesis testing we assume two possibilities, called hypotheses, about the state of nature. They must be mutually exclusive and collectively exhaustive events. Practically, the true state of nature is never known, but we can use statistical evidence to make judgement and reject one of the hypothesis tested.

One hypothesis is called a null hypothesis and denoted by H_0 and the other is called an alternative hypothesis and denoted by H_1. Suppose we want to test where a population parameters θ is equal to a specific value θ_0 against the opposite hypothesis that θ is different than θ_0. Formally we define null hypothesis and an alternative hypothesis as follows:

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

and we call this hypothesis testing as a two-sided alternative. Alternatively, we could test whether population parameters θ is larger or equal a specific value θ_0 against the opposite hypothesis that θ is lower than θ_0. In this case, we define null hypothesis and an alternative hypothesis as follows:

$$H_0 : \theta \geq \theta_0$$
$$H_1 : \theta < \theta_0$$

and we call this hypothesis testing as a one-sided alternative.

6.1. Test statistics

To determine whether H_0 can be rejected or not we construct a test statistics T, which has a known sampling distribution assuming that the null hypothesis H_0 were true. Given the sampling distribution, we can determine what is the probability of observing a certain outcome of T. If this probability is smaller than a predetermined value, α, called significance level, then we can reject the null hypothesis.

Significance level. The significance level α of a test is also the probability of making what is called *type I error*, i.e. the error of falsely rejecting the null hypothesis when it is true. To minimize the error, the significance level is therefore chosen to be small, usually to larger than 5%.

For a given significance level α, we can find the value, cv, such that, if the null hypothesis were true, obtaining a T value larger than cv has probability α, i.e.

$$P(T > \mathrm{cv}_\alpha | \mathrm{H}_0) = \alpha$$

cv_α are called *critical value* for the hypothesis test. What if the value of our test statistic, T, is larger than the value cv_α? This means there is, at most, an $\alpha 100\%$ chance that the null hypothesis is true therefore we reject the null hypothesis.

Test power. Suppose now that the alternative hypothesis H_1 is true. The power of a test is the probability of rejecting H_0 given that H_1 is true (i.e. the probability of correctly rejecting the null). The larger is the power of a test, the lower are the chances of committing what is called *type II error*, i.e. the error of failing to reject the null hypothesis H_0 when the alternative is true.

We can write the power of a test as

$$P(T > \mathrm{cv}_\alpha | \mathrm{H}_1)$$

for a one-sided test and

$$P(|T| > \mathrm{cv}_{\frac{\alpha}{2}} | \mathrm{H}_1)$$

for a two-sided test. What determines the power of a test? Typically the power increases as 1) the estimator takes values that are different than the value under the null hypothesis and 2) as the sample size n increases. A test is said to be consistent if its power approaches one as n increases.

P-values The p-value or probability value is the probability of obtaining test results at least as extreme as the results actually observed during the test, assuming that the null hypothesis is correct. Let $\hat{T}$ be the observed value for our test statistics T. Then the p-value is equal to i.e.

$$\text{p-value} = P(T > \hat{T} | \mathrm{H}_0)$$

If $\hat{T} = \text{cv}_\alpha$, it must be that p-value $= \alpha$. If $\hat{T} > \text{cv}_\alpha$ it must be that p-value $< \alpha$

6.2. Applications

Two-sided hypothesis test of population mean Consider a normally distributed random variable X, with population mean μ and standard deviation σ. We have shown that the distribution of the sample mean of an independent $x_1, ..., x_n$ is given by

$$\bar{x} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

and we know that

$$\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

where s is an estimator of σ. We want to test the claim that the population mean μ is equal to a certain value μ_0, i.e.

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

The test statistics under the null hypothesis is:

$$T = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Remember that under H_0, $\mu = \mu_0$, and therefore T follows a t_{n-1} distribution. This implies that the critical value is given by

$$\text{cv}_\alpha = t_{\frac{\alpha}{2}, n-1}$$

We reject the null hypothesis, H_0, if $|T| > \text{cv}_\alpha$.

One-sided hypothesis test of population mean It is also possible to compute a one-sided hypothesis test, where we test the alternative hypothesis that μ is either greater than or less than some specified value, i.e. we use either of the following set of hypotheses:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu > \mu_0$$

or

$$H_0 : \mu = \mu_0$$
$$H_2 : \mu < \mu_0$$

In both cases the test statistics under the null hypothesis is equal to the one computed above. Instead the rejection rules are different:

- For H_0 versus H_1, we reject if $T > \text{cv}_\alpha$
- For H_0 versus H_1, we reject if $T < -\text{cv}_\alpha$

Notice finally that the one-sided tests require use of one-sided critical values, $t_{\alpha,n-1}$, instead of $t_{\frac{\alpha}{2},n-1}$

Hypothesis Test of a difference in population means. Suppose that there exist two independent normal random variables X and Y, with a population expected value and standard deviation equal to μ_x and σ_x, μ_y and σ_y, respectively.

An independent random sample of size n_x is drawn from X and of size n_y is drawn from Y. The sample means and sample standard deviations for each sample are given by

- $\bar{x}$ and s_x for X
- $\bar{y}$ and s_y for Y

It follows that $\bar{x} \sim N(\mu_x, \frac{\sigma_x^2}{n_x})$ and $\bar{y} \sim N(\mu_y, \frac{\sigma_y^2}{n_y})$. Since X and Y are independent, then it must be that

$$\bar{x} - \bar{y} \sim \mathcal{N}\left(\mu_x - \mu_y, \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}\right)$$

Suppose now we test the null hypothesis

$$H_0 : \mu_x - \mu_y = \mu_0$$
$$H_1 : \mu_x - \mu_y \neq \mu_0$$

Under the null hypothesis. The test statistics

$$T = \frac{\bar{x} - \bar{y} - \mu_0}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

and since $\mu_x - \mu_y$ follows a normal distribution, T follows a t-distribution with v degree of freedom, equal to

$$v = \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)^2}{\frac{1}{n_x-1}\left(\frac{s_x^2}{n_x}\right)^2 + \frac{1}{n_y-1}\left(\frac{s_y^2}{n_y}\right)^2}$$

An alternative approach is to use the conservative degrees of freedom setting that v is equal to the smaller of $n_x - 1$ or $n_y - 1$.

6.3. Exercises

Exercise 1 A telemarketing group claims that, after training, employees will earn an average of 1500GBP in their first month of work. A random sample of 150 employees is drawn. Sample mean earnings for the first month of work were 1262GBP and the sample standard deviation was 432GBP. Test at the 5% significance level the null hypothesis that the population mean is 1500GBP, against the alternative that it is less than 1,500.

Exercise 2 Explain what the following hypothesis testing terms mean:

- Type I and Type II errors
- Test Power

Exercise 3 A manufacturer states that the mean number of matches in a matchbox is 50. The number of matches in a box is normally distributed. In a random sample of 20 boxes, we find that the sample mean number of matches is 49.3 and the sample standard deviation is 1.25. Test the claim that the population mean is equal to 50, against the alternative that it is not equal to 50, at a 5% level of significance.

Exercise 4 Consider a different manufacturer of matchboxes who also claims that the number of matches in a box has a population mean of 50. The number of matches in a box is normally distributed. In a random sample of 20 boxes, we find that the sample mean number of matches is 50.4 and the sample standard deviation is 1.25. Test the claim that the population mean is equal to 50, against the alternative that it is greater than 50, at a 5% level of significance.

Exercise 5 Now consider a third manufacturer of matchboxes who claims that the number of matches in a box has a population mean of 55. The number of matches in a box is normally distributed. In a random sample of 16 boxes, we find that the sample mean number of matches is 52 and the sample standard deviation is 1.5. Test the claim that the population mean is equal to 55, against the alternative that it is less than 55, at a 5% level of significance.

Exercise 6 A car company owns two factories. The manager of the company wants to know if there is any difference in the mean number of cars produced by each factory per day. She considers a random sample of 30 days. The sample mean output for Factory 1 is 420 and for Factory 2 is 408. The sample standard deviation for Factory 1 is 12.5 and for Factory 2 is 18. Test the hypothesis that Factory 1 and 2 have the same population mean, against the alternative that they have different population means, at a 5% level of significance.

Chapter 7

Linear regression

Consider the following linear relationship, or *population regression equation*:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where $i = 1, 2...n$ are individual observations from a sample, y_i and x_i are, respectively, dependent and explanatory variables, and where ϵ_i is an iid random error term. Finally the parameters β_0 and β_1 are unknown and have to be estimated.

7.1. OLS estimators

The Ordinary Least Squares (OLS) estimators of β_0 and β_1, $\hat{\beta}_0$ and $\hat{\beta}_1$, are obtained by fitting a line through the data minimising the sum of squared residuals, i.e.

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \quad S(\hat{\beta}_0, \hat{\beta}_1) = \min_{\hat{\beta}_0, \hat{\beta}_1} \quad \sum_{i=1}^{n} \hat{\epsilon_i}^2$$

where

$$\hat{\epsilon_i} = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

To derive the OLS estimator, we need to solve the following First Order Conditions (FOC), i.e.

$$\begin{aligned}
\frac{\partial S(\hat{\beta}_0, \hat{\beta}_1)}{\partial \hat{\beta}_0} = 0 \implies -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\
\frac{\partial S(\hat{\beta}_0, \hat{\beta}_1)}{\partial \hat{\beta}_1} = 0 \implies -2\sum_{i=1}^{n}[x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)] = 0
\end{aligned}$$

From the first FOC, we get:

$$
\begin{aligned}
-2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \implies \\
\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \implies \\
\sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\hat{\beta}_0 - \sum_{i=1}^{n}\hat{\beta}_1 x_i = 0 \implies \\
\sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1\sum_{i=1}^{n} x_i = 0 \implies \\
\hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{\beta}_1\frac{1}{n}\sum_{i=1}^{n} x_i \implies \\
\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}
\end{aligned}
$$

where $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Substituting the solution for $\hat{\beta}_0$ into the second FOC, we get:

$$
\begin{aligned}
-2\sum_{i=1}^{n}[x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)] = 0 \implies \\
\sum_{i=1}^{n}[x_i(y_i - (\overline{y} - \hat{\beta}_1\overline{x}) - \hat{\beta}_1 x_i)] = 0 \implies \\
\sum_{i=1}^{n}[x_i(y_i - \overline{y}) - \hat{\beta}_1 x_i(x_i - \overline{x})] = 0 \implies \\
\sum_{i=1}^{n}\hat{\beta}_1 x_i(x_i - \overline{x}) = \sum_{i=1}^{n} x_i(y_i - \overline{y}) \implies \\
\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i(y_i - \overline{y})}{\sum_{i=1}^{n} x_i(x_i - \overline{x})}
\end{aligned}
$$

Notice that:

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x}) = \sum_{i=1}^{n}(x_i x_i - 2\overline{x}x_i + \overline{x}\overline{x}) = \\
\sum_{i=1}^{n} x_i x_i - 2\sum_{i=1}^{n}\overline{x}x_i + \sum_{i=1}^{n}\overline{x}\overline{x} = \\
\sum_{i=1}^{n} x_i x_i - \overline{x}\sum_{i=1}^{n} x_i - \overline{x}\sum_{i=1}^{n} x_i + n\overline{x}\overline{x} = \\
\sum_{i=1}^{n} x_i x_i - \overline{x}\sum_{i=1}^{n} x_i - \overline{x}\sum_{i=1}^{n} x_i + \overline{x}\sum_{i=1}^{n} x_i = \\
\sum_{i=1}^{n} x_i x_i - \overline{x}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i(x_i - \overline{x})
\end{aligned}
$$

By the same argument,

$$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n} x_i(y_i - \overline{y})$$

Substituting back into the formula for $\hat{\beta}_1$, we get:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i(y_i - \overline{y})}{\sum_{i=1}^{n} x_i(x_i - \overline{x})} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x})} = \frac{\mathrm{cov}[x, y]}{\mathrm{var}[x]}$$

i.e. $\hat{\beta}_1$ is the ratio of the sample covariance between x and y to the sample variance of x. Since the variance is always positive, $\hat{\beta}_1$ will have the same sign as the covariance (and the correlation coefficient). Plugging $\hat{\beta}_1$ into $\hat{\beta}_0$, we get

$$\hat{\beta}_0 = \overline{y} - \frac{\mathrm{cov}[x, y]}{\mathrm{var}[x]}\overline{x}$$

7.1.1 Properties

Unbiasedness The OLS estimators are unbiased, i.e.

$$\mathrm{E}[\hat{\beta}_0] = \beta_0$$
$$\mathrm{E}[\hat{\beta}_1] = \beta_1$$

The Gauss-Markov Theorem. Among all linear unbiased estimators, OLS estimators have the smallest variance, i.e. OLS estiamators are efficient estimators. The variance of any other linear unbiased estimator must be larger than that of the OLS estimator. We say that OLS estimators are BLUE (best linear unbiased estimator)

Consistency The OLS estimators are consistent. Since they are also unbiased, this implies that:

$$\lim_{n\to\infty} \mathrm{VAR}[\hat{\beta}_0] = 0$$
$$\lim_{n\to\infty} \mathrm{VAR}[\hat{\beta}_1] = 1$$

7.1.2 Goodness of Fit

We can measure the goodness of fit of a regression model by comparing the variation that has been explained by the model to the total variation in the data. We define the Total Sum of Squares (TSS) to be the total (squared) variation of the y_i values about their mean $\bar{y}$, such that

$$\mathrm{TSS} = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

and we define the Explained Sum of Squares (ESS) to be the total (squared) variation of the fitted values $\hat{y}_i$ about their mean $\bar{y}$, such that

$$\text{ESS} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

Finally, we define the Residual Sum of Squares (RSS) to be the total (squared) unexplained variation in our model, i.e. the total squared difference between the y_i values and the fitted $\hat{y}_i$ values,

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (\epsilon_i)^2$$

It can be shown that $\text{TSS} = \text{ESS} + \text{RSS}$. Intuitively, we can think of the total variation of the data (TSS) being decomposed into a part that is explained by the fitted model (ESS) and a part that is unexplained by the fitted model (RSS).
Therefore, a measure of fitness is

$$\text{R}^2 = \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}$$

R^2 is called coefficient of determination. Notice that

$$0 \leq \text{ESS} \leq \text{TSS} \quad \text{and} \quad 0 \leq \text{RSS} \leq \text{TSS} \implies 0 \leq \text{R}^2 \leq 1$$

It can be shown that R^2 is equivalent to the squared sample correlation coefficient between y and x, i.e.

$$\text{R}^2 = \left(\frac{\text{COV}[x, y]}{s_x s_y} \right)^2$$

7.2. Hypothesis Testing of Regression Parameters

Let the regression model be:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Suppose we want to test claims made about the parameters β_1[1]. In particular, consider testing the claim that there is no relationship between x and y, i.e.

$$\text{H}_0 : \beta_1 = 0 \quad \text{versus} \quad \text{H}_0 : \beta_1 \neq 0$$

Assumption: $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

This assumption ensure that y_i is normally distributed as well. Since estimator $\hat{\beta}_i$ is

[1] The same argument applies for β_0

a linear function of y_i, it must also follow a normal distribution. Therefore,

$$\hat{\beta}_i \sim \mathcal{N}(\beta_1, \mathrm{VAR}[\beta_1])$$

and we can standardise $\hat{\beta}_i$ in the usual way (by subtracting its mean and dividing by its standard deviation), such that:

$$\frac{\hat{\beta}_i - \beta_1}{\sqrt{\mathrm{VAR}[\beta_1]}} \sim \mathcal{N}(0, 1)$$

It turns out that $\mathrm{VAR}[\beta_1]$, is a function of σ^2 which is unknown. Replacing σ^2 with a sample estimator changes the distribution of the above statistic from standard normal to Student's, i.e.

$$\frac{\hat{\beta}_i - \beta_1}{\sqrt{\widehat{\mathrm{VAR}}[\beta_1]}} \sim t_{n-2}$$

where $\widehat{\mathrm{VAR}}[\beta_1]$ is an estimator of $\mathrm{VAR}[\beta_1]$. This statistic follows a t-distribution with $n-2$ degrees of freedom (instead of $n-1$ in the population mean case) This is because we now estimate two parameters instead of one. Therefore, to test our null hypothesis against the alternative, we have to construct the test statistic, i.e.

$$\frac{\hat{\beta}_i}{\sqrt{\widehat{\mathrm{VAR}}[\beta_1]}}$$

and compare it to critical values from the $t_{\frac{\alpha}{2},t-2}$ distribution. Notice that we can also consider one-sided alternatives of the form

$$\mathrm{H}_0 : \beta_1 = 0 \quad \text{versus} \quad \mathrm{H}_1 : \beta_1 \geq 0$$

and test the hull hypothesis against this alternative using critical values from $t_{\alpha,t-2}$.

7.3. Exercises

Exercise 1 The weight (in kg) and height (in cm) of a sample of 5 individuals is reported below:

Individual	Weight	Height
1	83	181
2	70	175
3	63	165
4	82	193
5	75	178

Calculate the OLS estimates of the following linear regression model:

$$\text{weight}_i = \beta_0 + \beta_1 \text{height}_i + \epsilon_i$$

Exercise 2 The campaign manager for a local politician wants to know if the number of leaflets delivered during an election campaign affects the total number of votes received by a candidate standing for election. She collects data from the last 5 elections. She denotes the number of leaflets delivered as x and the number of votes the candidate received as y. She finds the following results:

$$\begin{aligned} \bar{x} &= 1280 \\ \bar{y} &= 1980 \\ VAR[x] &= 44.5 \\ COV[x, y] &= 25.2 \end{aligned}$$

Calculate the Ordinary Least Squares estimates of the parameters β_0 and β_1 in the following linear regression:

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t$$

Exercise 3 Consider the sample data and fitted regression considered in Exercise 2. You are now told that $VAR[y] = 29.7$. Compute R^2 for this regression.

Exercise 4 An economist collects data on GDP (x) and literacy rates (y) for a sample of 30 countries and estimates the following fitted regression model

$$y_i = 12 + 0.2x_i + \hat{\epsilon}_i$$

where the standard errors of $\hat{\beta}_0$ and $\hat{\beta}_1$ are 2.3 and 0.08 respectively.

- Test the claim that there is no relationship between GDP and literacy rates at a 5% significance level, against a two-sided alternative.
- Test the claim that β_1 is equal to 0.3 against the one-sided alternative that β_1 is less than 0.3 at a 5% level of significance

Bibliography

1. Newbold, P., Carlson, W. L., and Thorne, B. (2013). Statistics for business and economics. Boston, MA: Pearson.

2. Gujarati, D. N., and Porter, D. C. (1999). Essentials of econometrics (Vol. 2). Singapore: Irwin/McGraw-Hill.

3. Barrow, M. (2009). Statistics for economics, accounting and business studies. Pearson Education.

4. Salvatore, D., and Reagle, D. (2002). Statistics and Econometrics, Schaum's Outline Series.