STAT10430 - Statistics with Python
Descriptive Statistics

Dr. Áine Byrne
aine.byrne@ucd.ie

What is Statistics?

Statistics is the science of data. It concerns the collection, classification, organisation, analysis, interpretation and presentation of information.

Statistical Process

1. Hypothesis
2. Study Design
3. Collect Data
4. Analyse Data
    - Descriptive Statistics
    - Inference
5. Present Results

Population and Samples

- **Population:** Entire set of objects of interest.
- **Sample:** Subset of the set of objects of interest.
- e.g. Interested in the average height of UCD students.
    - Population: Entire UCD student body.
    - Sample: Students in one lecture theatre.
- We analyse samples to make inferences about a population.
- This is **statistical inference**.

Study Types

Designed Experiment

Researcher exerts control over the experimental units. e.g. Tensile strength of beams randomly assigned to different treatments.

Observational Study

Researcher observes the experimental units and records the variables of interest. e.g. RPM of an engine at the failure time of a component.

Study Types

- In a designed experiment we create differences in the explanatory variable and then examine the results. In an observational study we observe differences in the explanatory variable and then notice whether these are related to differences in the response variable.
- Experimental studies are complex to implement, but they are more informative. There can be ethical issues with experimental studies.

Data Types

Quantitative Data

Measurements that are recorded on a natural numerical scale.

Data Types

Quantitative Data

Measurements that are recorded on a natural numerical scale.

- Continuous Data are measurements which can fall anywhere on the real line. (Or an interval of the real line)
  e.g. Height, weight, volume . . .

Data Types

Quantitative Data

Measurements that are recorded on a natural numerical scale.

- Continuous Data are measurements which can fall anywhere on the real line. (Or an interval of the real line)
  e.g. Height, weight, volume ...

- Discrete Data are measurements which can only take one of a finite set of values.
  e.g. Number of server crashes in a week, number of coin flips until a head is observed, ...

Data Types

Qualitative Data (Categorical)

Measurements that cannot be recorded on a natural numerical scale.

Data Types

Qualitative Data (Categorical)

Measurements that cannot be recorded on a natural numerical scale.

- Nominal Data is qualitative data with no meaningful ordering. e.g. Eye colour, (Blue, Green, Brown...).

Data Types

Qualitative Data (Categorical)

Measurements that cannot be recorded on a natural numerical scale.

- Nominal Data is qualitative data with no meaningful ordering. e.g. Eye colour, (Blue, Green, Brown...).
- Ordinal Data is qualitative data which has an inherent order. e.g. Your grade at the end of the semester (B, B+, A-, etc.).

Data Types

Qualitative Data (Categorical)

Measurements that cannot be recorded on a natural numerical scale.

- Nominal Data is qualitative data with no meaningful ordering. e.g. Eye colour, (Blue, Green, Brown...).
- Ordinal Data is qualitative data which has an inherent order. e.g. Your grade at the end of the semester (B, B+, A-, etc.).
- Qualitative data is very common in survey data.
- Note that there is no concept of distance with ordinal data.

Data Types

What data types do the following have?

1. Number of students in UCD
2. Types of trees
3. Exchange rates
4. App ratings
5. Months of the year
6. Leaving Cert points
7. Daily temperature
8. Price of an apple
9. Age

Warning!

- The distinction between categorical or numerical data can be difficult to determine in some cases.
- Consider the app rating question on the previous slide.

Warning!

- The distinction between categorical or numerical data can be difficult to determine in some cases.
- Consider the app rating question on the previous slide.
- The ratings take the values 1,2,3,4,5. *It could be argued that these are a categorical value in disguise!*

Warning!

- The distinction between categorical or numerical data can be difficult to determine in some cases.
- Consider the app rating question on the previous slide.
- The ratings take the values 1,2,3,4,5. *It could be argued that these are a categorical value in disguise!*
- This issue can arise in a lot of survey applications. *For example someone might code*
    - 1 – Strongly Disagree
    - 2 – Disagree
    - 3 – No opinion
    - 4 – Agree
    - 5 – Strongly Agree

Warning!

- The distinction between categorical or numerical data can be difficult to determine in some cases.
- Consider the app rating question on the previous slide.
- The ratings take the values 1,2,3,4,5. *It could be argued that these are a categorical value in disguise!*
- This issue can arise in a lot of survey applications. *For example someone might code*
  - 1 – Strongly Disagree
  - 2 – Disagree
  - 3 – No opinion
  - 4 – Agree
  - 5 – Strongly Agree
- This is called a Likert scale.

Warning 2!

- The distinction between discrete and continuous numerical data can be ambiguous too.

Warning 2!

- The distinction between discrete and continuous numerical data can be ambiguous too.
    1. Suppose that the weight of an item is recorded to the nearest kilogram.
    2. Suppose the area of a piece of land is recorded to the nearest cm^2.

Warning 2!

- The distinction between discrete and continuous numerical data can be ambiguous too.
  1. Suppose that the weight of an item is recorded to the nearest kilogram.
  2. Suppose the area of a piece of land is recorded to the nearest cm^2.
- It could be argued that although weight/land area is continuous it is recorded in a discrete manner.
- However, it is always possible to say one object is heavier than the other or one land area is greater than the other. The chances of them being identical is infinitesimally small.

Descriptive Statistics

Numerical Summaries

- These report some numbers that provide information about the data
- There are many numerical summaries but the two main types are:
  - Measures of location: Where is the data “centered”?
  - Measures of spread: How spread out are the data?

Descriptive Statistics

Numerical Summaries

- These report some numbers that provide information about the data
- There are many numerical summaries but the two main types are:
  - Measures of location: Where is the data "centered"?
  - Measures of spread: How spread out are the data?

Graphical Summaries

- These provide a pictorial summary of the data. *"A picture is worth a thousand words."*
  - Barchart: Shows the distribution of categorical values.
  - Histogram: Shows the distribution of numerical values.
  - Scatter plot: Shows the relationship between variables

Examples: Graphical Summaries

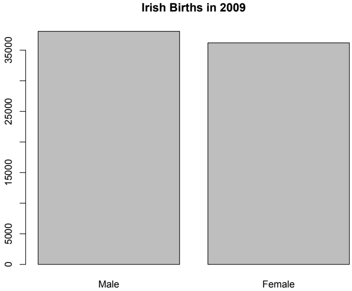
- It's worth keeping an eye out for good and bad examples of plots.
- Good places to start on the web are:
  - Hans Rosling's talks (`https://www.gapminder.org/videos/the-joy-of-stats/`)
  - Lecture by Ross Ihaka (`https://www.stat.auckland.ac.nz/~ihaka/120/Lectures/lecture03.pdf`)
  - Andrew Gelman's blog (`https://statmodeling.stat.columbia.edu/`)
  - Edward Tufte's webpage (`http://www.edwardtufte.com`)
  - Media sources like the Economist and the NY Times.

Barcharts (Example: Births)

- There were 74278 children born in Ireland in 2009.
- How did it break down into Male/Female?

Barcharts (Example: Births)

- There were 74278 children born in Ireland in 2009.
- How did it break down into Male/Female?



Barcharts

- These provide a graphical summary of **categorical data**.
- They are very easy to produce:
  1. Count the number of observations in each category (frequency).
  2. Draw a plot where each category is an equal width rectangle and the height is proportional to frequency.

Barcharts

- These provide a graphical summary of **categorical data**.
- They are very easy to produce:
  1. Count the number of observations in each category (frequency).
  2. Draw a plot where each category is an equal width rectangle and the height is proportional to frequency.
- The height of each bar could also be the relative frequency,

$$RF = \frac{Frequency}{N}$$

where N is the total number of observations.

Barcharts

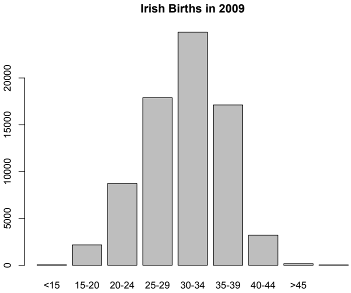
- These provide a graphical summary of **categorical data**.
- They are very easy to produce:
  1. Count the number of observations in each category (frequency).
  2. Draw a plot where each category is an equal width rectangle and the height is proportional to frequency.
- The height of each bar could also be the relative frequency,

$$RF = \frac{Frequency}{N}$$

  where N is the total number of observations.
- If your data is ordinal make sure that the rectangles are in a sensible order. *Warning: Software may reorder the categories into alphabetical order*

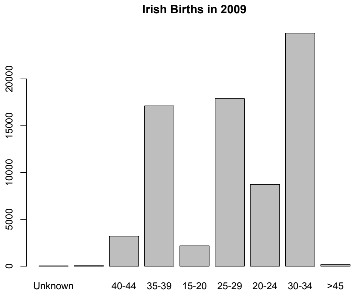
Example: Births 2

- If we look at the age of the mother, coded into categorical bands.



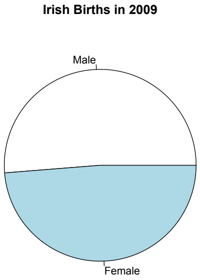
Example: Births 3

- This plot is technically equivalent to the previous one, but it's not very good!



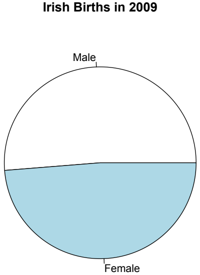
Piechart

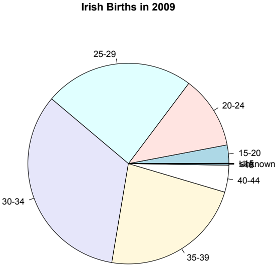
- A piechart is an alternative plot. The area of each slice (equivalently angle) is proportional to frequency.



Piechart

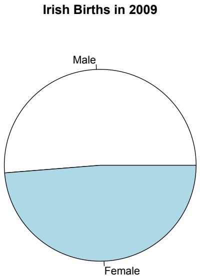
- A piechart is an alternative plot. The area of each slice (equivalently angle) is proportional to frequency.

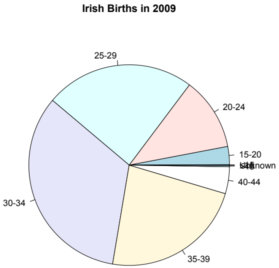




Piechart

- A piechart is an alternative plot. The area of each slice (equivalently angle) is proportional to frequency.





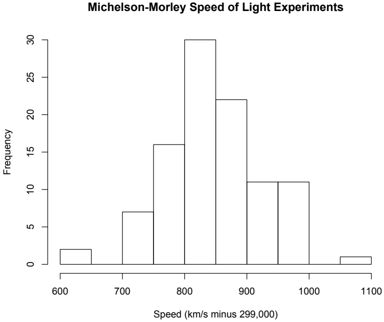
- Lots of people argue that they're a poor plot choice (https://visuanalyze.wordpress.com/2013/02/05/bad_piecharts/).

Histogram (Example: Speed of Light)

- In 1887 Michelson and Morley conducted five experiments to determine the speed of light. In each experiment they made twenty measurements of the speed of light.

Histogram (Example: Speed of Light)

- In 1887 Michelson and Morley conducted five experiments to determine the speed of light. In each experiment they made twenty measurements of the speed of light.
- A histogram of the one hundred measurements is as follows:



Histogram

- Technically, the barchart for the mother's age is a histogram.
- Histograms are used to represent numerical data.

Histogram

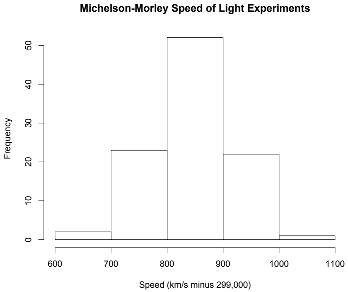
- Technically, the barchart for the mother's age is a histogram.
- Histograms are used to represent numerical data.
  1. Divide the range of the data into bins. *The bins are usually, but not necessarily equal width.*
  2. Count how many observations fall into each bin.
  3. Plot a rectangle for each bin, where the base length is proportional to the bin width and the **area** of the rectangle is proportional to frequency.

Example: Other Histograms

- The number of bins in a histogram can change its appearance.

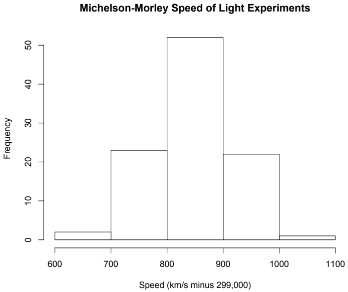
Example: Other Histograms

- The number of bins in a histogram can change its appearance.

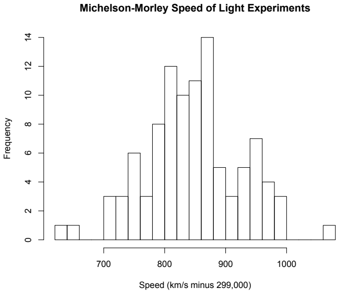


Example: Other Histograms

- The number of bins in a histogram can change its appearance.

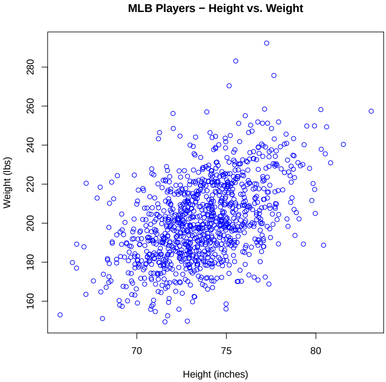


Michelson-Morley Speed of Light Experiments



Michelson-Morley Speed of Light Experiments

Scatterplot



Scatterplots

- If we have measurements on two variables, plotting one against the other can be useful for finding patterns.

Scatterplots

- If we have measurements on two variables, plotting one against the other can be useful for finding patterns.
- Scatterplots can be a useful graphical representation of the relationship between variables.

Scatterplots

- If we have measurements on two variables, plotting one against the other can be useful for finding patterns.
- Scatterplots can be a useful graphical representation of the relationship between variables.
- These plots will be very useful for regression analysis. (Later in the course)

Numerical Summaries

- We have two main types of numerical summaries:
  1. Measures of location
  2. Measures of spread

Numerical Summaries

- We have two main types of numerical summaries:
  1. Measures of location
  2. Measures of spread
- Some summaries (eg. quantiles) have a dual purpose.

Numerical Summaries

- We have two main types of numerical summaries:
  1. Measures of location
  2. Measures of spread
- Some summaries (eg. quantiles) have a dual purpose.
  1. Measures of location (or central tendency) give an idea of where the data are "centered".
  2. Measures of spread give an idea of the range of "most" of the data.

Location

- The **mean** or **average** is a measure of central tendency.

Location

- The **mean** or **average** is a measure of central tendency.
- Suppose we have a sample of size n from a population of size N:

$$x_1, x_2, \ldots, x_n$$

- The **population mean** is

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i.$$

Location

- The **mean** or **average** is a measure of central tendency.
- Suppose we have a sample of size n from a population of size N:

$$x_1, x_2, \ldots, x_n$$

- The **population mean** is

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i.$$

- The **sample mean** is

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

Location

- The **mean** or **average** is a measure of central tendency.
- Suppose we have a sample of size n from a population of size N:

$$x_1, x_2, \dots, x_n$$

- The **population mean** is

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i .$$

- The **sample mean** is

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

- The sample mean is sensitive to outliers.

Median

- The **median** is the middle value.

Median

- The **median** is the middle value.
- The **population median** has the same number of values greater or less than it.
- The **sample median** has the same number of values greater or less than it.

Median

- The **median** is the middle value.
- The **population median** has the same number of values greater or less than it.
- The **sample median** has the same number of values greater or less than it.
- In the case of an odd number of values, the sample median corresponds to the position $\frac{(n+1)}{2}$ in the ordered data.

Median

- The **median** is the middle value.
- The **population median** has the same number of values greater or less than it.
- The **sample median** has the same number of values greater or less than it.
- In the case of an odd number of values, the sample median corresponds to the position $\frac{(n+1)}{2}$ in the ordered data.
- In the case of an even number of values, there isn't a unique sample median. The convention is to find the average of observations in position $\frac{n}{2}$ and $\frac{(n+2)}{2}$ in the ordered data.

Median

- The **median** is the middle value.
- The **population median** has the same number of values greater or less than it.
- The **sample median** has the same number of values greater or less than it.
- In the case of an odd number of values, the sample median corresponds to the position $\frac{(n+1)}{2}$ in the ordered data.
- In the case of an even number of values, there isn't a unique sample median. The convention is to find the average of observations in position $\frac{n}{2}$ and $\frac{(n+2)}{2}$ in the ordered data.
- The sample median is insensitive to outliers

Mode

- The **mode** is the most common value.
- For numerical values, the mode may not be well defined.

Mode

- The **mode** is the most common value.
- For numerical values, the mode may not be well defined.
- It is possible to have more than one mode.

Mode

- The **mode** is the most common value.
- For numerical values, the mode may not be well defined.
- It is possible to have more than one mode.
- The word “mode” is also used to refer to the highest point in the histogram.

Mode

- The **mode** is the most common value.
- For numerical values, the mode may not be well defined.
- It is possible to have more than one mode.
- The word “mode” is also used to refer to the highest point in the histogram.
- The **sample mean**, **sample median** and **mode** are a examples of **statistics**, numbers calculated from the sample data which in some way summaries the data.

Example: Failure Times

Twenty four pumps were tested until they failed. The time to failure was recorded.

```
 6.0    8.6  17.8  18.0  27.5  33.5  50.5  51.5
69.0  74.0  74.0  89.0 109.0 118.0 119.0 138.0
141.0 144.0 146.0 150.0 151.0 153.0 153.1 153.2
```

- How can we summarise the data?

Example: CPI

The consumer price index was recorded for ten consecutive years.

```
3.295837 3.295837 3.401197 3.610918 3.871201
3.850148 3.784190 3.761200 3.713572 3.688880
```

- How can we summarise the data?

Example: Old Faithful Geyser

The time (in minutes) between eruptions of the Old Faithful geyser in Yellowstone national park were recorded for 272 eruptions.

79 54 74 62 85 55 88 85 51 . . . 60 75 81 46 90 46 74

- How can we represent the distribution of the time between eruptions?
- What characteristics does the distribution have?

Example: Speed of Light

- Michelson and Morley conducted five experiments in 1887 to study the speed of light.

- Their experiments yielded 100 measurements of the speed of light.
- How can we represent their measurements?

Measures of Spread

- Measures of spread give a numerical summary of the “typical range” of the data.

Measures of Spread

- Measures of spread give a numerical summary of the “typical range” of the data.
- A number of alternatives exist including:

Measures of Spread

- Measures of spread give a numerical summary of the “typical range” of the data.
- A number of alternatives exist including:
  1. **Range:** This records the full spread of the data.

Measures of Spread

- Measures of spread give a numerical summary of the "typical range" of the data.
- A number of alternatives exist including:
    1. **Range:** This records the full spread of the data.
    2. **Interquartile Range:** This records the spread of the middle of the data.

Measures of Spread

- Measures of spread give a numerical summary of the "typical range" of the data.
- A number of alternatives exist including:
  1. **Range:** This records the full spread of the data.
  2. **Interquartile Range:** This records the spread of the middle of the data.
  3. **Standard Deviation:** This measures how values differ from the mean value.

Range

- The **range** records the difference between the **minimum** and the **maximum** value.
- This is a good statistic because it records the full range of the data.

Range

- The **range** records the difference between the **minimum** and the **maximum** value.
- This is a good statistic because it records the full range of the data.
- However, it is **highly sensitive** to the extreme values in the data.

Range

- The **range** records the difference between the **minimum** and the **maximum** value.
- This is a good statistic because it records the full range of the data.
- However, it is **highly sensitive** to the extreme values in the data.

Data	Min	Max	Range
Failure	6	153.2	147.2
CPI	3.295837	3.850148	0.554311
Michelson-Morley	620	1070	450
Old Faithful	43	96	53

Range

- The **range** records the difference between the **minimum** and the **maximum** value.
- This is a good statistic because it records the full range of the data.
- However, it is **highly sensitive** to the extreme values in the data.

Data	Min	Max	Range
Failure	6	153.2	147.2
CPI	3.295837	3.850148	0.554311
Michelson-Morley	620	1070	450
Old Faithful	43	96	53

- Note that the range has the same “units” as the data.

Interquartile Range

- The **interquartile range (IQR)** records the difference between the 75% percentile and the 25% percentile.

Interquartile Range

- The **interquartile range (IQR)** records the difference between the 75% percentile and the 25% percentile.
- The 25% percentile (also known as the 1st Quartile) has 25% of the values less than it.
- The 75% percentile (also known as the 3rd Quartile) has 75% of the values less than it.
- It records the range of the middle 50% of the data.

Interquartile Range

- The **interquartile range (IQR)** records the difference between the 75% percentile and the 25% percentile.
- The 25% percentile (also known as the 1st Quartile) has 25% of the values less than it.
- The 75% percentile (also known as the 3rd Quartile) has 75% of the values less than it.
- It records the range of the middle 50% of the data.
- It is very robust to extreme values in the data.

Interquartile Range

- The **interquartile range (IQR)** records the difference between the 75% percentile and the 25% percentile.
- The 25% percentile (also known as the 1st Quartile) has 25% of the values less than it.
- The 75% percentile (also known as the 3rd Quartile) has 75% of the values less than it.
- It records the range of the middle 50% of the data.
- It is very robust to extreme values in the data.

Data	Q1	Q3	IQR
Failure	46.25	144.50	98.25
CPI	3.453627	3.871201	0.3248153
Michelson-Morley	807.5	892.5	85
Old Faithful	58	82	24

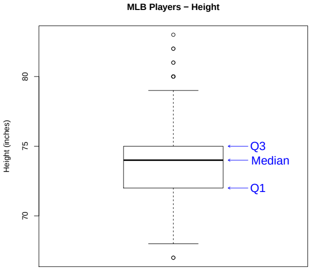
- Note that the IQR has the same "units" as the data.

Box Plots

- Box plots provide a useful graphical summary of the data.

Box Plots

- Box plots provide a useful graphical summary of the data.
- They show where there data are located and also indicate the spread of the data.



Standard Deviation

- The **standard deviation** records the root mean squared deviation of values from the mean.

Standard Deviation

- The **standard deviation** records the root mean squared deviation of values from the mean.
- The **population standard deviation** is

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}.$$

Standard Deviation

- The **standard deviation** records the root mean squared deviation of values from the mean.
- The **population standard deviation** is

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}.$$

- The **sample standard deviation** is

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}.$$

  - The reason for the mysterious -1 will be clear later!
- The larger the value of the standard deviation, the greater is the spread of data.

Standard Deviation (2)

- For our data we get:

Data	SD
Failure	54.1
CPI	0.219
Michelson-Morley	79.0
Old Faithful	13.5

- Note that the standard deviation (SD) has the same “units” as the data.

Standard Deviation (2)

- For our data we get:

Data	SD
Failure	54.1
CPI	0.219
Michelson-Morley	79.0
Old Faithful	13.5

- Note that the standard deviation (SD) has the same "units" as the data.

- The **variance** is the squared standard deviation.
  It does not have the same units!

Using Summaries in Conjunction

- In many situations we find that “most” of the data lie in the range

$$\text{Median} \pm 1.5\text{IQR}$$

and/or

$$\overline{x} \pm 2s$$

Using Summaries in Conjunction

- In many situations we find that “most” of the data lie in the range

$$\text{Median} \pm 1.5\text{IQR}$$

and/or

$$\overline{x} \pm 2s$$

- Later, we will see a theorem that explains why 95% of the data lie in the second range.

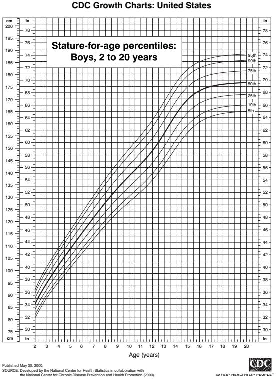
Percentiles

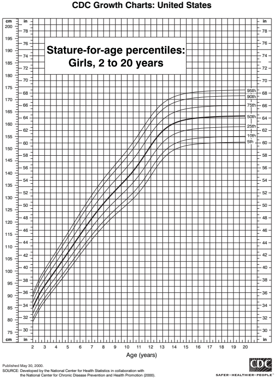
- **Percentiles** are used to give a value where a particular percentage of a sample (or population) fall below this value.

Percentiles

- **Percentiles** are used to give a value where a particular percentage of a sample (or population) fall below this value.
- The 1% percentile will have 1% of values below it and 99% of values above it.
- The 90% percentile will have 90% of values below it and 10% of values above it.

Growth Charts





STAT10430 - Statistics with Python
Probability

Dr. Áine Byrne
aine.byrne@ucd.ie

Overview

- Elementary probability.
- Events; Independence; Mutually exclusivity.
- Additive rule; Multiplicative rule.
- Conditional probability; Bayes Theorem.
- Multiplicative counting rule; Combination rule; Permutations rule.

Running example

A fair die is rolled once. The possible outcomes are:

$$\{1, 2, 3, 4, 5, 6\}$$

What is the probability of rolling 3?

Preliminary definitions

Experiment

An experiment is an act or process of observation that leads to a single outcome that cannot be predicted with certainty.

- Example: The experiment is rolling the die.

Preliminary definitions

Experiment

An experiment is an act or process of observation that leads to a single outcome that cannot be predicted with certainty.

- Example: The experiment is rolling the die.

Sample point

A sample point is the most basic outcome of an experiment

- Example: There are 6 sample points in this experiment:

$$1, 2, 3, 4, 5, 6.$$

Preliminary definitions

Sample Space

The sample space is the set of **ALL** sample points.

- Example: The sample space of this experiment is:

$$\{1, 2, 3, 4, 5, 6\}.$$

Preliminary definitions

Sample Space

The sample space is the set of **ALL** sample points.

- Example: The sample space of this experiment is:

$$\{1, 2, 3, 4, 5, 6\}.$$

Probability Axioms

Let p_i denote the probability of sample point i and let S be the sample space:

- $0 \leq p_i \leq 1$.
- $\mathbb{P}(S) = 1$.
- The probabilities of **all** sample points within a sample space must sum to 1.

$$\sum_i p_i = 1$$

Events

Event

An event is a collection of sample points. It is a subset of the sample space S.

- An event $\mathcal{A}$ occurs if any one of the sample points in $\mathcal{A}$ occur.
- Example: an event may be rolling an even number. If we roll, say, 2, the event occured.

Events

Event

An event is a collection of sample points. It is a subset of the sample space S.

- An event $\mathcal{A}$ occurs if any one of the sample points in $\mathcal{A}$ occur.
- Example: an event may be rolling an even number. If we roll, say, 2, the event occured.

Probability of an event.

The probability of an event $\mathcal{A}$ is calculated by summing the probabilities of the sample points in $\mathcal{A}$.

- Example: the probability of rolling an even number is

$$\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{1}{2}$$

Example: Loaded Die

- A die has been *loaded* so that the probability of side i coming up is proportional to i.
- What is the probability of rolling 3?
- If $\mathcal{A}$ is the event that either a 2 or a 3 comes up.
- What is $\mathbb{P}(\mathcal{A})$?

Set notation

An event that can be viewed as the composition as 2 or more events is called a compound event.

Union

The union of two events $\mathcal{A}$ and $\mathcal{B}$ is the event that occurs if either $\mathcal{A}$ or $\mathcal{B}$ (or both) occur. Denoted $\mathcal{A} \cup \mathcal{B}$.

- Example: $\mathcal{A}$ is the event that an even number is rolled, and $\mathcal{B}$ that a multiple of 3 is rolled. The sample space of $\mathcal{A} \cup \mathcal{B}$ is $\{2, 3, 4, 6\}$.

Set notation

An event that can be viewed as the composition as 2 or more events is called a compound event.

Union

The union of two events $\mathcal{A}$ and $\mathcal{B}$ is the event that occurs if either $\mathcal{A}$ or $\mathcal{B}$ (or both) occur. Denoted $\mathcal{A} \cup \mathcal{B}$.

- Example: $\mathcal{A}$ is the event that an even number is rolled, and $\mathcal{B}$ that a multiple of 3 is rolled. The sample space of $\mathcal{A} \cup \mathcal{B}$ is $\{2, 3, 4, 6\}$.

Intersection

The intersection of two events $\mathcal{A}$ and $\mathcal{B}$ is the event that occurs if both $\mathcal{A}$ and $\mathcal{B}$ occur. Denoted $\mathcal{A} \cap \mathcal{B}$.

- Example: For $\mathcal{A}$ and $\mathcal{B}$ given above, the sample space for $\mathcal{A} \cap \mathcal{B}$ is $\{6\}$.

Additive law

The probability of the union of events may be calculated without knowing the individual sample point probabilities.

Additive law

The probability of the union of two events $\mathcal{A}$ and $\mathcal{B}$ is:

$$\mathbb{P}(\mathcal{A} \cup \mathcal{B}) = \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{B}) - \mathbb{P}(\mathcal{A} \cap \mathcal{B})$$

Mutually exclusive events

Two events $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive if they cannot occur at the same time.

Additive law

The probability of the union of events may be calculated without knowing the individual sample point probabilities.

Additive law

The probability of the union of two events $\mathcal{A}$ and $\mathcal{B}$ is:

$$\mathbb{P}(\mathcal{A} \cup \mathcal{B}) = \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{B}) - \mathbb{P}(\mathcal{A} \cap \mathcal{B})$$

Mutually exclusive events

Two events $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive if they cannot occur at the same time.

- We can write $\mathcal{A} \cap \mathcal{B} = \emptyset$ when the events are mutually exclusive.

Additive law

The probability of the union of events may be calculated without knowing the individual sample point probabilities.

Additive law

The probability of the union of two events $\mathcal{A}$ and $\mathcal{B}$ is:

$$\mathbb{P}(\mathcal{A} \cup \mathcal{B}) = \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{B}) - \mathbb{P}(\mathcal{A} \cap \mathcal{B})$$

Mutually exclusive events

Two events $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive if they cannot occur at the same time.

- We can write $\mathcal{A} \cap \mathcal{B} = \emptyset$ when the events are mutually exclusive.
- If $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive events, $\mathbb{P}(\mathcal{A} \cap \mathcal{B}) = 0$ and so

$$\mathbb{P}(\mathcal{A} \cup \mathcal{B}) = \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{B})$$

Complementary events

Complement

The complement of an event $\mathcal{A}$ is the event that $\mathcal{A}$ does not occur. Denoted $\mathcal{A}'$

Complementary events

Complement

The complement of an event $\mathcal{A}$ is the event that $\mathcal{A}$ does not occur. Denoted $\mathcal{A}'$

Note: All the sample points in S are either in $\mathcal{A}$ or $\mathcal{A}'$, no sample point can be in both. Thus,

$$\begin{aligned} & \mathbb{P}(\mathcal{A}) + \mathbb{P}(\mathcal{A}') = 1 \\ \Rightarrow \quad & \mathbb{P}(\mathcal{A}) = 1 - \mathbb{P}(\mathcal{A}') \end{aligned}$$

This is a useful formula for computation.

Example: Dice

Two fair dice are rolled. Event $\mathcal{A}$ is that we observe a 5. Event $\mathcal{B}$ is that the dice sum to 7. Calculate:

- $\mathbb{P}(\mathcal{A} \cap \mathcal{B})$ and $\mathbb{P}(\mathcal{A} \cup \mathcal{B})$.
- $\mathbb{P}(\mathcal{A}')$.

Example: Dice

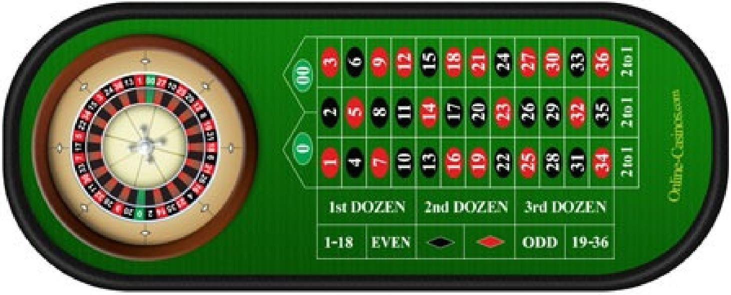
Two fair dice are rolled. Event $\mathcal{A}$ is that we observe a 5. Event $\mathcal{B}$ is that the dice sum to 7. Calculate:

- $\mathbb{P}(\mathcal{A} \cap \mathcal{B})$ and $\mathbb{P}(\mathcal{A} \cup \mathcal{B})$.
- $\mathbb{P}(\mathcal{A}')$.

S:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Example: Roulette


00
0
3 6 9 12 15 18 21 24 27 30 33 36 2 to 1
2 5 8 11 14 17 20 23 26 29 32 35 2 to 1
1 4 7 10 13 16 19 22 25 28 31 34 2 to 1
1st DOZEN
2nd DOZEN
3rd DOZEN
1-18
EVEN
ODD
19-36
Online-Casinos.com

- 38 slots, 18 red, 18 black, 2 green, 18 even, 18 odd.

Example: Roulette

- A – outcome is an odd number (0 and 00 are neither odd nor even).
- B – outcome is a red number.
- C – outcome is in the first dozen (1-12).

1. Define the events $A \cap B$ and $A \cup B$ as a specific sets of sample points.
2. Find $\mathbb{P}(A)$, $\mathbb{P}(B)$, $\mathbb{P}(A \cap B)$, $\mathbb{P}(A \cup B)$ and $\mathbb{P}(C)$ by summing the probabilities of the appropriate sample points.
3. Find $\mathbb{P}(A \cup B)$ using the additive rule. Are events A and B mutually exclusive?
4. Find $\mathbb{P}(A \cap B \cap C)$.

Conditional Probability

- Sometimes we are aware of extra information which might affect the outcome of an experiment. This extra information may then alter the probability of a particular event of interest.

Conditional Probability

- Sometimes we are aware of extra information which might affect the outcome of an experiment. This extra information may then alter the probability of a particular event of interest.
- Suppose we are interested in evaluating the probability that event $\mathcal{B}$ happens given that we know that event $\mathcal{A}$ has happened.

Conditional Probability

- Sometimes we are aware of extra information which might affect the outcome of an experiment. This extra information may then alter the probability of a particular event of interest.
- Suppose we are interested in evaluating the probability that event $\mathcal{B}$ happens given that we know that event $\mathcal{A}$ has happened.
- We write $\mathbb{P}(\mathcal{B}|\mathcal{A})$ for this.
- It is called the *conditional probability of $\mathcal{B}$ given $\mathcal{A}$.*

Bayes Theorem and Multiplication Rule

- Bayes Theorem states that

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

Bayes Theorem and Multiplication Rule

- Bayes Theorem states that

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

- The multiplication rule of probabilities states that

$$\begin{aligned} \mathbb{P}(\mathcal{B} \cap \mathcal{A}) &= \mathbb{P}(\mathcal{B}|\mathcal{A})\mathbb{P}(\mathcal{A}) \\ &= \mathbb{P}(\mathcal{A}|\mathcal{B})\mathbb{P}(\mathcal{B}) \end{aligned}$$

Very Useful Identity

- Let's consider the probability of event $\mathcal{A}$.

$$\mathbb{P}(\mathcal{A})$$

Very Useful Identity

- Let's consider the probability of event $\mathcal{A}$.

$$\mathbb{P}(\mathcal{A}) \;\; = \;\; \mathbb{P}\{(\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{B}')\}$$

Very Useful Identity

- Let's consider the probability of event $\mathcal{A}$.

$$\begin{aligned} \mathbb{P}(\mathcal{A}) &= \mathbb{P}\{(\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{B}')\} \\ &= \mathbb{P}(\mathcal{A} \cap \mathcal{B}) + \mathbb{P}(\mathcal{A} \cap \mathcal{B}') \end{aligned}$$

Very Useful Identity

- Let's consider the probability of event $\mathcal{A}$.

$$\begin{aligned}
\mathbb{P}(\mathcal{A}) &= \mathbb{P}\{(\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{B}')\} \\
&= \mathbb{P}(\mathcal{A} \cap \mathcal{B}) + \mathbb{P}(\mathcal{A} \cap \mathcal{B}') \\
&= \mathbb{P}(\mathcal{A}|\mathcal{B})\mathbb{P}(\mathcal{B}) + \mathbb{P}(\mathcal{A}|\mathcal{B}')\mathbb{P}(\mathcal{B}')
\end{aligned}$$

Very Useful Identity

- Let's consider the probability of event $\mathcal{A}$.

$$\begin{aligned}\mathbb{P}(\mathcal{A}) &= \mathbb{P}\{(\mathcal{A}\cap\mathcal{B})\cup(\mathcal{A}\cap\mathcal{B}')\}\\ &= \mathbb{P}(\mathcal{A}\cap\mathcal{B})+\mathbb{P}(\mathcal{A}\cap\mathcal{B}')\\ &= \mathbb{P}(\mathcal{A}|\mathcal{B})\mathbb{P}(\mathcal{B})+\mathbb{P}(\mathcal{A}|\mathcal{B}')\mathbb{P}(\mathcal{B}')\end{aligned}$$

- Hence, a very useful form of Bayes Theorem can be written as

$$\mathbb{P}(\mathcal{B}|\mathcal{A}) = \frac{\mathbb{P}(\mathcal{A}|\mathcal{B})\mathbb{P}(\mathcal{B})}{\mathbb{P}(\mathcal{A}|\mathcal{B})\mathbb{P}(\mathcal{B})+\mathbb{P}(\mathcal{A}|\mathcal{B}')\mathbb{P}(\mathcal{B}')}.$$

Example: Diagnostic Test

Suppose there is a rare disease which affects 1 person in every 1000 of the population. Fortunately a diagnostic medical test exists for the disease. It is a good test in that, if you have the disease, the test will be positive 95% of the time and if you do not have the disease it will be negative 99% of the time. If a patient tests positive for the disease, what is the probability that they actually have the disease?

Independence

Independence

Two events $\mathcal{A}$ and $\mathcal{B}$ are said to be independent if the occurence of $\mathcal{B}$ does not alter the probability that $\mathcal{A}$ has occurred. i.e. $\mathcal{A}$ and $\mathcal{B}$ are independent if:

$$\mathbb{P}(\mathcal{A}|\mathcal{B}) = \mathbb{P}(\mathcal{A})$$

Events which are not independent are said to be dependent.

Independence

Independence

Two events $\mathcal{A}$ and $\mathcal{B}$ are said to be independent if the occurence of $\mathcal{B}$ does not alter the probability that $\mathcal{A}$ has occurred. i.e. $\mathcal{A}$ and $\mathcal{B}$ are independent if:

$$\mathbb{P}(\mathcal{A}|\mathcal{B}) = \mathbb{P}(\mathcal{A})$$

Events which are not independent are said to be dependent.

- Combining the definition above with the multiplicative rule, it can be seen that if $\mathcal{A}$ and $\mathcal{B}$ are independent then:

  $$\mathbb{P}(\mathcal{A} \cap \mathcal{B}) = \mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B})$$

  The converse is also true, i.e. if $\mathbb{P}(\mathcal{A} \cap \mathcal{B}) = \mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B})$ then the events $\mathcal{A}$ and $\mathcal{B}$ are independent.

Mutual Exclusivity

- Suppose $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive events. If $\mathcal{B}$ occurs then $\mathcal{A}$ cannot occur simultaneously so $\mathbb{P}(\mathcal{A}|\mathcal{B}) = 0$.

Mutual Exclusivity

- Suppose $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive events. If $\mathcal{B}$ occurs then $\mathcal{A}$ cannot occur simultaneously so $\mathbb{P}(\mathcal{A}|\mathcal{B}) = 0$.

$\Rightarrow$ Mutually exclusive events are dependent events.

Example: Tyres

Three types of tyres (Types A, B and C) are independently tested for suitability for use in SUVs. The probabilities that each passes the test are 0.7, 0.6, and 0.5 respectively.

- What is the probability that they all fail the test?
- What is the probability that at least one passes?
- Granted that at least one passed, what is the probability that type B was the only one to do so?

Example: Corrosion

The independence of corrosion and the functional status of a machine component are to be investigated. Are they independent?

	Functioning	Malfunctioning
Corroded	0.2	0.4
Not corroded	0.3	0.1

Counting Rules

Multiplicative counting rule

Suppose we have k sets with n_1 elements in the first set, n_2 elements in the second, ..., n_k elements in the k^{th} set. If we wish to take a sample of size k consisting of 1 element from each set, the number of ways this sample can be formed is:

$$n_1 \times n_2 \times \ldots \times n_k$$

e.g. A password consists of 1 letter followed by 3 digits. How many possible passwords are there?

Counting Rules

Multiplicative counting rule

Suppose we have k sets with n_1 elements in the first set, n_2 elements in the second, ..., n_k elements in the k^{th} set. If we wish to take a sample of size k consisting of 1 element from each set, the number of ways this sample can be formed is:

$$n_1 \times n_2 \times \ldots \times n_k$$

e.g. A password consists of 1 letter followed by 3 digits. How many possible passwords are there?

$$\begin{aligned} \# \text{ passwords} &= \boxed{26} \times \boxed{10} \times \boxed{10} \times \boxed{10} \\ &= 26,000 \end{aligned}$$

Combinations Rule

Given a set of N elements, an unordered subset of these elements is called a combination.

Combinations rule

The number of combinations of size r which can be formed from a set of size N is:

$$\binom{N}{r} = \frac{N!}{r!(N-r)!}$$

Combinations Rule

Given a set of N elements, an unordered subset of these elements is called a combination.

Combinations rule

The number of combinations of size r which can be formed from a set of size N is:

$$\binom{N}{r} = \frac{N!}{r!(N-r)!}$$

e.g. The number of soccer teams which can be formed from a panel of size 22 is:

$$\binom{22}{11} = \frac{22!}{11!(22-11)!} = 705,432 \text{ teams.}$$

Example: Illegal Trading

A trading manager knows that 3 out of 10 traders under her supervision are making illegal trades. If she selects 2 workers at random what is the probability that they have both been trading illegally?

Example: Poker

1. How many 5 card hands may be dealt from a deck of 52 cards?
2. What is the probability of being dealt 3 of a kind in poker?
3. What is the probability of being dealt a full house in poker? (2 of one denomination and 3 of another)

Permutations Rule

The arrangement of elements of a set in a distinct order is called a permutation.

Permutations rule

The number of different permutations of size r which can be formed from a set of size N is:

$$P_r^N = \frac{N!}{(N-r)!}$$

Permutations Rule

The arrangement of elements of a set in a distinct order is called a permutation.

Permutations rule

The number of different permutations of size r which can be formed from a set of size N is:

$$P_r^N = \frac{N!}{(N-r)!}$$

e.g. 50 engineers are available to do 3 jobs. How many ways can the engineers be allocated to the jobs?

$$P_3^{50} = \frac{50!}{(50-3)!} = 117,600$$

Example: ID number

A student ID number consists of 8 digits.

1. How many unique ID numbers are there?
2. If each digit may only appear once per ID number, how many unique ID numbers can be created?
3. Assume the first 2 digits of an ID number correspond to the year a student joined the university. How many unique ID codes are there for students who started in 2019, if each digit may only appear once in the entire ID code?

STAT10430 - Statistics with Python
Discrete Random Variables

Dr. Áine Byrne
aine.byrne@ucd.ie

Overview

- Discrete Random Variables
- Expected Value
- Variance
- Probability mass functions
- Binomial
- Hypergeometric
- Poisson

Random Variables

A random variable is a variable which assumes numerical values associated with the random outcomes of an experiment, where one (an only one) numerical value is assigned to each sample point.

Random Variables

A random variable is a variable which assumes numerical values associated with the random outcomes of an experiment, where one (an only one) numerical value is assigned to each sample point.

For example,

1. Number of defective items in a batch.
2. Number of cars crossing a bridge in a day.
3. Highest daily temperature in Dublin.
4. Trading price of a gold bullion each day.

Types of random variables

Two types

There are two types or random variables:

- A random variable is said to be discrete if it can assume only a countable number of values.

Types of random variables

Two types

There are two types or random variables:

- A random variable is said to be discrete if it can assume only a countable number of values.
- A random variable that can assume values corresponding to any of the points contained in one or more intervals is called continuous.

Types of random variables

Two types

There are two types or random variables:

- A random variable is said to be discrete if it can assume only a countable number of values.
- A random variable that can assume values corresponding to any of the points contained in one or more intervals is called continuous.

- Examples 1 and 2 on the previous slide are discrete random variables.
- Examples 3 and 4 on the previous slide are continuous random variables.

Discrete probability distribution

The probability distribution for a discrete random variable X can be represented by a formula, a table, or a graph, which provides the probabilities $p(x)$ corresponding to each and every value of x:

$$\mathbb{P}(X = x) = p(x)$$

Discrete probability distribution

The probability distribution for a discrete random variable X can be represented by a formula, a table, or a graph, which provides the probabilities $p(x)$ corresponding to each and every value of x:

$$\mathbb{P}(X = x) = p(x)$$

The probability distribution function for a discrete random variable is also known as probability mass function.

Discrete probability distribution

The probability distribution for a discrete random variable X can be represented by a formula, a table, or a graph, which provides the probabilities $p(x)$ corresponding to each and every value of x:

$$\mathbb{P}(X = x) = p(x)$$

The probability distribution function for a discrete random variable is also known as probability mass function.

For any discrete probability distribution the following must be true:

1. $0 \leq p(x) \leq 1$ for all x.
2. $\sum_x p(x) = 1$ where the summation is over all possible values of x.

Example: Discrete probability distribution

- X – Number observed after rolling a fair die:

	$X = 1$	$X = 2$	$X = 3$	$X = 4$	$X = 5$	$X = 6$
$p(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Example: Discrete probability distribution

- X – Number observed after rolling a fair die:

	$X = 1$	$X = 2$	$X = 3$	$X = 4$	$X = 5$	$X = 6$
$p(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

- Y – Number of heads observed in two coin tosses:

	$Y = 0$	$Y = 1$	$Y = 2$
$p(y)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Cumulative distribution (discrete case)

Cumulative distribution function

The cumulative distribution function of random variable X is given by the cumulative probabilities $F(x)$:

$$F(x) = \mathbb{P}(X \leq x) = \sum_{u=0}^{x} p(u)$$

Example: Cumulative distribution (discrete)

- X – Number observed after rolling a fair die:

	$X \leq 1$	$X \leq 2$	$X \leq 3$	$X \leq 4$	$X \leq 5$	$X \leq 6$
$F(x)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	1

Example: Cumulative distribution (discrete)

- X – Number observed after rolling a fair die:

	$X \leq 1$	$X \leq 2$	$X \leq 3$	$X \leq 4$	$X \leq 5$	$X \leq 6$
$F(x)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	1

- Y – Number of heads observed in two coin tosses:

	$Y \leq 0$	$Y \leq 1$	$Y \leq 2$
$F(y)$	$\frac{1}{4}$	$\frac{3}{4}$	1

Expected value of discrete random variables

Expected Value

The mean or expected value of a discrete random variable is:

$$\mu = \mathbb{E}[X] = \sum_{x} x \; p(x)$$

Expected value of discrete random variables

Expected Value

The mean or expected value of a discrete random variable is:

$$\mu = \mathbb{E}[X] = \sum_{x} x \, p(x)$$

For example, consider rolling a die again and let X be the number observed:

$$\mathbb{E}[X] \;\; = \;\; \sum_{x} x \, p(x) = 1 \times \left(\frac{1}{6}\right) + 2 \times \left(\frac{1}{6}\right) + \ldots + 6 \times \left(\frac{1}{6}\right) = 3.5$$

Expected value of discrete random variables

Expected Value

The mean or expected value of a discrete random variable is:

$$\mu = \mathbb{E}[X] = \sum_x x\ p(x)$$

For example, consider rolling a die again and let X be the number observed:

$$\begin{aligned}\mathbb{E}[X] &= \sum_x x\ p(x) = 1 \times \left(\frac{1}{6}\right) + 2 \times \left(\frac{1}{6}\right) + \ldots + 6 \times \left(\frac{1}{6}\right) = 3.5 \\ \mathbb{E}[Y] &= \sum_y y\ p(y) = 0 \times \left(\frac{1}{4}\right) + 1 \times \left(\frac{1}{2}\right) + 2 \times \left(\frac{1}{4}\right) = 1\end{aligned}$$

Expected value of discrete random variables

- Note: The expected value is the mean of the probability distribution or a measure of central tendency.

Expected value of discrete random variables

- Note: The expected value is the mean of the probability distribution or a measure of central tendency.

- The expected value of a function of a random variable $g(X)$ can be found similarly:

$$\mathbb{E}[g(X)] = \sum_{x} g(x)p(x)$$

Expected value of discrete random variables

- Note: The expected value is the mean of the probability distribution or a measure of central tendency.

- The expected value of a function of a random variable $g(X)$ can be found similarly:

$$\mathbb{E}[g(X)] = \sum_x g(x)p(x)$$

  e.g. From the previous example, let $g(X) = X^2$:

$$\mathbb{E}[X^2] = \sum_x x^2 p(x) = 1^2 \times \left(\frac{1}{6}\right) + 2^2 \times \left(\frac{1}{6}\right) + \ldots + 6^2 \times \left(\frac{1}{6}\right) = 15.17$$

Properties of Expectation:

Expectation:

1. $\mathbb{E}[c] = c$ where c is a constant.

Properties of Expectation:

Expectation:

1. $\mathbb{E}[c] = c$ where c is a constant.

2. $\mathbb{E}[c\, g(X)] = c\, \mathbb{E}[g(X)]$,

   in particular $\mathbb{E}[c\, X] = c\, \mathbb{E}[X]$

Properties of Expectation:

Expectation:

1. $\mathbb{E}[c] = c$ where c is a constant.

2. $\mathbb{E}[c\, g(X)] = c\, \mathbb{E}[g(X)]$,

   in particular $\mathbb{E}[c\, X] = c\, \mathbb{E}[X]$

3. If $g_1(X)$, $g_2(X), \ldots, g_k(X)$ be k functions of X. Then

$$\mathbb{E}\left[g_1(X) + \ldots + g_k(X)\right] = \mathbb{E}\left[g_1(X)\right] + \ldots \mathbb{E}\left[g_k(X)\right]$$

Variance of a random variable

The variance measures the spread of a probability distribution.

Variance

The variance of a discrete random variable X is:

$$\sigma^2 = \mathbb{E}\left[(x-\mu)^2\right] = \sum_x (x-\mu)^2 p(x) = \sum_x x^2 p(x) - \mu^2$$

- A sometimes more convenient formula for variance is:

$$\sigma^2 = \mathbb{E}[X^2] - \mu^2$$

Variance of a random variable

The variance measures the spread of a probability distribution.

Variance

The variance of a discrete random variable X is:

$$\sigma^2 = \mathbb{E}\left[(x-\mu)^2\right] = \sum_x (x-\mu)^2 p(x) = \sum_x x^2 p(x) - \mu^2$$

- A sometimes more convenient formula for variance is:

$$\sigma^2 = \mathbb{E}[X^2] - \mu^2$$

The standard deviation of a discrete random variable is the square root of the variance:

$$\sigma = \sqrt{\sigma^2}$$

Properties of Variance:

Variance:

1. $Var(c) = 0$ where c is a constant.

Properties of Variance:

Variance:

1. $Var(c) = 0$ where c is a constant.

2. $Var(a_1X_1 + \ldots + a_mX_m) = a_1^2 Var(X_1) + \ldots a_m^2 Var(X_m)$
   if $X_1, \ldots, X_m$ are **independent** random variables.

Example: Rolling one Die

- Let X be the number displayed after rolling one die.
- From previous calculations we know that:
  - $\mathbb{E}[X] = 3.5$
  - $\mathbb{E}[X^2] = 15.17$
- What is the standard deviation of X?

Example: Car Ratings

Cars are rated on a scale of 1-5 stars with regard to the level of protection they offer in a head-on collision. 98 cars were tested and the star allocations are summarised in the following table:

$*$	$**$	$***$	$****$	$*****$
0	4	17	59	18

- Calculate the probability distribution for Y, the number of stars obtained by one of the cars selected at random.
- Calculate $\mathbb{P}(Y \leq 3)$.
- Hence calculate $\mathbb{P}(Y > 3)$.
- What is the expected value of Y?
- Calculate the standard deviation of Y.

Trading Example:

The number of trades on which a trader makes a loss per day, Y, is assumed to have a mean and variance of 4. If the profit made by the trader per day is given by $X = 1000 - 10Y - 5Y^2$.

1. Find the expected profit per day.
2. Let O be a random variable denoting overhead costs, which has standard deviation 50. If the profit made by a trading team was given by $P = 7X - 1.5O - 10$, what is the standard deviation of P? Assume that O and X are independent and $\mathbb{E}[X^2] = 800,000$.

Binomial Distribution

- Suppose an experiment is conducted where there are a fixed number (n) of independent trials.

Binomial Distribution

- Suppose an experiment is conducted where there are a fixed number (n) of independent trials.
- For each trial we observe either a success or a failure.

Binomial Distribution

- Suppose an experiment is conducted where there are a fixed number (n) of independent trials.
- For each trial we observe either a success or a failure.
- Further, suppose that the probability $\mathbb{P}(\text{Success}) = p$ for all trials. Also, $\mathbb{P}(\text{Failure}) = q = 1 - p$.

Binomial Distribution

- Suppose an experiment is conducted where there are a fixed number (n) of independent trials.
- For each trial we observe either a success or a failure.
- Further, suppose that the probability $\mathbb{P}(\text{Success}) = p$ for all trials. Also, $\mathbb{P}(\text{Failure}) = q = 1 - p$.
- Then, the number of observed successes, X, has a probability mass function (pmf) equal to

$$p(x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ for } x = 0, 1, \ldots, n.$$

Binomial Distribution

- Suppose an experiment is conducted where there are a fixed number (n) of independent trials.
- For each trial we observe either a success or a failure.
- Further, suppose that the probability $\mathbb{P}(\text{Success}) = p$ for all trials. Also, $\mathbb{P}(\text{Failure}) = q = 1 - p$.
- Then, the number of observed successes, X, has a probability mass function (pmf) equal to

$$p(x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ for } x = 0, 1, \ldots, n.$$

- We write $X \sim \text{binomial}(n, p)$.

Examples: Binomial

Binomial

- Flip a coin 10 times, Y is the number of heads observed.
- 100 consumers are selected at random and given a blind taste test. Each states which of two brands (A and B) of chocolate they prefer. Y is the number who prefer brand A.

Examples: Binomial

Binomial

- Flip a coin 10 times, Y is the number of heads observed.
- 100 consumers are selected at random and given a blind taste test. Each states which of two brands (A and B) of chocolate they prefer. Y is the number who prefer brand A.

Not Binomial

- In the run up to an election every person in 5 city neighbourhoods are asked which of two candidates they will vote for.
  - This is not binomial since the people are not chosen at random and so are unlikely to be independent. It is likely that people who live in the same neighbourhood will have similar income levels, education etc. and so will vote in a similar way.

Binomial Distribution: Explained

The p.m.f. is:

$$p(x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ for } x = 0, 1, \ldots, n.$$

- p^x is the probability of x successes.

Binomial Distribution: Explained

The p.m.f. is:

$$p(x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ for } x = 0, 1, \ldots, n.$$

- p^x is the probability of x successes.
- $(1-p)^{n-x}$ is the probability of $n - x$ failures.

Binomial Distribution: Explained

The p.m.f. is:

$$p(x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ for } x = 0, 1, \ldots, n.$$

- p^x is the probability of x successes.
- $(1-p)^{n-x}$ is the probability of $n - x$ failures.
- $\binom{n}{x}$ is the number of ways of observing x successes in n trials.

Binomial Distribution: Properties

If $X \sim \text{binomial}(n, p)$ then:

$$\mathbb{E}(X) = np$$

Binomial Distribution: Properties

If $X \sim \text{binomial}(n, p)$ then:

$$\mathbb{E}(X) = np$$
$$\mathbb{V}\text{ar}(X) = np(1 - p)$$

Binomial Distribution: Properties

If $X \sim \text{binomial}(n, p)$ then:

$$\mathbb{E}(X) = np$$

$$\mathbb{V}\text{ar}(X) = np(1-p)$$

$$F(x) = \mathbb{P}(X \leq x) = \sum_{y=0}^{x} \binom{n}{y} p^y (1-p)^{n-y}, \text{ for } x = 0, 1, \ldots, n.$$

Binomial Distribution: Example

A financial broker sells products to customers.

- Suppose he sells twenty products.
- Each product has probability $p = 0.9$ of making money for the customer.

Binomial Distribution: Example

A financial broker sells products to customers.

- Suppose he sells twenty products.
- Each product has probability $p = 0.9$ of making money for the customer.

What is the probability that at least 18 of the products make money?

Binomial Distribution: Example (Revisited)

- What is the probability that at most 2 products lose money?

- **Result:** If $X \sim \text{binomial}(n, p)$ and if $Y = n - X$. Then,
$$Y \sim \text{binomial}(n, 1 - p).$$

Hypergeometric Random Variables

Characteristics of a Hypergeometric Random Variable

1. The experiment consists of randomly drawing n elements without replacement from a set of N elements, s of which are 'special' and $(N - s)$ are regular.
2. The hypergeometric random variable X is the number of special items in the draw of n elements.

Hypergeometric Random Variables

Characteristics of a Hypergeometric Random Variable

1. The experiment consists of randomly drawing n elements without replacement from a set of N elements, s of which are 'special' and $(N - s)$ are regular.
2. The hypergeometric random variable X is the number of special items in the draw of n elements.

Examples:

- The lottery – 49 numbers in total (N), 6 special numbers (s), you pick 6 (n) and X is the number of special numbers you match on your ticket.

Hypergeometric Random Variables

Characteristics of a Hypergeometric Random Variable

1. The experiment consists of randomly drawing n elements without replacement from a set of N elements, s of which are 'special' and $(N - s)$ are regular.
2. The hypergeometric random variable X is the number of special items in the draw of n elements.

Examples:

- The lottery – 49 numbers in total (N), 6 special numbers (s), you pick 6 (n) and X is the number of special numbers you match on your ticket.
- The government inspect bridges each year. Out of a total of 20 bridges 5 are not up to code. If a sample of 8 of the bridges are inspected then the number of defective bridges in that sample is a hypergeometric random variable.

The Hypergeometric Distribution

Hypergeometric Distribution

$$\mathbb{P}(X = x) = \frac{\binom{s}{x}\binom{N-s}{n-x}}{\binom{N}{n}}$$

$$\mu = \frac{ns}{N} \qquad \sigma^2 = n\left(\frac{s}{N}\right)\left(\frac{N-s}{N}\right)\left(\frac{N-n}{N-1}\right)$$

where,

- $N =$ Total number of elements.
- $s =$ Number of special items in N elements.
- $n =$ Number of elements drawn.
- $x =$ Number of special items in the n elements.

Example: Catalysts

An experiment is conducted to select a suitable catalyst for the commercial production of ethylenediamine (EDA), a product used in soaps. Suppose a chemical engineer randomly selects 3 catalysts for testing from a group of 10 catalysts, 6 of which have low acidity and 4 of which have high acidity.

- Find the probability that no highly acidic catalyst is selected.
- Find the probability that exactly one highly-acidic catalyst is selected.
- How many highly acidic acids would you expect to be chosen?
- What is the standard deviation of the number of highly acidic catalysts chosen?

Box Example:

A box contains 6 balls of which 4 are black and 2 are white. 3 balls are selected without replacement.

1. What is the probability that we will obtain 3 blacks?

2. What is the probability that we will exactly 1 black ball?

Poisson Distribution

- Suppose an experiment is conducted where the number of events are counted for:

Poisson Distribution

- Suppose an experiment is conducted where the number of events are counted for:
  - a fixed time period
  - a given length
  - a given area
  - a given volume
  - . . .

Poisson Distribution

- Suppose an experiment is conducted where the number of events are counted for:
  - a fixed time period
  - a given length
  - a given area
  - a given volume
  - ...

- The probability that an event occurs in a given unit of time/length/area/volume is the same for all units of the same size.

Poisson Distribution

- Suppose an experiment is conducted where the number of events are counted for:
  - a fixed time period
  - a given length
  - a given area
  - a given volume
  - ...

- The probability that an event occurs in a given unit of time/length/area/volume is the same for all units of the same size.

- The number of events that occur in disjoint units of time/length/area/volume are independent.

Poisson Random Variables

Examples

- The number of vehicles crossing a bridge in a day.
- The number of jobs sent to a computer processor in an hour.
- The number of times a neuron spikes in a minute.
- The number of particles of a pollutant in a litre of river water.

Poisson Probability Distribution

The probability distribution for a Poisson random variable Y is given by:

$$\mathbb{P}(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

where

- $\lambda =$ rate events occur during a given unit of time, area or volume.
- The mean and variance of a Poisson random variable are:
  - $\mathbb{E}[Y] = \mu = \lambda$
  - $\mathrm{Var}[Y] = \sigma^2 = \lambda$

Poisson Distribution: Properties

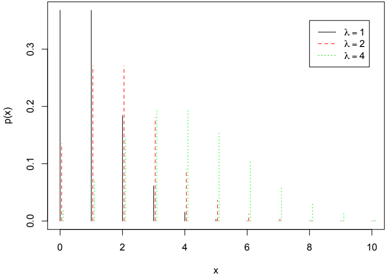
- The parameter λ is called the rate parameter.
- It turns out that

$$\mathbb{E}(X) = \lambda = \mu \text{ and } \mathbb{V}\text{ar}(X) = \lambda = \sigma^2.$$

Poisson Distribution: Properties

- The parameter λ is called the rate parameter.
- It turns out that

$$\mathbb{E}(X) = \lambda = \mu \text{ and } \mathbb{V}\text{ar}(X) = \lambda = \sigma^2.$$



Poisson Distribution: Property

- Suppose we sum two Poisson random variables, then the sum is also Poisson.

- That is, if

$$X \sim \text{Poisson}(\lambda) \text{ and } Y \sim \text{Poisson}(\mu),$$

then

$$X + Y \sim \text{Poisson}(\lambda + \mu).$$

Example: Cables

Suppose that faults occurring in cables being manufactured follow a Poisson process with a rate of 2 faults per 100 m.

1. What is the probability of no faults in 100 m of cable?
2. What is the probability of one fault in 100 m of cable?
3. What is the probability of two or more faults in 100 m of cable?
4. What is the probability of no faults in 200 m of cable?

Example: Births

Births in a hospital occur randomly at an average rate of 3.8 births per hour.

1. What is the probability of observing 6 births in a given hour at the hospital?
2. What is the probability of observing no births in a given hour at the hospital?
3. What is the probability of observing 6 births in a 2 hour period?

STAT10430 - Statistics with Python
Continuous Random Variables

Dr. Áine Byrne
aine.byrne@ucd.ie

Overview

- Continuous Random Variables
- Expected Value
- Variance
- Continuous Probability Distributions
  - Probability density functions.
  - Uniform.
  - Exponential.
  - Weibull.
  - Normal.

Random Variables (recap)

A random variable is a variable which assumes numerical values associated with the random outcomes of an experiment, where one (an only one) numerical value is assigned to each sample point.

Random Variables (recap)

A random variable is a variable which assumes numerical values associated with the random outcomes of an experiment, where one (an only one) numerical value is assigned to each sample point.

For example,

1. Number of defective items in a batch.
2. Number of cars crossing a bridge in a day.
3. Highest daily temperature in Dublin.
4. Trading price of a gold bullion each day.

Types of random variables (recap)

Two types

There are two types or random variables:

- A random variable is said to be discrete if it can assume only a countable number of values.

Types of random variables (recap)

Two types

There are two types or random variables:

- A random variable is said to be discrete if it can assume only a countable number of values.
- A random variable that can assume values corresponding to any of the points contained in one or more intervals is called continuous.

Types of random variables (recap)

Two types

There are two types or random variables:

- A random variable is said to be discrete if it can assume only a countable number of values.
- A random variable that can assume values corresponding to any of the points contained in one or more intervals is called continuous.

- Examples 1 and 2 on the previous slide are discrete random variables.
- Examples 3 and 4 on the previous slide are continuous random variables.

Continuous Random Variables

- Recall: A continuous random variable is one which can assume values corresponding to any of the points contained in one or more intervals.

Continuous Random Variables

- Recall: A continuous random variable is one which can assume values corresponding to any of the points contained in one or more intervals.

- The graphical form of the probability distribution for a continuous random variable X is a smooth curve called a probability density function (pdf) and is denoted by $f(x)$.

Continuous Random Variables

- Recall: A continuous random variable is one which can assume values corresponding to any of the points contained in one or more intervals.

- The graphical form of the probability distribution for a continuous random variable X is a smooth curve called a probability density function (pdf) and is denoted by $f(x)$.

For $f(x)$ to be a valid density we must have:

$$f(x) \geq 0 \; \forall \; x \text{ and } \int_{-\infty}^{+\infty} f(x) \, dx = 1$$

Continuous Random Variables

- The probability that X takes a value between two points a and b is given by the area under the density curve between these points.

Continuous Random Variables

- The probability that X takes a value between two points a and b is given by the area under the density curve between these points.

- Hence the total area under the curve must be 1 in order to satisfy the second probability axiom. $(\mathbb{P}(S) = 1)$

Continuous Random Variables

- The probability that X takes a value between two points a and b is given by the area under the density curve between these points.

- Hence the total area under the curve must be 1 in order to satisfy the second probability axiom. $(\mathbb{P}(S) = 1)$

- Since there is no area over a single point:

  $$\mathbb{P}(X = x) = 0 \text{ for all } x$$

  Thus, $\mathbb{P}(a < x < b) = \mathbb{P}(a \leq x \leq b)$.

Cumulative Distribution Function

- The cumulative distribution function, $F(x)$ gives the area under the density curve to the left of a point x.

Cumulative Distribution Function

- The cumulative distribution function, $F(x)$ gives the area under the density curve to the left of a point x.

- It is found by integrating the density function between the lower limit of the range of the random variable and x.

Cumulative Distribution Function

- The cumulative distribution function, $F(x)$ gives the area under the density curve to the left of a point x.

- It is found by integrating the density function between the lower limit of the range of the random variable and x.

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(u)du$$

- Note: The lower limit of the range of a random variable is not always $-\infty$.

Mean and Variance

- Expected value of X:

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

Mean and Variance

- Expected value of X:

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

- Expected value of $g(x)$:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)\ f(x) dx$$

Mean and Variance

- Expected value of X:

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

- Expected value of $g(x)$:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)\ f(x) dx$$

- As in the discrete case:

$$\text{Var}(X) = \sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

Mean and Variance

- Expected value of X:

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

- Expected value of $g(x)$:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)\ f(x) dx$$

- As in the discrete case:

$$\text{Var}(X) = \sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

- Note that the integration above should take place over the range of X. This may not always be $(-\infty, +\infty)$

The Uniform Distribution

- Continuous random variables that have equally likely outcomes over their range of possible values possess a uniform probability distribution.

The Uniform Distribution

- Continuous random variables that have equally likely outcomes over their range of possible values possess a uniform probability distribution.

- If X is a uniform random variable, we write $X \sim U(a, b)$.

The Uniform Distribution

- Continuous random variables that have equally likely outcomes over their range of possible values possess a uniform probability distribution.
- If X is a uniform random variable, we write $X \sim U(a, b)$.

Probability density function

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if} \quad a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

The Uniform Distribution

Cumulative distribution function

$$F(x) = \begin{cases} 0 & \text{if} \quad x < a \\ \dfrac{x-a}{b-a} & \text{if} \quad a \leq x \leq b \\ 1 & \text{if} \quad x > b \end{cases}$$

Properties:

- Mean: $\mu = \frac{a+b}{2}$
- Variance: $\sigma^2 = \frac{(b-a)^2}{12}$
- $\mathbb{P}(c < x < d) = (d-c)/(b-a)$, where $a \leq c < d \leq b$

Example: Uniform

A company's rolling machine is producing sheets of steel of varying thickness. The thickness Y is a uniform random variable taking values between 1.5 and 2 milimeters. Any sheets less than 1.6 milimeters thick must be scrapped.

1. Calculate the mean and standard deviation of Y. Graph the probability density function and mark the mean as well as 1 and 2 standard deviation intervals around the mean on the horizontal axis.
2. If this distribution models the situation correctly, what proportion of sheets would you expect to be scrapped.

Exponential Distribution

The exponential distribution $X \sim Exp(\lambda)$ is an example of a waiting time distribution.

Examples:

- Time between server failures
- Time spent waiting for a train

Probability density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if} \quad x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where the parameter $\lambda > 0$.

Expected value: $\mu = \frac{1}{\lambda}$

Variance: $\sigma^2 = \frac{1}{\lambda^2}$

Probabilities and the Exponential Distribution

- Note that the standard deviation of an exponentially distributed random variable is equal to its mean.

Probabilities and the Exponential Distribution

- Note that the standard deviation of an exponentially distributed random variable is equal to its mean.
- As with all continuous random variables, probabilities are found by integrating the density function over the interval of interest.

Probabilities and the Exponential Distribution

- Note that the standard deviation of an exponentially distributed random variable is equal to its mean.
- As with all continuous random variables, probabilities are found by integrating the density function over the interval of interest.
- It can be convenient to use the cumulative distribution function instead.

Probabilities and the Exponential Distribution

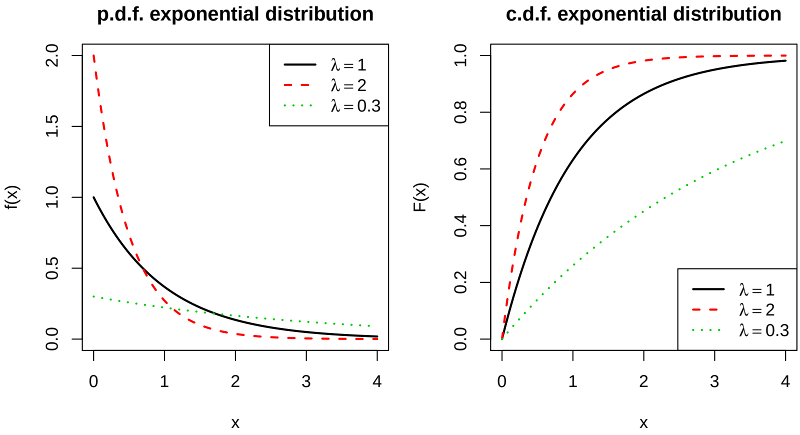
- Note that the standard deviation of an exponentially distributed random variable is equal to its mean.
- As with all continuous random variables, probabilities are found by integrating the density function over the interval of interest.
- It can be convenient to use the cumulative distribution function instead.

Cumulative distribution function

$$F(x) = \mathbb{P}(X \leq x) = 1 - e^{-\lambda x}$$

$$\mathbb{P}(X > x) = e^{-\lambda x}$$

Exponential Distribution



Example: Exponential – Toll Bridge

Suppose that the length of time (in minutes) between cars approaching a particular barrier at a toll bridge is exponentially distributed with mean 2. What is the probability that at least 3 minutes pass without a car approaching the barrier?

The Memoryless Property

- The exponential distribution is said to be memoryless since:

$$\mathbb{P}(X > a + b | X > a) = \mathbb{P}(X > b)$$

The Memoryless Property

- The exponential distribution is said to be memoryless since:

$$\mathbb{P}(X > a + b | X > a) = \mathbb{P}(X > b)$$

- This has important implications for the exponential's role as a waiting time distribution.

The Memoryless Property

- The exponential distribution is said to be memoryless since:

$$\mathbb{P}(X > a + b | X > a) = \mathbb{P}(X > b)$$

- This has important implications for the exponential's role as a waiting time distribution.
- Only the future waiting time is considered in an exponential distribution. It does not take into account the amount of time which has already elapsed.
  - For server failures this is realistic.
  - For human lifetimes it is not.

Poisson and Exponential Distributions

- The Poisson distributions is a discrete counting process for the number of events that occur in an interval.

- If events occur according to a Poisson process with rate λ then the probability of k events in an interval of duration t is:

$$\mathbb{P}(X = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

Poisson and Exponential Distributions

- The Poisson distributions is a discrete counting process for the number of events that occur in an interval.

- If events occur according to a Poisson process with rate λ then the probability of k events in an interval of duration t is:

$$\mathbb{P}(X = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

- Consider the random variable t_0, the time until the first event in a Poisson process.

- If we are interested in the probability that t_0 exceeds some given time t^* then we require 0 events to occur in the interval $[0, t^*]$

Poisson and Exponential Distributions

- From above this probability is:

$$\mathbb{P}(X = 0) = \frac{(\lambda t^*)^0 e^{-\lambda t^*}}{0!} = e^{-\lambda t^*}$$

Poisson and Exponential Distributions

- From above this probability is:

$$\mathbb{P}(X = 0) = \frac{(\lambda t^*)^0 e^{-\lambda t^*}}{0!} = e^{-\lambda t^*}$$

- But for an exponential random variable Y with rate λ:

$$\mathbb{P}(Y > t^*) = e^{-\lambda t^*}$$

Poisson and Exponential Distributions

- From above this probability is:

$$\mathbb{P}(X = 0) = \frac{(\lambda t^*)^0 e^{-\lambda t^*}}{0!} = e^{-\lambda t^*}$$

- But for an exponential random variable Y with rate λ:

$$\mathbb{P}(Y > t^*) = e^{-\lambda t^*}$$

- Thus the time until the first event in a Poisson process with rate parameter λ per unit time follows an exponential distribution with rate parameter λ.
- The time between consecutive events also follows an exponential distribution with rate parameter λ.

Example: Server log-ons

User log-ons to a server can be modelled as a Poisson process with a mean of 25 log-ons per hour. What is the probability that there are no log-ons in a 6 minute interval? What is the probability that the time until the next log-on is between 2 and 3 minutes?

Example: Ice creams

The number of ice creams sold at a seaside stall follows a Poisson distribution with a rate of 20 per day. What is the probability it takes less than half an hour to sell the first ice-cream of the day? (Assume the working day is 8 hours.)

The Weibull Distribution

The Weibull distribution is another waiting time distribution.

If $X \sim \text{Weibull}(\alpha, \beta)$ then the density function is:

$$f(x) = \begin{cases} \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}} & \text{if} \quad x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where the parameters $\alpha > 0$, $\beta > 0$.

This distribution allows for the number of failures to vary with time.

- The longer you use your computer the more likely it is that a component will fails.

The Weibull Distribution

The shape parameter, α, describes how the failure rate changes over time.

$\alpha < 1$: High failure rate to begin with. Decreasing over time. e.g. High infant mortality.

The Weibull Distribution

The shape parameter, α, describes how the failure rate changes over time.

$\alpha < 1$: High failure rate to begin with. Decreasing over time.
e.g. High infant mortality.

$\alpha = 1$: Failure rate is constant over time.
e.g. Random external events are causing mortality or failure.

The Weibull Distribution

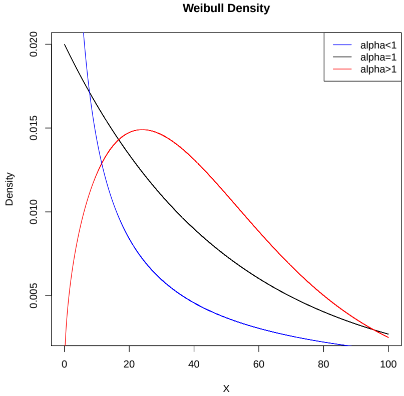
The shape parameter, α, describes how the failure rate changes over time.

$\alpha < 1$: High failure rate to begin with. Decreasing over time.
e.g. High infant mortality.

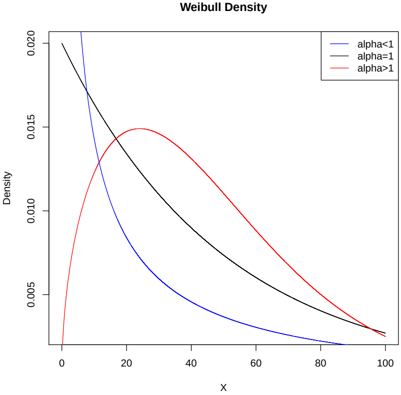
$\alpha = 1$: Failure rate is constant over time.
e.g. Random external events are causing mortality or failure.

$\alpha > 1$: Failure rate increases with time.
e.g. Systems which are affected by aging, parts which are more likely to fail as time goes on.

The Weibull Distribution



The Weibull Distribution



- A Weibull random variable with $\alpha = 1$ is an exponential random variable with rate parameter $\frac{1}{\beta}$. Hence the exponential distribution is a special case of the Weibull distribution.

Example:

A company requires a new ventilation system to be installed to meet government regulations. If it is installed within 100 days the cost will be €10,000. If it is installed after the 100 day mark the cost is reduced to €8,000 but there is a 60% chance the government will find out and fine them €1,000. Calculate the expected cost of installing the system if the installation time is t where:

$$t \sim \text{Weibull}(\alpha = 1/4, \beta = 4)$$

The Normal Distribution

The normal distribution is one of the most useful and widely used distributions in statistics.

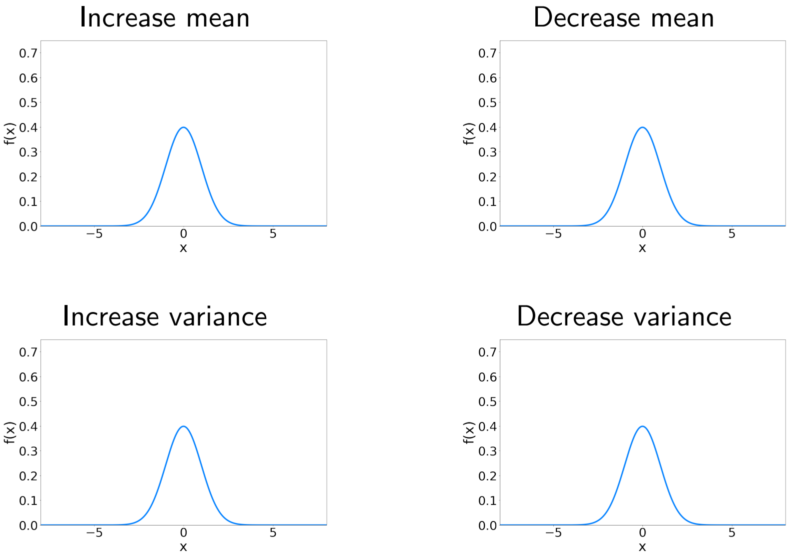
The Normal Distribution

The density function of a normal random variable Y is given by:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2\sigma^2)}$$

The parameters μ and σ^2 are the mean and variance of the normal random variable Y.

Exponential The Normal Distribution



Increase mean

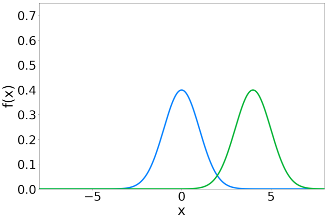
Decrease mean

Increase variance

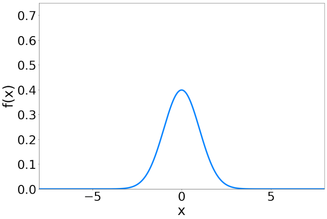
Decrease variance

Exponential The Normal Distribution

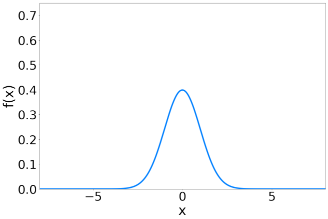
Increase mean



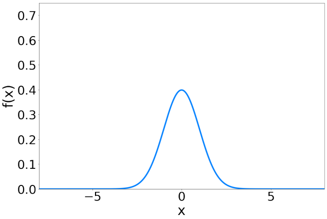
Decrease mean



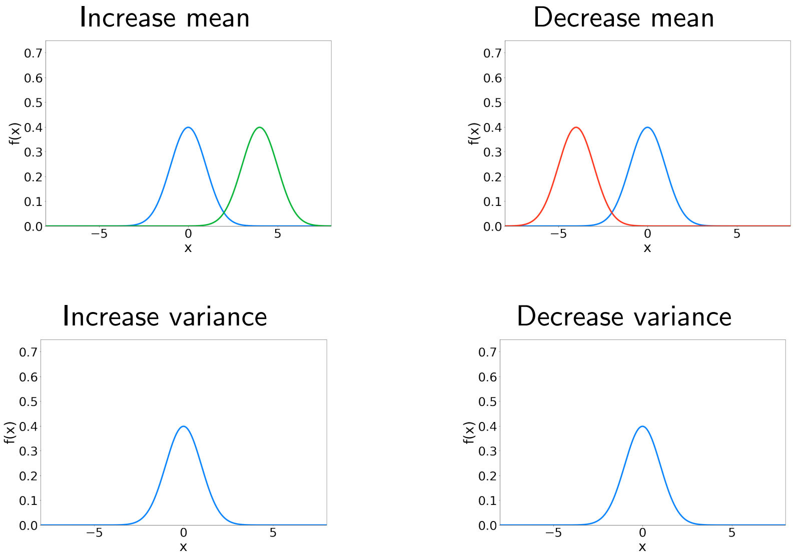
Increase variance



Decrease variance



Exponential The Normal Distribution

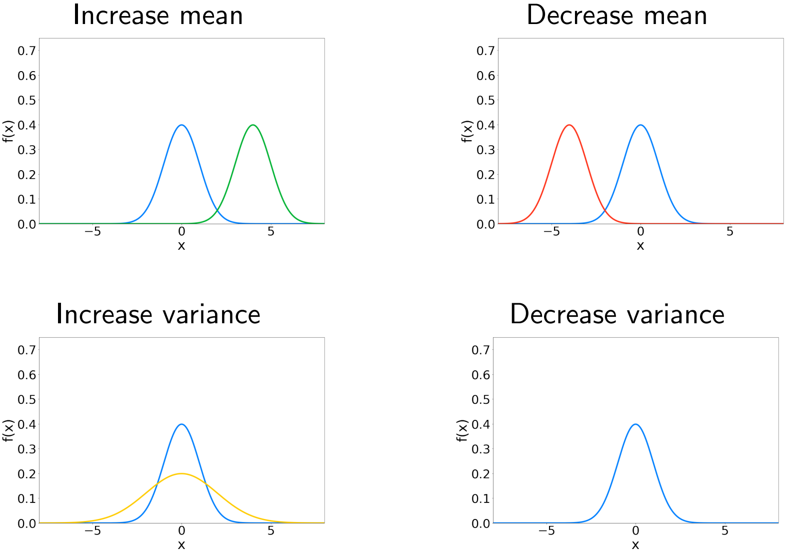

Increase mean

Decrease mean

Increase variance

Decrease variance

Exponential The Normal Distribution



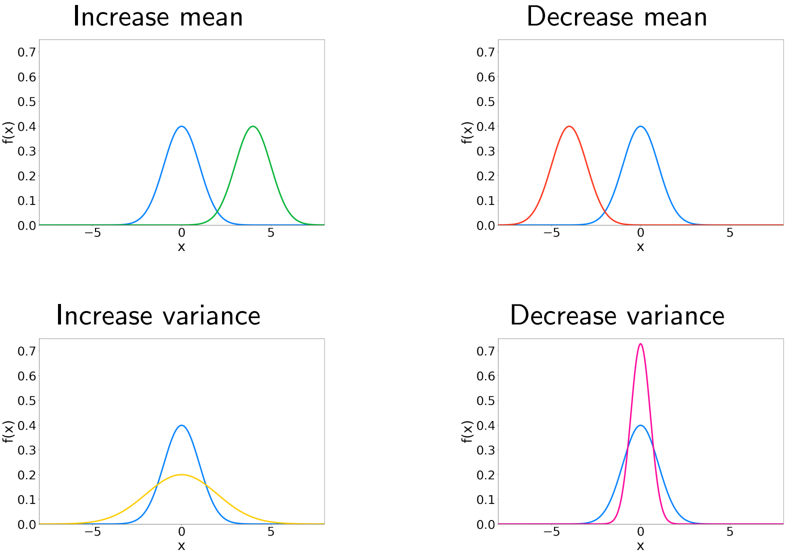
Increase mean

Decrease mean

Increase variance

Decrease variance

Exponential The Normal Distribution



Increase mean

Decrease mean

Increase variance

Decrease variance

The Normal Distribution

- The location is governed by μ.

The Normal Distribution

- The location is governed by μ.
- The spread is governed by σ.

The Normal Distribution

- The location is governed by μ.
- The spread is governed by σ.
- The height is also governed by σ since the area under each curve must be 1.

The Normal Distribution

- The location is governed by μ.
- The spread is governed by σ.
- The height is also governed by σ since the area under each curve must be 1.

- $\mu \pm \sigma$ includes 68% of observations.
- $\mu \pm 1.96\sigma$ includes 95% of observations.
- $\mu \pm 2.58\sigma$ includes 99% of observations.

The Standard Normal Distribution

- There is no closed form expression for the integral of the normal density function. Approximations are obtained using numerical methods and these are tabulated.

The Standard Normal Distribution

- There is no closed form expression for the integral of the normal density function. Approximations are obtained using numerical methods and these are tabulated.

A normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ is called a standard normal distribution. It is often denoted by Z.

The Standard Normal Distribution

- There is no closed form expression for the integral of the normal density function. Approximations are obtained using numerical methods and these are tabulated.

A normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ is called a standard normal distribution. It is often denoted by Z.

- Note: Since the normal distribution is symmetric. If $X \sim N(0, \sigma^2)$ then

$$\mathbb{P}(X < -a) = \mathbb{P}(X > a)$$

The Normal Distribution

If Y is a normal random variable with mean μ and variance σ^2,

$$Y \sim N(\mu, \sigma^2)$$

then,

$$Z = \frac{Y - \mu}{\sigma}$$

is a standard normal random variable.

$$Z \sim N(0, 1)$$

The Normal Distribution

If Y is a normal random variable with mean μ and variance σ^2,

$$Y \sim N(\mu, \sigma^2)$$

then,

$$Z = \frac{Y - \mu}{\sigma}$$

is a standard normal random variable.

$$Z \sim N(0, 1)$$

- This fact will be used to find probabilities from any normal distribution.

Example:

Suppose $X \sim N(\mu = 5, \sigma^2 = 3)$.

- Calculate $\mathbb{P}(X \leq 7)$.
- Calculate $\mathbb{P}(X \geq 6)$.
- Calculate $\mathbb{P}(X \leq 2)$.
- Calculate $\mathbb{P}(2.5 \leq X \leq 7)$

Example: Paper Friction

In a study to investigate the paper feeding process in a photocopier the coefficient of friction is a proportion which measures the degree of friction between adjacent sheets of paper in the paper stack. This coefficient is assumed to be normally distributed with mean 0.55 and standard deviation 0.013. During system operation the friction coefficient is measured at randomly selected times.

1. Find the probability that the friction coefficient falls between 0.53 and 0.56.
2. Is it likely to observe a friction coefficient below 0.52? Explain.

The Normal Distribution

- It can also be useful to use the normal tables in reverse.
- Suppose $X \sim N(2, 2)$, for what value a is $\mathbb{P}(X \leq a) = 0.1$?

$$\begin{aligned} \mathbb{P}(X \leq a) &= 0.1 \\ \mathbb{P}\left(\frac{X-2}{\sqrt{2}} \leq \frac{a-2}{\sqrt{2}}\right) &= 0.1 \end{aligned}$$

The Normal Distribution

- It can also be useful to use the normal tables in reverse.
- Suppose $X \sim N(2, 2)$, for what value a is $\mathbb{P}(X \leq a) = 0.1$?

$$\begin{aligned} \mathbb{P}(X \leq a) &= 0.1 \\ \mathbb{P}\left(\frac{X-2}{\sqrt{2}} \leq \frac{a-2}{\sqrt{2}}\right) &= 0.1 \end{aligned}$$

From tables $\mathbb{P}(Z \geq 1.28) = 0.1$ and since the standard normal distribution is symmetric this means that $\mathbb{P}(Z \leq -1.28) = 0.1$. So,

$$\begin{aligned} \frac{a-2}{\sqrt{2}} &= -1.28 \\ \Rightarrow a &= (-1.28)(\sqrt{2}) + 2 = 0.19 \end{aligned}$$

Example: Exam Results

Scores on an examination are assumed to be normally distributed with a mean of 78 and a variance of 36.

- Suppose that students scoring in the top 10% of this distribution are to receive an A grade. What is the minimum score a student must achieve to obtain an A grade?
- What must be the cutoff for passing the exam if the examiner wants only the lowest 25% of all scores to fail?
- Find, approximately, what proportion of students have scores 5 or more points above the passing cutoff.

Example: Skyscraper

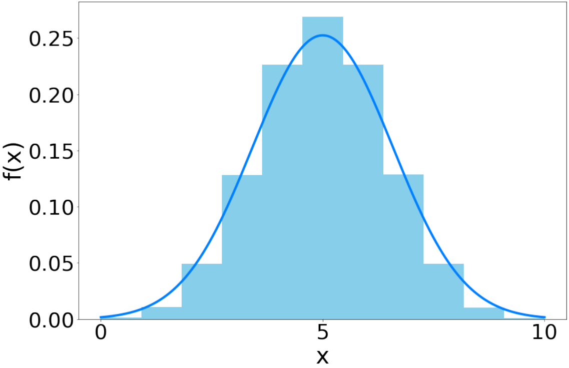
The lifetime of steel joists in a skyscraper are independent and normally distributed with a mean of 50 years and a standard deviation of 6 years.

1. What is the probability that a steel joist lasts more than 55 years?
2. What proportion of the steel joists would you expect to last between 42 and 58 years?
3. In a random sample of 4 joists what is the probability that 2 last less than 55 years?

Normal Distribution and Binomial Distribution

For large *n* (number of trials), the binomial distribution can be approximated using a normal distribution.

$$\text{binomial}(n, p) \sim N(np, np(1-p))$$



Summary

	Discrete RV	Continuous RV
	$x = 0, 1, 2, \dots$	$-\infty < x < \infty$
p.d.f.	$p(x)$	$f(x)$
Properties:	$0 \leq p(x) \leq 1$	$f(x) \geq 0$
	$\sum_{x=0}^{\infty} p(x) = 1$	$\int_{-\infty}^{\infty} f(x)\ dx = 1$
$\mathbb{P}(X = x)$	$p(x)$	0
c.d.f. $\mathbb{P}(X \leq x)$	$F(x) = \sum_{u=0}^{x} p(u)$	$F(x) = \int_{-\infty}^{x} f(u)\ du$
$\mathbb{P}(a \leq X \leq b)$ for $a \leq b$	$F(b) - F(a-1)$	$F(b) - F(a)$
$\mu \equiv \mathbb{E}(X)$	$\sum_{x=0}^{\infty} x\ p(x)$	$\int_{-\infty}^{\infty} x\ f(x)\ dx$
$\mathbb{E}(g(X))$	$\sum_{x=0}^{\infty} g(x)\ p(x)$	$\int_{-\infty}^{\infty} g(x)\ f(x)\ dx$
$\text{Var}(X) = \mathbb{E}(X^2) - \mu^2$	$\sum_{x=0}^{\infty} x^2\ p(x) - \mu^2$	$\int_{-\infty}^{\infty} x^2\ f(x)\ dx - \mu^2$

STAT10430 - Statistics with Python
Statistical Inference

Dr. Áine Byrne
aine.byrne@ucd.ie

Overview

- Estimators.
- Central Limit Theorem.
- Confidence Intervals.
- Hypothesis Testing.

Sample Statistics

- We are often interested in estimating a descriptive property of a population
    - e.g. Population mean, or population variance.

Sample Statistics

- We are often interested in estimating a descriptive property of a population
    - e.g. Population mean, or population variance.
- Estimate the property using a sample statistic calculated from a sample from the population.

Sample Statistics

- We are often interested in estimating a descriptive property of a population
  - e.g. Population mean, or population variance.

- Estimate the property using a sample statistic calculated from a sample from the population.

Parameters:	Sample Statistics:
Fixed values of population characteristics. e.g. Mean (μ), variance (σ^2)	A quantity calculated from the sample values. e.g. Sample mean ($\bar{X}$), sample variance (s^2)

Sample Statistics

- We are often interested in estimating a descriptive property of a population
  - e.g. Population mean, or population variance.

- Estimate the property using a sample statistic calculated from a sample from the population.

Parameters:
Fixed values of population characteristics.
e.g. Mean (μ), variance (σ^2)

Sample Statistics:
A quantity calculated from the sample values.
e.g. Sample mean ($\bar{X}$), sample variance (s^2)

- A sample statistic is a descriptive measure of a sample from the population.

Sampling Distribution

- Sample statistics are random variables since their value will vary from one sample to the next.

- A **sampling distribution** is the probability distribution of a sample statistic calculated from a sample of size n.

Sampling Distribution

- Sample statistics are random variables since their value will vary from one sample to the next.
- A **sampling distribution** is the probability distribution of a sample statistic calculated from a sample of size n.
- The sampling distribution describes how the statistic varies from one sample to the next.

Sampling Distribution

- Sample statistics are random variables since their value will vary from one sample to the next.
- A **sampling distribution** is the probability distribution of a sample statistic calculated from a sample of size n.
- The sampling distribution describes how the statistic varies from one sample to the next.
- In the light bulbs example the sample mean $\bar{X}$ was a good estimator of the population mean μ. The larger the sample, the closer the sample mean was to the population mean.

Comparing Estimators

- A statistic is an unbiased estimator if the mean of the sampling distribution is equal to the quantity of interest. E.g:

$$\mathbb{E}[\bar{X}] = \mu \implies \bar{X} \text{ is unbiased estimator of } \mu$$

Comparing Estimators

- A statistic is an unbiased estimator if the <u>mean</u> of the sampling distribution is equal to the quantity of interest.
  E.g:
  $$\mathbb{E}[\bar{X}] = \mu \implies \bar{X} \text{ is unbiased estimator of } \mu$$

- The standard error of a statistic is the <u>standard</u> <u>deviation</u> of the sampling distribution.
  E.g: The standard error of the mean can be expressed as:
  $$SE_{\bar{X}} = sd[\bar{X}] = \sqrt{Var[\bar{X}]} = \frac{\sigma}{\sqrt{n}}$$

Comparing Estimators

- A statistic is an unbiased estimator if the mean of the sampling distribution is equal to the quantity of interest. E.g:

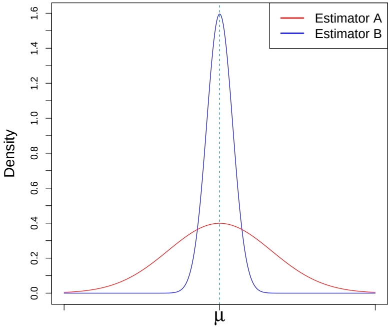
$$\mathbb{E}[\bar{X}] = \mu \implies \bar{X} \text{ is unbiased estimator of } \mu$$

- The standard error of a statistic is the standard deviation of the sampling distribution.
  E.g: The standard error of the mean can be expressed as:

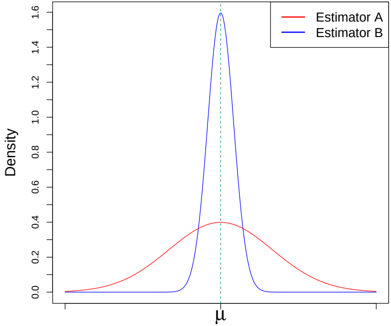
$$SE_{\bar{X}} = sd[\bar{X}] = \sqrt{Var[\bar{X}]} = \frac{\sigma}{\sqrt{n}}$$

- Generally speaking it is desirable for an estimator to be unbiased and to have a small standard error.

Comparing Estimators



Comparing Estimators



- Both estimators are unbiased.
- Estimator B is preferable due to its smaller standard error.

Example Estimators:

- $\bar{X}$ is an unbiased estimator of the population mean μ.
- s^2 is an unbiased estimator of the population variance σ^2.

Example Estimators:

- $\bar{X}$ is an unbiased estimator of the population mean μ.
- s^2 is an unbiased estimator of the population variance σ^2.

$$\mathbb{E}[s^2] = \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \sigma^2$$

Example Estimators:

- $\bar{X}$ is an unbiased estimator of the population mean μ.

- s^2 is an unbiased estimator of the population variance σ^2.

$$\mathbb{E}[s^2] = \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \sigma^2$$

- Thus for the light bulb example:
  - $\bar{X}$ provides an unbiased estimator of the mean lifetime of any light bulb.
  - s^2 provides an unbiased estimate of the variance of the lifetime of light bulbs.

Central Limit Theorem

The central limit theorem makes inferences about the sample mean easy.

Central Limit Theorem (CLT)

Given a sample of n independent observations from a population with mean μ and variance σ^2 then for n sufficiently large:

$$\bar{X} \ \dot{\sim} \ N\left(\mu, \frac{\sigma^2}{n}\right)$$

Central Limit Theorem

The central limit theorem makes inferences about the sample mean easy.

Central Limit Theorem (CLT)

Given a sample of n independent observations from a population with mean μ and variance σ^2 then for n sufficiently large:

$$\bar{X} \ \dot{\sim} \ N\left(\mu, \frac{\sigma^2}{n}\right)$$

- The larger the sample size n the better the approximation.

Central Limit Theorem

The central limit theorem makes inferences about the sample mean easy.

Central Limit Theorem (CLT)

Given a sample of n independent observations from a population with mean μ and variance σ^2 then for n sufficiently large:

$$\bar{X} \ \dot{\sim} \ N\left(\mu, \frac{\sigma^2}{n}\right)$$

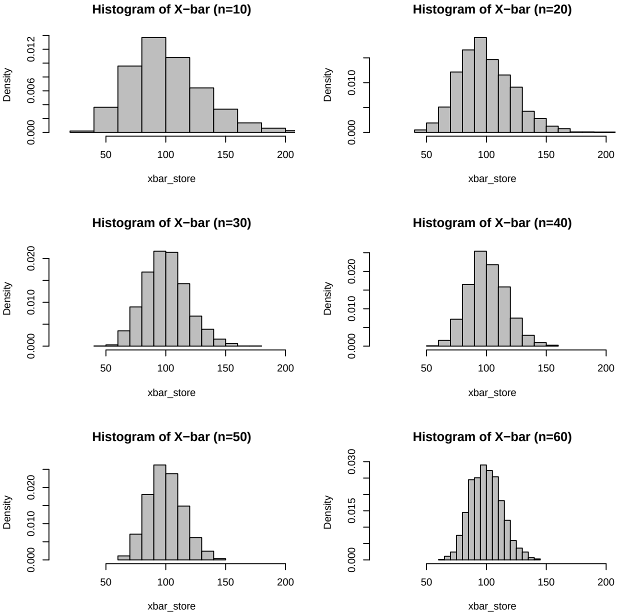
- The larger the sample size n the better the approximation.
- As a rule of thumb, for $n \geq 30$ the normal approximation will be reasonable.

Example: Formula 1

A Formula 1 racing team are interested in the distance covered (km) by a particular tyre, under reasonable weather conditions, until performance times suffer due to tyre wear. It is assumed that this distance follows an exponential distribution with unknown mean μ. The team test 50 sets of tyres and record the distance covered until tyre wear affects performance times.

- The team would like to estimate the unknown population mean parameter μ.

Example: Formula 1



Histogram of X−bar (n=10)

Histogram of X−bar (n=20)

Histogram of X−bar (n=30)

Histogram of X−bar (n=40)

Histogram of X−bar (n=50)

Histogram of X−bar (n=60)

Central Limit Theorem

- Note that the CLT applies when X follows any distribution. Thus for n sufficiently large $\bar{X}$ is approximately normal even if the underlying sample is not.

Central Limit Theorem

- Note that the CLT applies when X follows any distribution. Thus for n sufficiently large $\bar{X}$ is approximately normal even if the underlying sample is not.
- The CLT justifies the assumption of a normal distribution for any random variable which can be viewed as the sum of a large number of of independent and identically distributed quantities.

Central Limit Theorem

- Note that the CLT applies when X follows any distribution. Thus for n sufficiently large $\bar{X}$ is approximately normal even if the underlying sample is not.
- The CLT justifies the assumption of a normal distribution for any random variable which can be viewed as the sum of a large number of of independent and identically distributed quantities.
- For example measurement error is often assumed to be normally distributed since each error may be thought of as the sum of many smaller errors.

Example: CD Manufacture

Suppose that a sample of 50 items is taken from several batches of CDs produced in a particular factory and each is tested for defects. It is found that 10 CDs are defective.

- Give an unbiased estimator of the proportion of CDs produced by this factory which are defective.
- The company claim that less than 10% of their CDs are defective. If this is true what is the probability of observing more than 9 defective items in this sample.
- In light of this calculation comment on the company's claim.

Sampling Distribution of Sample Proportion

- In the previous example we used $\bar{P}$ to estimate the probability of a CD being defective.

- Thus by the CLT the sampling distribution of population proportion is approximately normal:

$$\bar{P} \dot{\sim} N\left(p, \frac{p(1-p)}{n}\right).$$

Confidence Intervals

- A confidence interval for a population parameter is an interval which almost certainly contains the true parameter.

Confidence Intervals

- A confidence interval for a population parameter is an interval which almost certainly contains the true parameter.
- Almost certainly usually means 95% certain.

If we are given a large sample of values from a population then a $100(1-\alpha)\%$ confidence interval for the population mean, μ, is:

$$\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Confidence Intervals

- A confidence interval for a population parameter is an interval which almost certainly contains the true parameter.
- Almost certainly usually means 95% certain.

If we are given a large sample of values from a population then a $100(1-\alpha)\%$ confidence interval for the population mean, μ, is:

$$\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

- $Z_{\frac{\alpha}{2}}$ is the value of the standard normal random variable Z such that $\mathbb{P}(Z > Z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$.

Confidence Intervals

- To form a 95% confidence interval $\alpha = 0.05$ and $Z_{\frac{\alpha}{2}} = 1.96$. This value is obtained from the standard normal percentage points table.

Confidence Intervals

- To form a 95% confidence interval $\alpha = 0.05$ and $Z_{\frac{\alpha}{2}} = 1.96$. This value is obtained from the standard normal percentage points table.

- The confidence level of the interval can be adjusted by changing the value of α and hence the value of $Z_{\frac{\alpha}{2}}$. A higher the degree of confidence for a particular sample will lead to a wider interval.

Confidence Intervals

- To form a 95% confidence interval $\alpha = 0.05$ and $Z_{\frac{\alpha}{2}} = 1.96$. This value is obtained from the standard normal percentage points table.
- The confidence level of the interval can be adjusted by changing the value of α and hence the value of $Z_{\frac{\alpha}{2}}$. A higher the degree of confidence for a particular sample will lead to a wider interval.
- Notice that the confidence interval defined on the previous slide will vary from sample to sample. 95% of such intervals will contain the true mean value.

Confidence Intervals

- To form a 95% confidence interval $\alpha = 0.05$ and $Z_{\frac{\alpha}{2}} = 1.96$. This value is obtained from the standard normal percentage points table.

- The confidence level of the interval can be adjusted by changing the value of α and hence the value of $Z_{\frac{\alpha}{2}}$. A higher the degree of confidence for a particular sample will lead to a wider interval.

- Notice that the confidence interval defined on the previous slide will vary from sample to sample. 95% of such intervals will contain the true mean value.

- The formula above is a direct result of the central limit theorem.

Example: C.I. for non-normal population

A random sample of 100 observations from a non-normally distributed population possesses a mean of 83.2 and a standard deviation of 6.4.

- Find a 95% confidence interval for the population mean μ and interpret the interval.
- Find a 99% confidence interval for μ.
- Comment on the width of the intervals.

Hypothesis Testing

- Hypothesis testing is one of the most widely used statistical procedures.

Hypothesis Testing

- Hypothesis testing is one of the most widely used statistical procedures.
- The question which hypothesis tests seek to answer is:

Is the relationship observed in the sample clear enough to be called statistically significant, or could it have been due to chance?

Hypothesis Testing

- Hypothesis testing is one of the most widely used statistical procedures.
- The question which hypothesis tests seek to answer is:

Is the relationship observed in the sample clear enough to be called statistically significant, or could it have been due to chance?

Examples:

- Dublin Bus says that the 145 bus comes every 10 minutes. You often wait longer than 10 minutes for the 145 and want to determine if Dublin Bus need to change their timetable.
- A pharmaceutical company make a new drug and want to test whether or not it is better than a drug currently on the market.

Basic Steps for Testing Hypotheses

1. Determine the null and alternative hypotheses.
2. Select an appropriate test and collect the data.
3. Execute the test.
4. Make a decision.

Step 1: Determine hypotheses.

There are always two hypotheses.

1. Null hypothesis (H_0): usually says that nothing changed or happened.

   e.g. That there is no relationship or that the relationship is due to chance.

Step 1: Determine hypotheses.

There are always two hypotheses.

1. Null hypothesis (H_0): usually says that nothing changed or happened.
   e.g. That there is no relationship or that the relationship is due to chance.
2. Alternative hypothesis (H_A): the research hypothesis. e.g. The researcher suspects that the status quo belief is incorrect and that there is indeed a relationship.

Step 1: Determine hypotheses.

There are always two hypotheses.

1. Null hypothesis (H_0): usually says that nothing changed or happened.
   e.g. That there is no relationship or that the relationship is due to chance.
2. Alternative hypothesis (H_A): the research hypothesis. e.g. The researcher suspects that the status quo belief is incorrect and that there is indeed a relationship.

The researcher needs to be quite sure before they reject the null hypothesis in favour of the alternative.

e.g. In a trial situation the hypotheses are:

$$H_0: \text{ Defendant is innocent. vs. } H_A: \text{ Defendant is guilty.}$$

Example: Determine hypotheses.

The average height of adults in Ireland is known to be 175 cm and you wish to test whether or not it has changed.

- The null hypothesis is that there is no change.

$$H_0 : \mu = 175 \text{ cm}$$

- Depending on what exactly you wish to test, there are 3 options for the alternative hypothesis.
  a) Upper one-tailed test: Height of Irish people has increased.

  $$H_A : \mu > 175 \text{ cm}$$

  b) Lower one-tailed test: Height of Irish people has decreased.

  $$H_A : \mu < 175 \text{ cm}$$

  c) Two-tailed test: Height of Irish people has changed.

  $$H_A : \mu \neq 175 \text{ cm}$$

Exercise: Determine hypotheses.

Last year the average GPA of students graduating with a Computer Science degree was 3.12. You wish to test whether the studentdfv s graduating this year obtained better results.

1. State the null hypothesis.
2. State the alternative hypothesis.

Step 2: Select appropriate test

- The decision in a hypothesis test is based on a single number summary of the observed data. This summary is called the test statistic.

- There are many different test statistics and the one used depends on the situation.

Step 3: Execute test

In order to decide if the results could be due to chance the following question is asked:

What is the probability that our observed test statistic was drawn the null distribution?

Step 3: Execute test

In order to decide if the results could be due to chance the following question is asked:

What is the probability that our observed test statistic was drawn the null distribution?

p-value

The *p*-value is computed by assuming the null hypothesis is true, and then asking how likely we would be to observe results at least as extreme as we have observed.

Step 3: Execute test

In order to decide if the results could be due to chance the following question is asked:

What is the probability that our observed test statistic was drawn the null distribution?

p-value

The *p*-value is computed by assuming the null hypothesis is true, and then asking how likely we would be to observe results at least as extreme as we have observed.

- The p-value does **not** give the probability that the null hypothesis is true
- A low p-value indicates an improbable event

Step 4: Make decision.

Once we know how likely/unlikely the observed test statistic is we face two choices:

Choice 1: The *p*-value is not small enough to convincingly rule out chance so we fail to reject the null hypothesis.

Step 4: Make decision.

Once we know how likely/unlikely the observed test statistic is we face two choices:

Choice 1: The p-value is not small enough to convincingly rule out chance so we fail to reject the null hypothesis.

Choice 2: The p-value is small enough to convincingly rule out chance so we reject the null hypothesis and accept the alternative hypothesis.

Step 4: Make decision.

Once we know how likely/unlikely the observed test statistic is we face two choices:

Choice 1: The *p*-value is not small enough to convincingly rule out chance so we fail to reject the null hypothesis.

Choice 2: The *p*-value is small enough to convincingly rule out chance so we reject the null hypothesis and accept the alternative hypothesis.

A *p*-value of less than a cut off point called the level of significance (α) is considered small enough to reject the null hypothesis. The standard level of significance is $\alpha = 0.05$.

Step 4: Make decision.

Courtroom example:

Choice 1: There is not enough evidence to prove the defendant is not innocent so he/she is not guilty

- Fail to reject the null hypothesis

Choice 2: There is enough evidence to rule out the possibility the defendant is innocent so he/she is guilty.

- Reject the null hypothesis and accept the alternative hypothesis

Example: Make decision.

Consider the Dublin bus example from slide 17.

- State the null and alternative hypotheses.
- What are the possible outcomes of your test?

Type I and Type II errors

Type I error

Rejecting the null hypothesis H_0 when it is in fact true is a type I error.

Courtroom example: Sending a innocent person to jail.

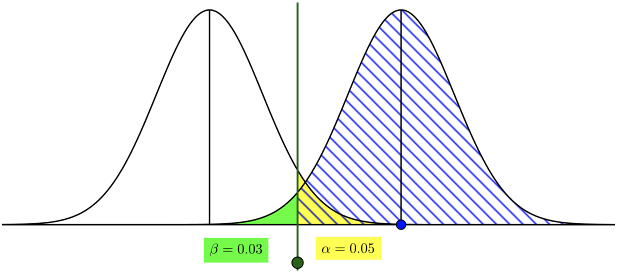
Type I error

Accepting the null hypothesis H_0 when it is in fact false is a type II error.

Courtroom example: Letting a guilty person walk free.

Type I and Type II errors

- The probability of making a type I error is equal to the significance level α.
- The probability of making a type II error depends on a number of things, such as the sample size and the significance level α.
- Decreasing the significance level α decreases the chance of making a Type I error, but increases the chance of making a Type II error.



Type I and Type II errors

- A type I error is often deemed more severe than a type II error. By rejecting the null hypothesis, you are changing the status quo.
- A type II error deems that there is not sufficient evidence to change the status quo at this time and nothing changes.

Type I and Type II errors

- A type I error is often deemed more severe than a type II error. By rejecting the null hypothesis, you are changing the status quo.
- A type II error deems that there is not sufficient evidence to change the status quo at this time and nothing changes.

Example: A pharmaceutical company is testing a new drug.

- Type I error: They concluded that the drug works, even though it does not. The drug is rolled out and patients switch from their old medication to this new ineffective medication.
- Type II error: There is not enough evidence to prove that the drug work, even though it does. Patients continue on their old medication and the pharmaceutical company continues its research and testing.

Z-Test

The Z-test is the most basic hypothesis test.

Condition for Z-test

The sample size is large (i.e. $n \geq 30$).

The population standard deviation is known.

Z-Test

The Z-test is the most basic hypothesis test.

Condition for Z-test

The sample size is large (i.e. $n \geq 30$).

The population standard deviation is known.

The test statistic for a Z-test is a Z-score:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

where μ_0 is the assumed population mean, $\bar{x}$ is the sample mean, σ is the population standard deviation and n is the sample size.

Z-Test for Population Means

1. Specify the null and alternative hypotheses.

   Null Hypothesis: The population mean is equal to some prior supposed value

   $$H_0 : \mu = \mu_0$$

   Alternative Hypothesis: There are 3 possibilities here.

   a) $H_A : \mu > \mu_0$, this is a one-tailed (upper tailed) test.

   b) $H_A : \mu < \mu_0$, this is a one-tailed (lower tailed) test.

   c) $H_A : \mu \neq \mu_0$, this is a two-tailed test.

Z-Test for Population Means

1. Specify the null and alternative hypotheses.

   Null Hypothesis: The population mean is equal to some prior supposed value

   $$H_0 : \mu = \mu_0$$

   Alternative Hypothesis: There are 3 possibilities here.

   a) $H_A : \mu > \mu_0$, this is a one-tailed (upper tailed) test.
   b) $H_A : \mu < \mu_0$, this is a one-tailed (lower tailed) test.
   c) $H_A : \mu \neq \mu_0$, this is a two-tailed test.

2. Compute the Z-score:

   $$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Z-Test Test for Population Mean

3. Compute the p-value by looking up the probability $P(Z < z)$ or $P(Z > z)$ in your statistical tables.

Z-Test Test for Population Mean

3. Compute the p-value by looking up the probability $P(Z < z)$ or $P(Z > z)$ in your statistical tables.

4. The threshold for our p-value depends on the significance level α and on the alternative hypothesis.

   a) If $P(Z > z) < \alpha$, we reject the null hypothesis

   b) If $P(Z < z) < \alpha$, we reject the null hypothesis

   c) If $P(Z > z) < \frac{\alpha}{2}$ or $P(Z < z) < \frac{\alpha}{2}$, we reject the null hypothesis

Example: Student Rent

A property developer claims that the average rental income per room in student accommodation is at most €5,000 per year with a standard deviation of €735. A random sample of 50 students were asked how much their annual rent was and the average was €5200. Do the sample results support the investors claim? (use $\alpha = 0.05$)

Example: Road Usage

A stretch of road is to be upgraded due to the volume of heavy freight traffic using this route. The local corporation say that the average number of heavy-duty vehicles using this road per hour is 71 with a standard deviation of 13.3. The engineers believe this number underestimates the true number. To test this, 50 one hour periods are selected over the course of the month and the number of heavy vehicles using the road in each hour are counted. The average number per hour is 74.1. Does the data support the engineers?

1. Test using a significance level of $\alpha = 0.1$.
2. Test using a significance level of $\alpha = 0.01$.

Rejection regions and *p*-values

- The rejection region is the region where the *p*-value is less than the significance level α and it depends on the alternative hypothesis.
  - a) Rejection region: $z > z_\alpha$.
  - b) Rejection region: $z < -z_\alpha$.
  - c) Rejection region: $z > z_{\frac{\alpha}{2}}$ or $z < -z_{\frac{\alpha}{2}}$
- Determining the rejection region for a test and finding that the test statistic is in it is equivalent to calculating the *p*-value for the test and finding that it is less than the desired confidence level α.

Z-Test for the Difference Between Two Means

- Sometimes it is of interest to compare the means of two populations to see if they are significantly different.

Z-Test for the Difference Between Two Means

- Sometimes it is of interest to compare the means of two populations to see if they are significantly different.
- e.g. Code efficiency. Two programmes designed to do the same job are compared to see if one is faster than the other. It is of interest to see if there is a significant difference between the means of the populations (denoted μ_1 and μ_2 respectively).

Z-Test for the Difference Between Two Means

- Sometimes it is of interest to compare the means of two populations to see if they are significantly different.
- e.g. Code efficiency. Two programmes designed to do the same job are compared to see if one is faster than the other. It is of interest to see if there is a significant difference between the means of the populations (denoted μ_1 and μ_2 respectively).
- Hypothesis tests and confidence intervals are available for this type of comparison. We focus on the case where we have large sample sizes.

Confidence Interval for $(\mu_1 - \mu_2)$

- By the CLT the sampling distribution of $(\bar{X}_1$ - $\bar{X}_2)$ is approximately:

$$(\bar{X}_1 - \bar{X}_2) \ \dot{\sim} \ N\left[(\mu_1 - \mu_2), \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)\right]$$

Confidence Interval for $(\mu_1 - \mu_2)$

- By the CLT the sampling distribution of $(\bar{X}_1$ - $\bar{X}_2)$ is approximately:

$$(\bar{X}_1 - \bar{X}_2) \ \dot{\sim} \ N\left[(\mu_1 - \mu_2), \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)\right]$$

- From this we can derive the form of a confidence interval.

If we are given large samples of values from two populations then a $100(1-\alpha)\%$ confidence interval for the difference between the population means, $(\mu_1 - \mu_2)$, is:

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Z-Test for $(\mu_1 - \mu_2)$

1. Specify the null and alternative hypotheses.

   Null Hypothesis:

   $$H_0 : (\mu_1 - \mu_2) = d_0$$

   Alternative Hypothesis:

   a) $H_A : (\mu_1 - \mu_2) > d_0$, this is a one-tailed (upper tailed) test.
   b) $H_A : (\mu_1 - \mu_2) < d_0$, this is a one-tailed (lower tailed) test.
   c) $H_A : (\mu_1 - \mu_2) \neq d_0$, this is a two-tailed test.

Z-Test for $(\mu_1 - \mu_2)$

1. Specify the null and alternative hypotheses.

   Null Hypothesis:

   $$H_0 : (\mu_1 - \mu_2) = d_0$$

   Alternative Hypothesis:

   a) $H_A : (\mu_1 - \mu_2) > d_0$, this is a one-tailed (upper tailed) test.
   b) $H_A : (\mu_1 - \mu_2) < d_0$, this is a one-tailed (lower tailed) test.
   c) $H_A : (\mu_1 - \mu_2) \neq d_0$, this is a two-tailed test.

2. The test statistic used in this situation is also a Z-score:

   $$z = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Z-Test for $(\mu_1 - \mu_2)$

3. Compute the p-value by looking up the probability associated with $Z < z$ or $Z > z$ in your statistical tables.

Z-Test for $(\mu_1 - \mu_2)$

3. Compute the p-value by looking up the probability associated with $Z < z$ or $Z > z$ in your statistical tables.

4. As with the previous Z-test, the p-value threshold depends on the significance level α and on the alternative hypothesis.
   a) If $P(Z > z) < \alpha$ we reject the null hypothesis
   b) If $P(Z < z) < \alpha$ we reject the null hypothesis
   c) If $P(Z > z) < \frac{\alpha}{2}$ or $P(Z < z) < \frac{\alpha}{2}$ we reject the null hypothesis

Z-Test for $(\mu_1 - \mu_2)$

3. Compute the p-value by looking up the probability associated with $Z < z$ or $Z > z$ in your statistical tables.

4. As with the previous Z-test, the p-value threshold depends on the significance level α and on the alternative hypothesis.
   a) If $P(Z > z) < \alpha$ we reject the null hypothesis
   b) If $P(Z < z) < \alpha$ we reject the null hypothesis
   c) If $P(Z > z) < \frac{\alpha}{2}$ or $P(Z < z) < \frac{\alpha}{2}$ we reject the null hypothesis

Remember: We can also think in terms of rejection regions and rejects the null hypothesis if

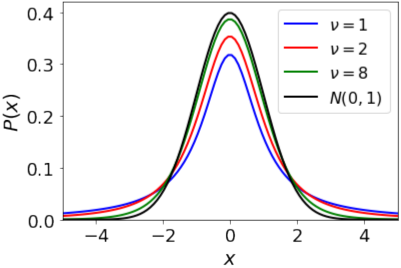
- $z > Z_\alpha$, for a upper one-tailed test.
- $z < -Z_\alpha$, for a lower one-tailed test.
- $z > Z_{\frac{\alpha}{2}}$ or $z < -Z_{\frac{\alpha}{2}}$, for a two-tailed test.

Example: Manual Dexterity and Sport

In a study to investigate the relationship between manual dexterity and sport, the manual dexterity of two groups of children was tested. The first group consisted of a random sample of 37 children who do not play sport and the sample mean and population standard deviation of this group was 31.68 and 4.56 respectively. The sample mean and population standard deviation of the second group was 32.19 and 4.34 respectively. The second group consisted of a random sample of 37 children who do play sport. Test the hypothesis that there is no difference between the mean manual dexterity scores ($H_0 : \mu_1 = \mu_2$) versus the alternative that that those who participate in sport have a higher average score ($H_A : \mu_1 < \mu_2$). Use $\alpha = 0.05$.

Student's t-test

- One of the most commonly used statistical test.
- Valid for smaller samples.
- Uses the sample standard deviation rather than the population standard deviation.
- Test statistic follows a Student's t-distribution



Degrees of freedom

- The degrees of freedom of a system are the number of things that can change freely.
- In statistics, this corresponds to the number of observations minus the number of constraints.
- If you have a set of 5 numbers with mean 10, you can choose the first four numbers freely, e.g. 4, 21, 7, 12. The value of the last number is *constrained* as the mean must be 10, hence, the fifth number must be 5. The degrees of freedom is 4.
- If we know the mean of a set of n numbers, the degrees of freedom is $n - 1$, the number of observations (n) minus the number of constraints (1).

t-test for Population Mean

1. Specify the null and alternative hypotheses.

   Null Hypothesis: The population mean is equal to some prior supposed value

   $$H_0 : \mu = \mu_0$$

   Alternative Hypothesis: As before there are 3 options.

   a) $H_A : \mu > \mu_0$, this is a one-tailed (upper tailed) test.
   b) $H_A : \mu < \mu_0$, this is a one-tailed (lower tailed) test.
   c) $H_A : \mu \neq \mu_0$, this is a two-tailed test.

t-test for Population Mean

1. Specify the null and alternative hypotheses.

   Null Hypothesis: The population mean is equal to some prior supposed value

   $$H_0 : \mu = \mu_0$$

   Alternative Hypothesis: As before there are 3 options.

   a) $H_A : \mu > \mu_0$, this is a one-tailed (upper tailed) test.
   b) $H_A : \mu < \mu_0$, this is a one-tailed (lower tailed) test.
   c) $H_A : \mu \neq \mu_0$, this is a two-tailed test.

2. The test statistic in this situation is a t-score:

   $$t_{n-1} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}},$$

   where the subscript $n - 1$ corresponds to the degrees of freedom.

t-test for Population Mean

3. Compute the p-value by looking up the probability associated with $T_{n-1} < t_{n-1}$ or $T_{n-1} > -t_{n-1}$ in your statistical tables. Pay particular attention to the degrees of freedom $n-1$. The probability values are different for different degrees of freedom.

t-test for Population Mean

3. Compute the p-value by looking up the probability associated with $T_{n-1} < t_{n-1}$ or $T_{n-1} > -t_{n-1}$ in your statistical tables. Pay particular attention to the degrees of freedom $n-1$. The probability values are different for different degrees of freedom.

4. As before, the threshold for the p-value depends on the significance level α and on the alternative hypothesis.
   a) If $P(Z > z_{\mu}) < \alpha$ we reject the null hypothesis
   b) If $P(Z < z_{\mu}) < \alpha$ we reject the null hypothesis
   c) If $P(Z > z_{\mu}) < \frac{\alpha}{2}$ or $P(Z < z_{\mu}) < \frac{\alpha}{2}$ we reject the null hypothesis

Example: Chocolate bars

A chocolate bar manufacturer states that their chocolate bars are 100 g. You are suspicious that the chocolate bars are lighter than the stated weight, and wish to test this theory. You buy 20 bars and weight them. The mean weight of your sample is 97.5 g and the standard deviation is 5 g.

1. What are the null and alternative hypothesis?
2. Compute the test statistic.
3. Is their sufficient evidence to reject the null hypothesis at a significance level of 0.05?
4. What about at a significance level of 0.01?

t-test for $(\mu_1 - \mu_2)$

- The null and alternative hypotheses are the same as for the Z-test.
- The degrees of freedom are $n_1 + n_2 - 2$, where n_1 is the size of the sample from distribution 1 and n_2 is the size of the sample from distribution 2.
- The test statistic is

$$t_{n_1+n_2-2} = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

- Compute the appropriate p-value and compare to the significance level.

Example: Study Techniques

You are conducting a study to see if students do better when they study all at once or in intervals. One group of 12 participants took a test after studying continuously for three hours. The other group of 12 participants took a test after studying for six thirty minute sessions. The first group had a mean score of 75 and a variance of 120. The second group had a mean score of 86 and a variance of 100.

1. What is the calculated t value? Are the mean test scores of these two groups significantly different at the .05 level?
2. What would the t value be if there were only 6 participants in each group? Would the scores be significant at the .05 level?

Notes:

- There is a correspondence between confidence intervals and hypothesis tests.

Notes:

- There is a correspondence between confidence intervals and hypothesis tests.

- Suppose we have a 95% confidence interval for a population parameter θ. This interval is equivalent to rejecting a null hypothesis that θ lies outside this interval at the 5% level.

Notes:

- There is a correspondence between confidence intervals and hypothesis tests.

- Suppose we have a 95% confidence interval for a population parameter θ. This interval is equivalent to rejecting a null hypothesis that θ lies outside this interval at the 5% level.

- There are many other hypothesis tests. For further examples see any introductory statistics text.

Notes:

- There is a correspondence between confidence intervals and hypothesis tests.

- Suppose we have a 95% confidence interval for a population parameter θ. This interval is equivalent to rejecting a null hypothesis that θ lies outside this interval at the 5% level.

- There are many other hypothesis tests. For further examples see any introductory statistics text.

- Each test has conditions which must be satisfied for the test to be valid.

STAT10430 - Statistics with Python
Simple Linear Regression

Dr. Áine Byrne
aine.byrne@ucd.ie

Deterministic Relationship

- A deterministic is one in which the value of one variable, Y, is completely determined by the value of another variable, X.

- In a deterministic relationship there is no allowance for error.

Deterministic Relationship

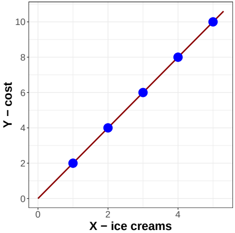
- A deterministic is one in which the value of one variable, Y, is completely determined by the value of another variable, X.
- In a deterministic relationship there is no allowance for error.
- e.g. Suppose the price of an ice-cream is 2 euro. Then the price paid, Y, for X ice-creams is:
$$Y = 2X$$

Deterministic Relationship

- A deterministic is one in which the value of one variable, Y, is completely determined by the value of another variable, X.

- In a deterministic relationship there is no allowance for error.

- e.g. Suppose the price of an ice-cream is 2 euro. Then the price paid, Y, for X ice-creams is:
$$Y = 2X$$



Probabilistic Relationship

- A probabilistic model is one which allows for unexplained variation or random error.

Probabilistic Relationship

- A probabilistic model is one which allows for unexplained variation or random error.
- Consists of a deterministic component and a random error component.

Probabilistic Relationship

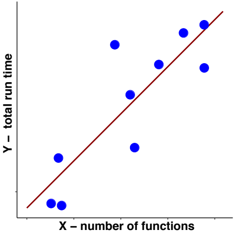
- A probabilistic model is one which allows for unexplained variation or random error.
- Consists of a deterministic component and a random error component.
- e.g. The amount of time to run a piece of code, Y, depends on number of functions, X, but also varies randomly from run to run.
  - $Y = 1.5X + \epsilon$
  - $Y =$ Deterministic Component $+$ Random Error

Probabilistic Relationship

- A probabilistic model is one which allows for unexplained variation or random error.

- Consists of a deterministic component and a random error component.

- e.g. The amount of time to run a piece of code, Y, depends on number of functions, X, but also varies randomly from run to run.
  - $Y = 1.5X + \epsilon$
  - $Y =$ Deterministic Component $+$ Random Error



Probabilistic Model

- The general form of a probabilistic model is:

$$Y = \text{Deterministic Component} + \text{Random Error}$$

where Y is the variable of interest.

Probabilistic Model

- The general form of a probabilistic model is:

$$Y = \text{Deterministic Component} + \text{Random Error}$$

  where Y is the variable of interest.

- It is assumed that the mean value of the random error is 0. This is equivalent to assuming that the mean of Y is equal to the deterministic component.

Probabilistic Model

- The general form of a probabilistic model is:

$$Y = \text{Deterministic Component} + \text{Random Error}$$

where Y is the variable of interest.

- It is assumed that the mean value of the random error is 0. This is equivalent to assuming that the mean of Y is equal to the deterministic component.
- We will focus on fitting straight line models. That is models whose deterministic component is a straight line.
- Plotting one variable against the other is helpful for deciding what kind of relationship exists between them.

Running Example: Grades

Students who do well on their continuous assessment tend to also do well on their final exam. The following data gives the final exam score for 12 students, Y, and their continuous assessment score, X. Plot the data. Is the relationship linear?

X	28	95	65	89	54	40	68	41	55	82	12	70
Y	38	80	54	95	36	31	77	58	58	71	30	56

Simple Linear Regression

The form of the model we will fit is:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

where

- Y – Dependent or response variable.
- X – Independent or predictor variable.
- $\mathbb{E}(Y) = \beta_0 + \beta_1 X =$ Deterministic Component.
- β_0 – Intercept of the line.
- β_1 – Slope of the line.
- ϵ – Random error.

Interpreting the Model Parameters

β_0

- β_0 is the intercept of the regression line, the point at which it crosses the y-axis.
- If the range of X includes 0 then β_0 is the expected value of Y when $X = 0$. If 0 is not in the range of X then we cannot make this interpretation.

β_1

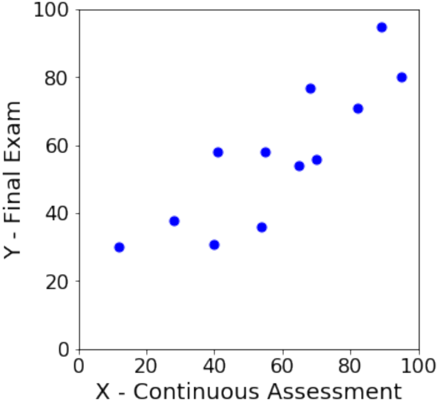
- β_1 is the slope of the regression line.
- It is the expected change in Y for every 1 unit increase in X.

Fitting the Line

- Consider the grades data $(X_i, Y_i), i = 1, \ldots, n$.

Fitting the Line

- Consider the grades data $(X_i, Y_i), i = 1, \ldots, n$.



Least Squares Estimation

- We will estimate $\mathbb{E}(Y_i) = \beta_0 + \beta_1 X_i$ with

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

so we need to find $\hat{\beta}_0$ and $\hat{\beta}_1$ which estimate β_0 and β_1.

Least Squares Estimation

- We will estimate $\mathbb{E}(Y_i) = \beta_0 + \beta_1 X_i$ with

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

  so we need to find $\hat{\beta}_0$ and $\hat{\beta}_1$ which estimate β_0 and β_1.

Least Squares Estimation:

- $Y_i - \hat{Y}_i$ – Difference between fitted line and observed value (error).

Least Squares Estimation

- We will estimate $\mathbb{E}(Y_i) = \beta_0 + \beta_1 X_i$ with
$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$
so we need to find $\hat{\beta}_0$ and $\hat{\beta}_1$ which estimate β_0 and β_1.

Least Squares Estimation:

- $Y_i - \hat{Y}_i$ – Difference between fitted line and observed value (error).
- $(Y_i - \hat{Y}_i)^2$ – Squared error.

Least Squares Estimation

- We will estimate $\mathbb{E}(Y_i) = \beta_0 + \beta_1 X_i$ with

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

so we need to find $\hat{\beta}_0$ and $\hat{\beta}_1$ which estimate β_0 and β_1.

Least Squares Estimation:

- $Y_i - \hat{Y}_i$ – Difference between fitted line and observed value (error).
- $(Y_i - \hat{Y}_i)^2$ – Squared error.
- $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ – Error sum of squares (SS_E).

Least Squares Estimation

- We will estimate $\mathbb{E}(Y_i) = \beta_0 + \beta_1 X_i$ with

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

so we need to find $\hat{\beta}_0$ and $\hat{\beta}_1$ which estimate β_0 and β_1.

Least Squares Estimation:

- $Y_i - \hat{Y}_i$ – Difference between fitted line and observed value (error).
- $(Y_i - \hat{Y}_i)^2$ – Squared error.
- $\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$ – Error sum of squares (SS_E).

- Idea: Find the line (i.e. $\hat{\beta}_0$ and $\hat{\beta}_1$) which minimises the error sum of squares.

Least Squares Estimates

Calculus is used to find formulae for $\hat{\beta}_0$ and $\hat{\beta}_1$ which minimise the SS_E. These are presented in the box below.

Least Squares Estimates

Calculus is used to find formulae for $\hat{\beta}_0$ and $\hat{\beta}_1$ which minimise the SS_E. These are presented in the box below.

Formulae:

- $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$
- $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

Where

- $S_{xy} = \sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{N} X_i Y_i - n\bar{X}\bar{Y}$
- $S_{xx} = \sum_{i=1}^{N} (X_i - \bar{X})^2 = \sum_{i=1}^{N} X_i^2 - n\bar{X}^2$

Derivation: https://uk.sagepub.com/sites/default/files/upm-assets/17668_book_item_17668.pdf

Example: Grades

X	28	95	65	89	54	40	68	41	55	82	12	70
Y	38	80	54	95	36	31	77	58	58	71	30	56

- Work out $\bar{X}$, $\bar{Y}$, $\sum_{i=1}^{n} X_i^2$ and $\sum_{i=1}^{n} X_i Y_i$
- Compute S_{xx} and S_{xy}.
- Use these values to calculate $\hat{\beta}_0$ and $\hat{\beta}_1$.

Assumptions of the Model

In fitting the simple linear regression model several assumptions are made. These center around the random component ϵ_i.

1. The mean of the probability distribution of ϵ_i is 0. This assumption implies that $\mathbb{E}(Y_i) = \beta_0 + \beta_1 X_i$.

Assumptions of the Model

In fitting the simple linear regression model several assumptions are made. These center around the random component ϵ_i.

1. The mean of the probability distribution of ϵ_i is 0. This assumption implies that $\mathbb{E}(Y_i) = \beta_0 + \beta_1 X_i$.

2. The variance of the probability distribution of ϵ_i is σ^2 for all values of X_i. That is $Var(\epsilon_i) = \sigma^2$ regardless of the value of X_i.

Assumptions of the Model

In fitting the simple linear regression model several assumptions are made. These center around the random component ϵ_i.

1. The mean of the probability distribution of ϵ_i is 0. This assumption implies that $\mathbb{E}(Y_i) = \beta_0 + \beta_1 X_i$.

2. The variance of the probability distribution of ϵ_i is σ^2 for all values of X_i. That is $Var(\epsilon_i) = \sigma^2$ regardless of the value of X_i.

3. The random component ϵ_i is normally distributed.

Assumptions of the Model

In fitting the simple linear regression model several assumptions are made. These center around the random component ϵ_i.

1. The mean of the probability distribution of ϵ_i is 0. This assumption implies that $\mathbb{E}(Y_i) = \beta_0 + \beta_1 X_i$.

2. The variance of the probability distribution of ϵ_i is σ^2 for all values of X_i. That is $Var(\epsilon_i) = \sigma^2$ regardless of the value of X_i.

3. The random component ϵ_i is normally distributed.

4. The value of ϵ associated with one value of Y has no effect on the value of ϵ associated with other Y values. That is ϵ_i and ϵ_j are independent for $i \neq j$.

Benefits and Limitations of Linear Regression

Benefits:

- The model allows us to make predictions for the value of the response variable Y corresponding to values of the predictor variable X which are not in the data set.

Benefits and Limitations of Linear Regression

Benefits:

- The model allows us to make predictions for the value of the response variable Y corresponding to values of the predictor variable X which are not in the data set.

- We can use the model to assess the strength of the relationship between the two variables (more later).

Benefits and Limitations of Linear Regression

Benefits:

- The model allows us to make predictions for the value of the response variable Y corresponding to values of the predictor variable X which are not in the data set.
- We can use the model to assess the strength of the relationship between the two variables (more later).
- The model allows a clear understanding of the relationship between X and Y that is easily explained.

Benefits and Limitations of Linear Regression

Limitations:

- We cannot predict, with confidence, the values of Y for values of X outside the range of the original data set.

Benefits and Limitations of Linear Regression

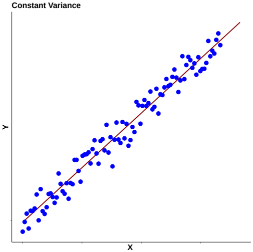
Limitations:

- We cannot predict, with confidence, the values of Y for values of X outside the range of the original data set.
- Later we will make inferences about the model parameter β_1. These inferences are only valid if the assumption of constant variance for ϵ_i is satisfied.

Benefits and Limitations of Linear Regression

Limitations:

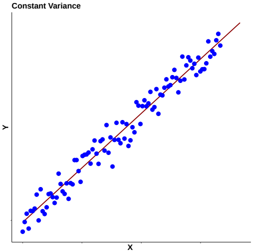
- We cannot predict, with confidence, the values of Y for values of X outside the range of the original data set.
- Later we will make inferences about the model parameter β_1. These inferences are only valid if the assumption of constant variance for ϵ_i is satisfied.

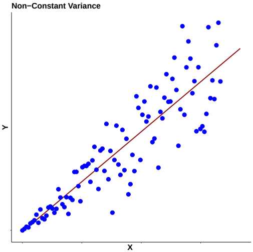


Benefits and Limitations of Linear Regression

Limitations:

- We cannot predict, with confidence, the values of Y for values of X outside the range of the original data set.
- Later we will make inferences about the model parameter β_1. These inferences are only valid if the assumption of constant variance for ϵ_i is satisfied.

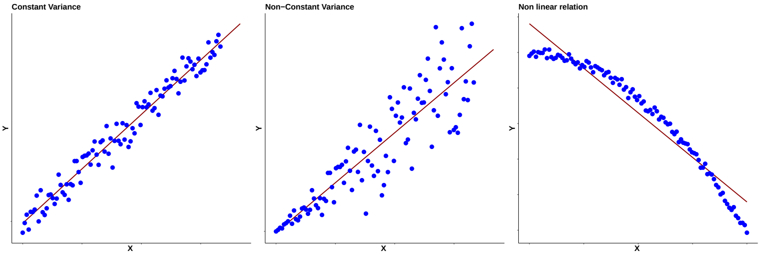




Benefits and Limitations of Linear Regression

Limitations:

- We cannot predict, with confidence, the values of Y for values of X outside the range of the original data set.
- Later we will make inferences about the model parameter β_1. These inferences are only valid if the assumption of constant variance for ϵ_i is satisfied.



An Estimator of σ^2

- In order to construct confidence intervals and perform hypothesis tests for β_1 we need an estimate of σ^2 the variance of ϵ_i.

An Estimator of σ^2

- In order to construct confidence intervals and perform hypothesis tests for β_1 we need an estimate of σ^2 the variance of ϵ_i.

- The mean square for error (MS_E) is an unbiased estimator of σ^2.

An Estimator of σ^2

- In order to construct confidence intervals and perform hypothesis tests for β_1 we need an estimate of σ^2 the variance of ϵ_i.

- The mean square for error (MS_E) is an unbiased estimator of σ^2.

Estimation of σ^2

$$MS_E = \frac{SS_E}{n-2}$$

where $SS_E = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = S_{yy} - \hat{\beta}_1 S_{xy}$
and $S_{yy} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2$.

An Estimator of σ^2

- In order to construct confidence intervals and perform hypothesis tests for β_1 we need an estimate of σ^2 the variance of ϵ_i.

- The mean square for error (MS_E) is an unbiased estimator of σ^2.

Estimation of σ^2

$$MS_E = \frac{SS_E}{n-2}$$

where $SS_E = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = S_{yy} - \hat{\beta}_1 S_{xy}$
and $S_{yy} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2$.

- This estimate is crucial for assessing the utility of the model.

Example: Grades

X	28	95	65	89	54	40	68	41	55	82	12	70
Y	38	80	54	95	36	31	77	58	58	71	30	56

On slide 11, we found $\bar{Y} = 57$, $S_{xy} = 4876$ and $\hat{\beta}_1 = 0.712$.

- Calculate $\sum_{i=1}^{n} Y_i^2$.
- Use the above values to compute S_{yy}.
- Calculate SS_E.
- What is the mean square error MS_E?

Assessing Model Utility

- We are interested in investigating if Y really does depend on X or not.

Assessing Model Utility

- We are interested in investigating if Y really does depend on X or not.

- Assessing the utility of the linear regression model boils down to making inferences about the slope β_1.

Assessing Model Utility

- We are interested in investigating if Y really does depend on X or not.
- Assessing the utility of the linear regression model boils down to making inferences about the slope β_1.
- If the predictor variable X is unrelated to the response variable Y then $\mathbb{E}(Y) = \beta_0 + \beta_1 X_i$ will remain unchanged as X changes, i.e. $\beta_1 = 0$.

Assessing Model Utility

- We are interested in investigating if Y really does depend on X or not.

- Assessing the utility of the linear regression model boils down to making inferences about the slope β_1.

- If the predictor variable X is unrelated to the response variable Y then $\mathbb{E}(Y) = \beta_0 + \beta_1 X_i$ will remain unchanged as X changes, i.e. $\beta_1 = 0$.

- Now that we have assumed the form of the probability distribution of ϵ_i and have an estimate of σ^2 we can perform a hypothesis test to see if β_1 is significantly different from 0.

A Test of Model Utility

We use a t-test to test model utility

1. Specify the null and alternative hypotheses.

   Null hypothesis: Y does not depend on X

   $$H_0 : \beta_1 = 0$$

   Alternative hypothesis:

   a) Positive linear relationship. $H_A : \beta_1 > 0$, upper one-tailed test.
   b) Negative linear relationship. $H_A : \beta_1 < 0$, lower one-tailed test.
   c) Linear relationship. $H_A : \beta_1 \neq 0$, two-tailed test.

A Test of Model Utility

We use a t-test to test model utility

1. Specify the null and alternative hypotheses.

   Null hypothesis: Y does not depend on X

   $$H_0 : \beta_1 = 0$$

   Alternative hypothesis:

   a) Positive linear relationship. $H_A : \beta_1 > 0$, upper one-tailed test.
   b) Negative linear relationship. $H_A : \beta_1 < 0$, lower one-tailed test.
   c) Linear relationship. $H_A : \beta_1 \neq 0$, two-tailed test.

2. The test statistic is a t statistic:

$$t_{n-2} = \frac{\hat{\beta}_1}{\sqrt{\frac{MS_E}{S_{xx}}}}$$

A Test of Model Utility

3. Compute the p-value by looking up the probability $P(T_{n-2} < t_{n-2})$ or $P(T_{n-2} > t_{n-2)}$ in your statistical tables.

A Test of Model Utility

3. Compute the p-value by looking up the probability $P(T_{n-2} < t_{n-2})$ or $P(T_{n-2} > t_{n-2)}$ in your statistical tables.

4. As before, the threshold for the p-value depends on the significance level α and on the alternative hypothesis.
   a) If $P(T_{n-2} > t_{n-2}) < \alpha$ we reject the null hypothesis
   b) If $P(T_{n-2} < t_{n-2}) < \alpha$ we reject the null hypothesis
   c) If $P(T_{n-2} > t_{n-2}) < \frac{\alpha}{2}$ or $P(T_{n-2} < t_{n-2}) < \frac{\alpha}{2}$ we reject the null hypothesis

A Test of Model Utility

3. Compute the p-value by looking up the probability $P(T_{n-2} < t_{n-2})$ or $P(T_{n-2} > t_{n-2)}$ in your statistical tables.

4. As before, the threshold for the p-value depends on the significance level α and on the alternative hypothesis.
   a) If $P(T_{n-2} > t_{n-2}) < \alpha$ we reject the null hypothesis
   b) If $P(T_{n-2} < t_{n-2}) < \alpha$ we reject the null hypothesis
   c) If $P(T_{n-2} > t_{n-2}) < \frac{\alpha}{2}$ or $P(T_{n-2} < t_{n-2}) < \frac{\alpha}{2}$ we reject the null hypothesis

Conditions:

- The four assumptions we made about ϵ must be satisfied.

Example: Grades

From slide 11 and 16, we have $\hat{\beta}_1 = 0.712$, $S_{xx} = 6852.25$ and $MS_E = 131.63$. Perform a t-test to assess if there is sufficient evidence for a linear relationship.

Confidence Interval for β_1

- It is also possible to calculate a confidence for β_1 using the t statistic above.

Confidence Interval for β_1

A $100(1-\alpha)\%$ confidence interval for β_1 is:

$$\hat{\beta}_1 \pm t_{n-2,\frac{\alpha}{2}}\sqrt{\frac{MS_E}{S_{xx}}}$$

- The four assumptions made about ϵ earlier are required for this to be a valid confidence interval.

Example: Grades

As with the t-test, we have $\hat{\beta}_1 = 0.712$, $S_{xx} = 6852.25$ and $MS_E = 131.63$. Compute the 95% confidence interval for β_1.

Predicting from the model

- Suppose we wish to estimate the value of the response variable for a new value of X, X^*, which is within the range of X.

Predicting from the model

- Suppose we wish to estimate the value of the response variable for a new value of X, X^*, which is within the range of X.

- We can obtain an estimate of the mean of Y when $X = X^*$ by simply plugging X^* in to the estimated regression equation.

$$\hat{Y}^* = \hat{\beta}_0 + \hat{\beta}_1 X^*$$

Example: Grades

Estimate the grade in the final exam of a student who received 54% in their continuous assessment.

Coefficient of Correlation

Coefficient of Correlation

The coefficient of correlation, r,is a measure of the strength of a *linear* relationship between two variables X and Y.

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

Coefficient of Correlation

Coefficient of Correlation

The coefficient of correlation, r,is a measure of the strength of a *linear* relationship between two variables X and Y.

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

The correlation coefficient is scaleless and assumes a value between -1 and $+1$.

Coefficient of Correlation

- A value of r near 0 means there is little or no relationship between X and Y.

Coefficient of Correlation

- A value of r near 0 means there is little or no relationship between X and Y.
- Values of r near 1 or -1 imply a strong relationship between X and Y.

Coefficient of Correlation

- A value of r near 0 means there is little or no relationship between X and Y.
- Values of r near 1 or -1 imply a strong relationship between X and Y.
- If $r > 0$ then the variables are positively correlated, i.e. as X increases Y increases.

Coefficient of Correlation

- A value of r near 0 means there is little or no relationship between X and Y.
- Values of r near 1 or -1 imply a strong relationship between X and Y.
- If $r > 0$ then the variables are positively correlated, i.e. as X increases Y increases.
- If $r < 0$ then the variables are negatively correlated, i.e. as X increases Y decreases.

Coefficient of Determination

Coefficient of Determination

The coefficient of determination is:

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

It represents the proportion of the total sample variability around $\bar{Y}$ which is explained by the linear relationship between Y and X

Coefficient of Determination

Coefficient of Determination

The coefficient of determination is:

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

It represents the proportion of the total sample variability around $\bar{Y}$ which is explained by the linear relationship between Y and X

- R^2 is always between 0 and 1.

Coefficient of Determination

Coefficient of Determination

The coefficient of determination is:

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

It represents the proportion of the total sample variability around $\bar{Y}$ which is explained by the linear relationship between Y and X

- R^2 is always between 0 and 1.
- $R^2 \approx 0 \Rightarrow$ Poor Fit, $R^2 \approx 1 \Rightarrow$ Good Fit.

Coefficient of Determination

Coefficient of Determination

The coefficient of determination is:

$$R^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}}$$

It represents the proportion of the total sample variability around $\bar{Y}$ which is explained by the linear relationship between Y and X

- R^2 is always between 0 and 1.
- $R^2 \approx 0 \Rightarrow$ Poor Fit, $R^2 \approx 1 \Rightarrow$ Good Fit.
- $R^2 = 0.8$ means that 80% of the variation in Y is explained by its linear relationship with X.

Example: Grades

Compute the coefficient of correlation and coefficient of determination for the grades example. We previously worked out $S_{xx} = 6852.25$, $S_{xy} = 4876.0$ and $S_{yy} = 4788.0$.

Checking the Model Assumptions

- Model : $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

Checking the Model Assumptions

- Model : $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$
- Under the model it is assumed that:
  - $\epsilon_i \sim N(0, \sigma^2)$ for $i = 1, \dots, n$.

Checking the Model Assumptions

- Model : $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

- Under the model it is assumed that:
    - $\epsilon_i \sim N(0, \sigma^2)$ for $i = 1, \ldots, n$.
    - ϵ_i and ϵ_j are independent if $i \neq j$.

- $\hat{\epsilon}_i = Y_i - \hat{Y}_i$ – Estimate of ϵ_i (residuals).

Checking the Model Assumptions

- Model : $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$
- Under the model it is assumed that:
  - $\epsilon_i \sim N(0, \sigma^2)$ for $i = 1, \ldots, n$.
  - ϵ_i and ϵ_j are independent if $i \neq j$.
- $\hat{\epsilon}_i = Y_i - \hat{Y}_i$ – Estimate of ϵ_i (residuals).
- Inferences made about the model parameters are invalid if these assumptions are false.

Checking the Model Assumptions

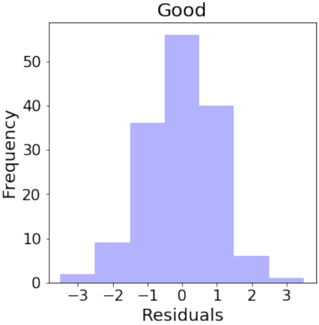
- Model : $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$
- Under the model it is assumed that:
  - $\epsilon_i \sim N(0, \sigma^2)$ for $i = 1, \ldots, n$.
  - ϵ_i and ϵ_j are independent if $i \neq j$.
- $\hat{\epsilon}_i = Y_i - \hat{Y}_i$ – Estimate of ϵ_i (residuals).
- Inferences made about the model parameters are invalid if these assumptions are false.
- Use plots to check assumptions are valid.

Checking the Model Assumptions

Histogram of Residuals: This plot should be roughly bell shaped if the normality assumption is true. This requires a reasonable amount of data.

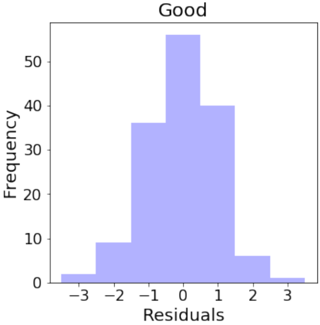
Checking the Model Assumptions

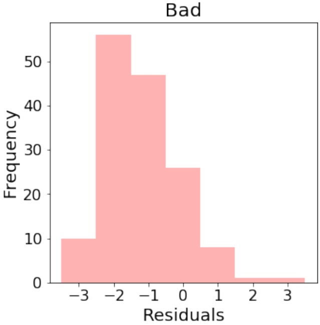
Histogram of Residuals: This plot should be roughly bell shaped if the normality assumption is true. This requires a reasonable amount of data.



Checking the Model Assumptions

Histogram of Residuals: This plot should be roughly bell shaped if the normality assumption is true. This requires a reasonable amount of data.





Checking the Model Assumptions

Scatterplot of Residuals Against X:

- This plot should show a random scatter of points about the x-axis if the residuals have mean 0.

Checking the Model Assumptions

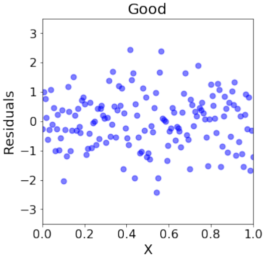
Scatterplot of Residuals Against X:

- This plot should show a random scatter of points about the x-axis if the residuals have mean 0.
- If the constant variance assumption is satisfied, then the variance of the scatter should not grow or shrink, it should be constant.

Checking the Model Assumptions

Scatterplot of Residuals Against X:

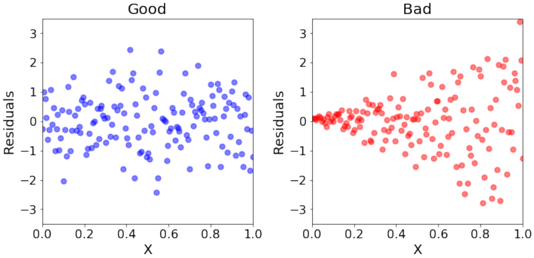
- This plot should show a random scatter of points about the x-axis if the residuals have mean 0.
- If the constant variance assumption is satisfied, then the variance of the scatter should not grow or shrink, it should be constant.



Checking the Model Assumptions

Scatterplot of Residuals Against X:

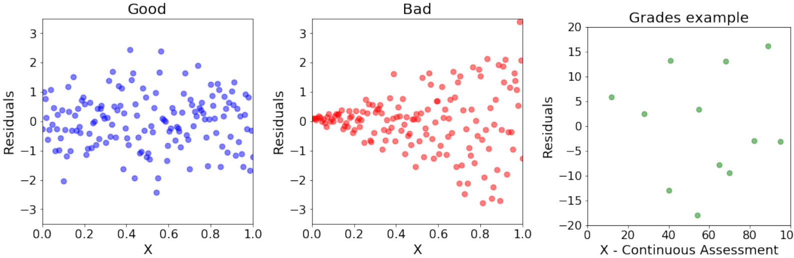
- This plot should show a random scatter of points about the x-axis if the residuals have mean 0.
- If the constant variance assumption is satisfied, then the variance of the scatter should not grow or shrink, it should be constant.



Checking the Model Assumptions

Scatterplot of Residuals Against X:

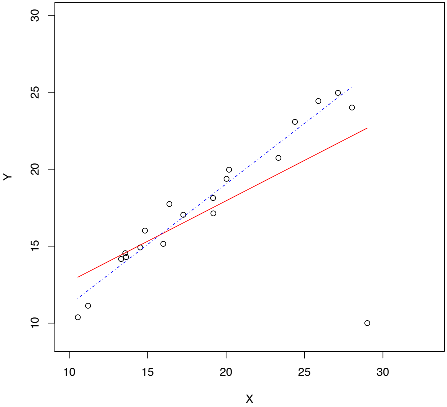
- This plot should show a random scatter of points about the x-axis if the residuals have mean 0.
- If the constant variance assumption is satisfied, then the variance of the scatter should not grow or shrink, it should be constant.



Outliers

- An observation which lies significantly away from the bulk of the data is known as an outlier.
- Outliers can be influential.
- Outliers are sometimes excluded from the data for more accurate results.
- It is important to plot the data first.

Outliers



Complete Example:

Research was conducted to investigate the relationship between the price of a particular product and the number of units sold. The results are shown below:

Price (€)	25	27	29	30	35	39	42	45
Units sold	260	250	246	240	227	232	200	196

1. Draw a scatter plot to verify the assumption that the relationship is linear. Let X be the price and Y the units sold.
2. Fit a straight line to the data using the method of least squares.
3. Is there a statistically significant relationship between the variables? Use $\alpha = 0.05$.
4. Calculate the coefficient of determination R^2.