University of Essex

Online

Research Methods and Professional Practice – June 2022

Seminar 4

University of Essex

Online

Agenda

- Seminar 4
- Quantitative methods
- Inferential Statistics—formative assignment
- Next Steps

University of Essex

Online

Tutor Contact

Dr Steph Paladini

Email: stefania.paladini@kaplan.com

Office Hours:

Monday 5pm - 7pm (17:00 - 19:00)

1.	Seminar 1 - Introduction
3.	**Seminar 2 - Peer review Activity**
4.	**Seminar 3 - Case Study: Privacy**
8.	**Seminar 4 - Inferential Statistics Workshop**
9.	**Seminar 5 - Workshop on Presenting Results**
11.	**Seminar 6 - e-Portfolio Preparation**

University of Essex

Online

Module Structure

Link to the calendar:
https://www.my-course.co.uk/course/view.php?id=8475&section=2

		Weighting
Literature Review	- Critically evaluate existing literature, research design and methodology for a chosen topic and so produce a literature review on this topic. - 2000 words	30%
Research Proposal Presentation	Presentation: about 15 minutes, you can work against a minimum of 1500 to a maximum of 2250 words in your oral presentation /transcript.	30%
Individual e-Portfolio	The strict word count limit for the e-portfolio applies to the **reflection only - 1,000 words**	40%

University of Essex

Online

Assessment -- Formative

Formative and e-Portfolio Activities

To aid your development of an in-depth understanding of the syllabus, regular formative assessment is provided via case studies, exercises, and reflective commentaries. Furthermore, the formative feedback received will enable you to develop your understanding of what is required for the summative assessments. Although not all of these activities are weighted components that count towards your module grade, they are designed to help you gain a deep understanding of the module content and provide a foundation for understanding.

We recommend that you participate in all formative activities and make frequent entries to your **e-portfolio which is assessed in this module**.

You will find the main e-portfolio and formative activities below. Please make sure that you read the guidance in each unit to fully understand weekly study requirements. Also read the full e-Portfolio guidance for the final submission in unit 12.

Unit(s)	Component	Deadline	e-Portfolio/Formative
1 - 3	Collaborative discussion 1	End of unit 3	e-Portfolio
1	Reasoning Quiz	End of unit 1	Formative
1	Reflective Activity 1	End of unit 1	e-Portfolio/Formative
4	e-Portfolio Activity: Literature Review Outline	End of unit 4	e-Portfolio/Formative
6	Wiki Activity	End of unit 6	Formative
6	e-Portfolio Update	End of unit 6	e-Portfolio/Formative
7 - 9	Collaborative discussion 2	End of unit 9	e-Portfolio
8	e-Portfolio Activity: Research Proposal Outline	End of unit 8	e-Portfolio/Formative
8 - 9	Statistical Worksheet Submissions	End of unit 10	e-Portfolio
12	Self Test Quiz	End of unit 12	Formative

University of Essex

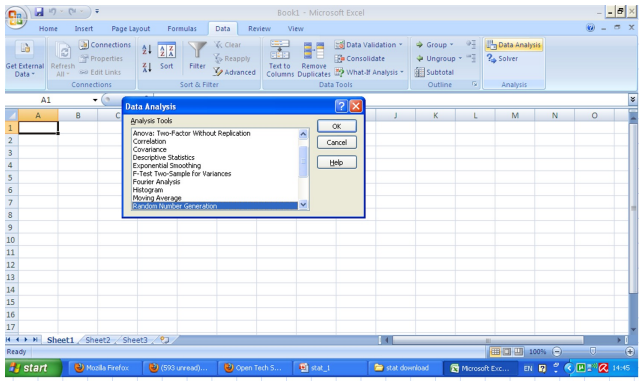
Online

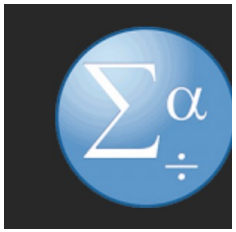
Research Methods and Professional Practice – Unit

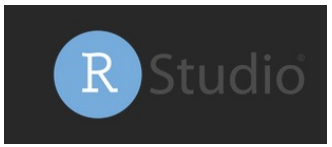
Today's Discussion:

- Quantitative Methods [what they are]
- The use of (Descriptive) Statistics to summarise data
- Inferential Statistics: what do we need to know?

+ Something about our Research Proposal Outline







University of Essex

Online

Research Methods and Professional Practice – Statistics

Summary: Using statistics - Four examples of its uses

* Sum up /make sense of a group of data with one or two figures: **Descriptive statistics**

*Making sense of the world from a little piece: **Inferential statistics**

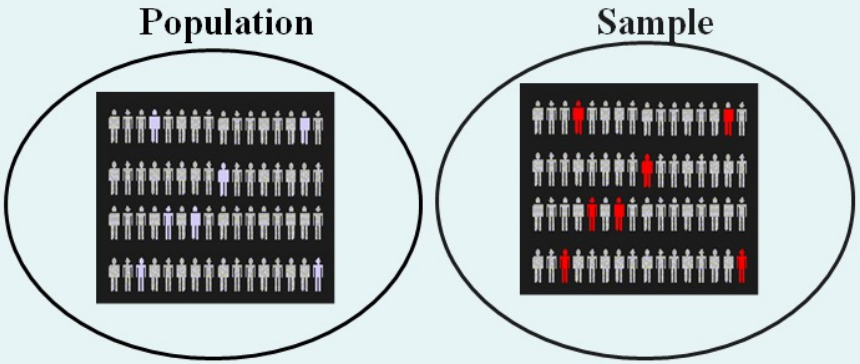
* More than a chance-related difference: **Statistical hypothesis testing**

*Finding associations between variables: **Correlation and regression**

(Allyn and Bacon 2010)

DESCRIPTIVE STATISTICS – it collects, analyses, explains and presents *data.* It describes a present or past situation.

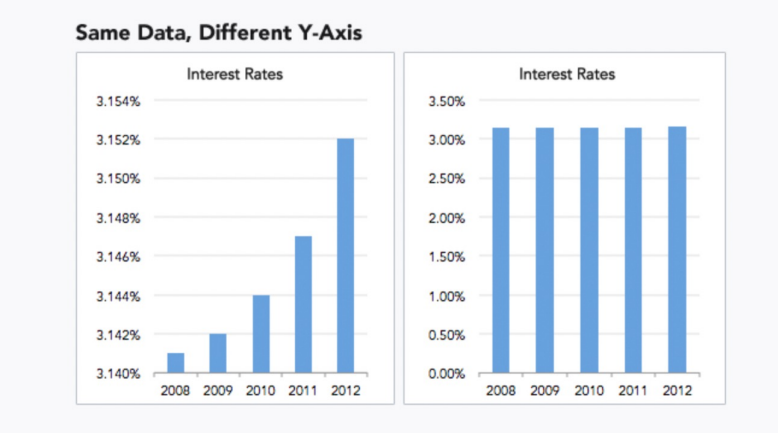
INFERENTIAL STATISTICS – it uses data collected by a small samples to make conclusions / previsions about a bigger group.



University of Essex

Online

Research Methods and Professional Practice – Statistics



Data Cleaning

Whatever ways you choose to collect data, you may find irregularities in the values you collect such as undefined or impossible values. For a categorical variable, an undefined value would be a value that does not represent one of the categories defined for the variable. For a numerical variable, an impossible value would be a value that falls outside a defined range of possible values for the variable. For a numerical variable without a defined range of possible values, you might also find **outliers**, values that seem excessively different from most of the rest of the values. Such values may or may not be errors, but they demand a second review.

Values that are *missing* are another type of irregularity. A **missing value** is a value that was not able to be collected (and therefore not available to analysis). For example, you would record a nonresponse to a survey question as a missing value. You can represent missing values in Minitab by using an asterisk value for a numerical variable or by using a blank value for a categorical variable, and such values will be properly excluded from analysis. The more limited Excel has no special values that represent a missing value. When using Excel, you must find and then exclude missing values manually.

Presenting data for reports: check-list

Goals for effective data presentation:

- Present data to display essential information
- Communicate complex ideas clearly and accurately
- Avoid distortion that might convey the wrong message

→ common mistakes:

- Unequal histogram interval widths
- Compressing or distorting the vertical axis
- Providing no zero point on the vertical axis
- Failing to provide a relative basis in comparing data between groups

University of Essex

Online

Research Methods and Professional Practice – Statistics

Statistics is "the branch of mathematics that transform numbers into useful information for decision makers"

(Berenson et al., 2009)

Has statistics an image problem?
[A:Watch the Ted Talk]

→ Main applications:

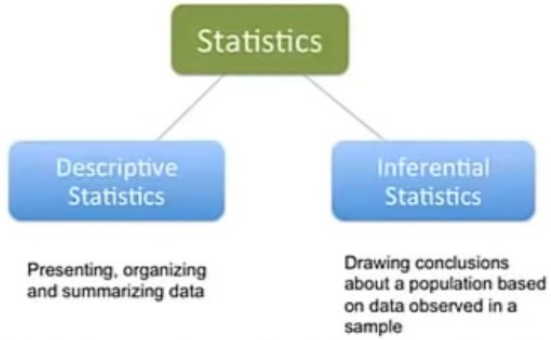
- summarise and draw conclusion business data

- make forecasts

- improving business processes

Ted Talks:
Smith, A. (2016) Why you should love Statistics. Ted Talks.
https://www.ted.com/talks/alan_smith_why_you_should_love_statistics?language=en

Statistics- the collection, organization, & interpretation of data
There are two categories of Statistics



Descriptive Statistics: values that describe the characteristics of a sample or population

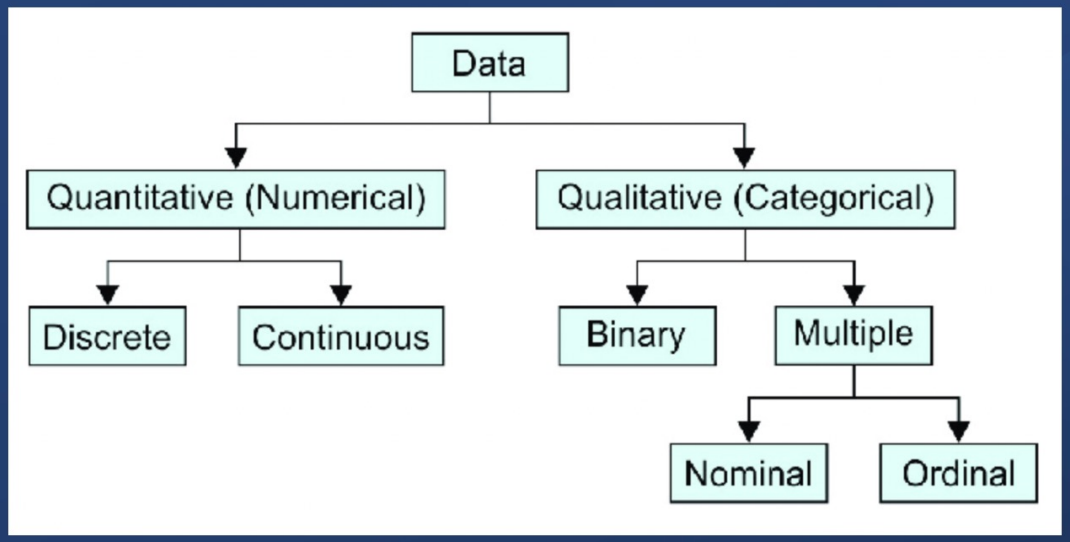
Inferential Statistics: values that infer results of a sample to the population from which the sample is drawn

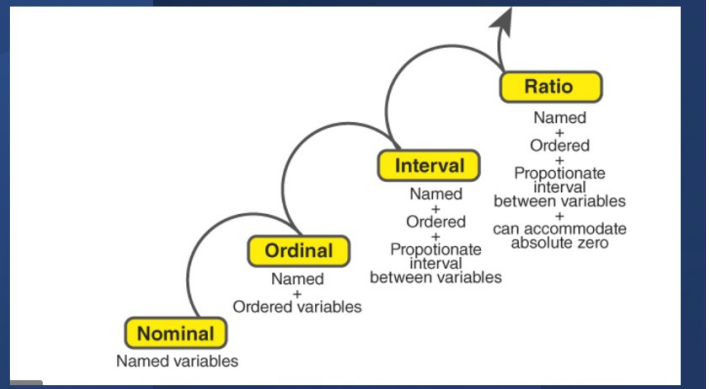
University of Essex

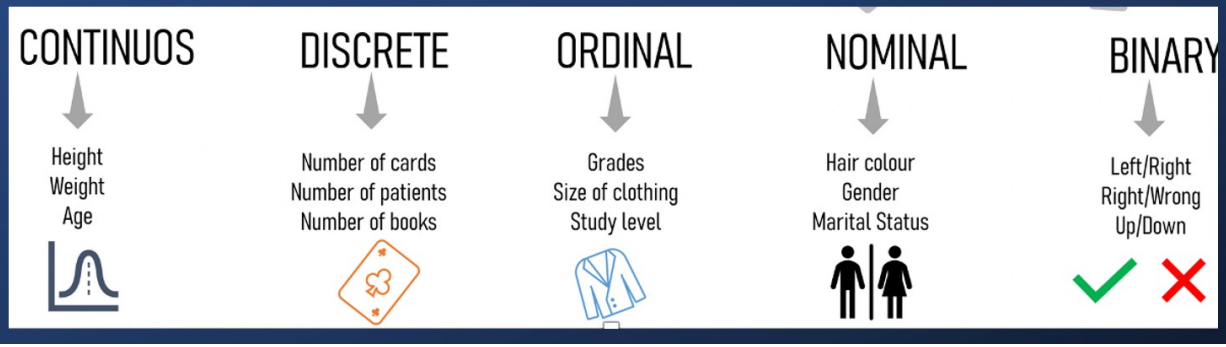
Online

Research Methods and Professional Practice – Statistics

Tricky stuff: **Is Age Discrete or Continuous?**





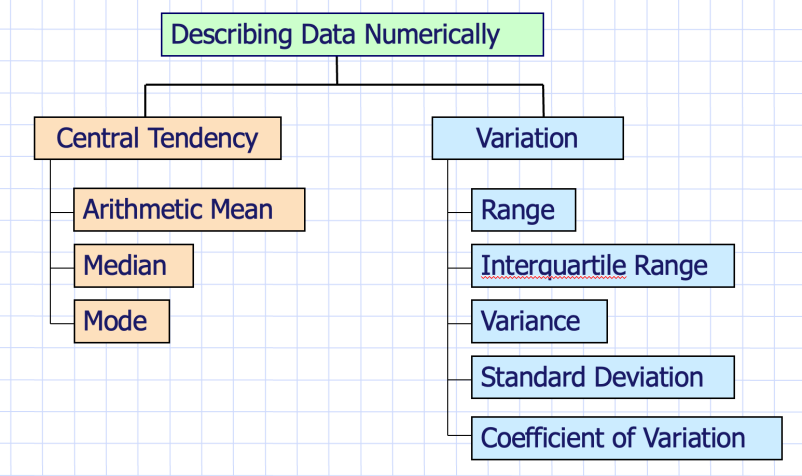


Name	Gender	Age	Weight	Height
James	Male	27	75.1	Short
Barra	Male	32	98.3	Short
Sarah	Female	34	63.5	Medium
Bill	Male	23	87.2	Tall
Peter	Male	27	75.1	Short
Chloe	Female	32	98.3	Short
Ben	Male	34	63.5	Medium
Anna	Female	23	87.2	Tall

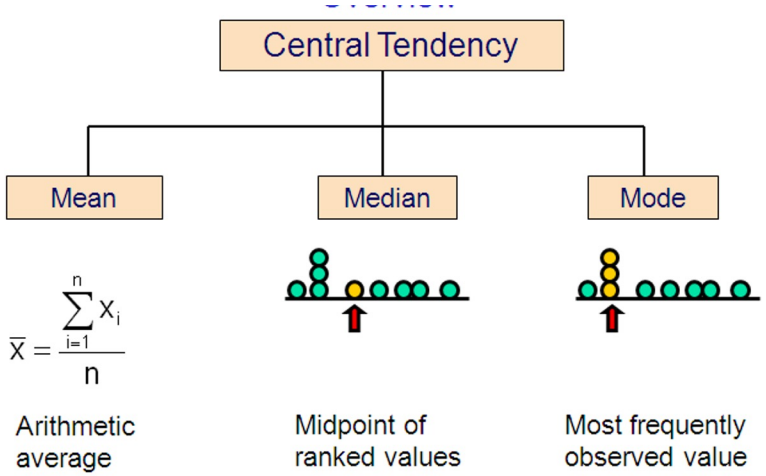
University of Essex

Online

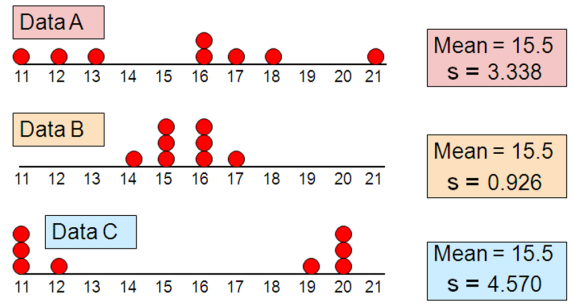
Research Methods and Professional Practice – Statistics

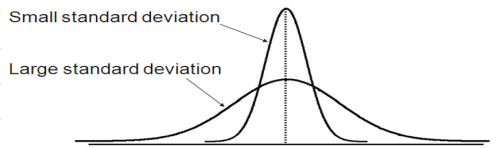


- The more the data are concentrated, the smaller the range, variance, and standard deviation.
- If the values are all the same (no variation), all these measures will be zero.



Why calculate the SD is useful? Which information can give us?

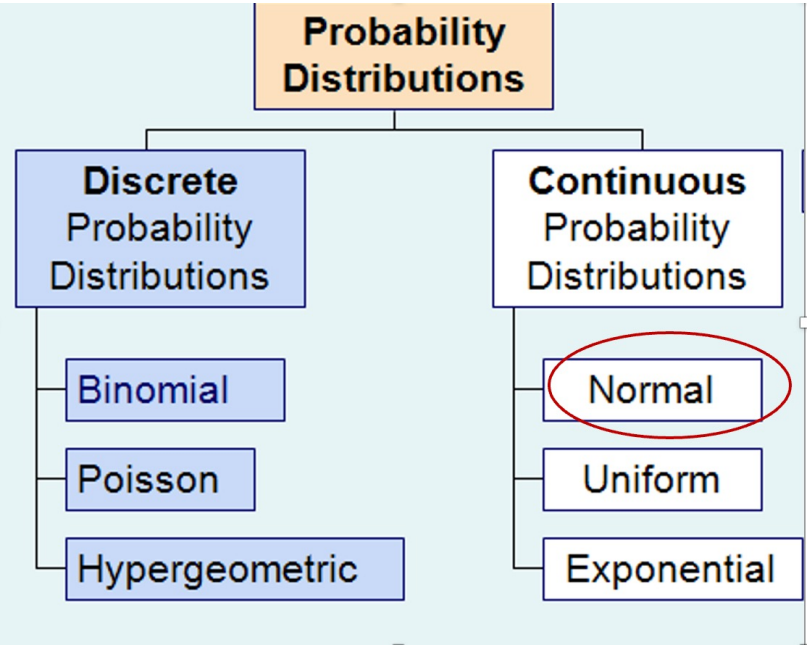


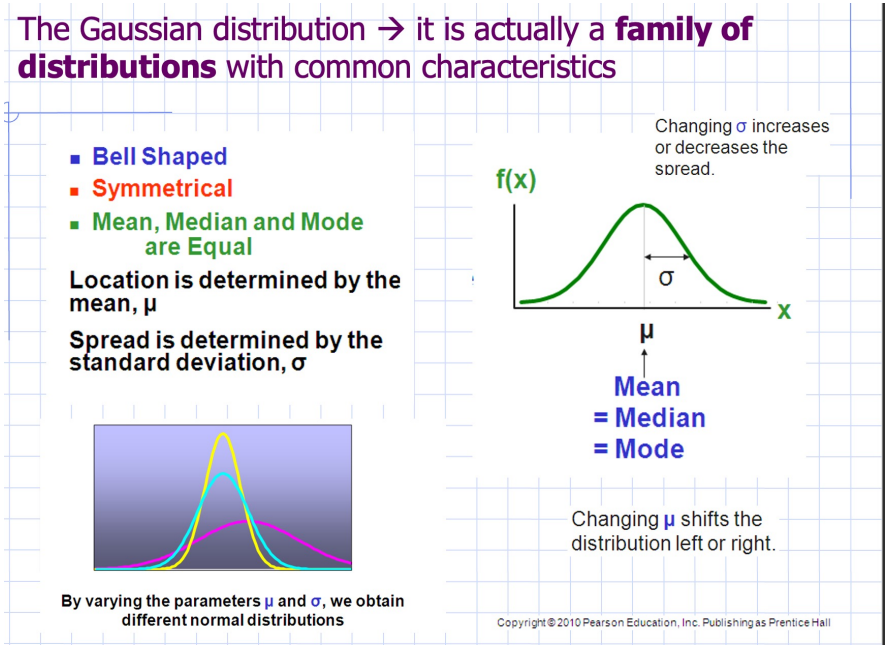


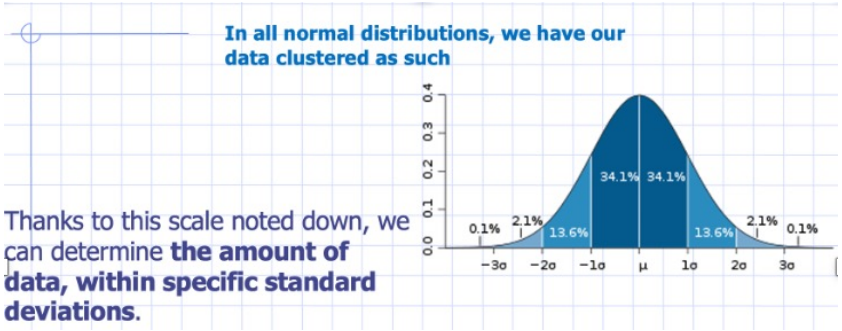
University of Essex

Online

Research Methods and Professional Practice – Statistics

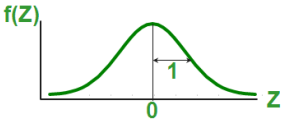






- Any normal distribution (with any mean and variance combination) can be transformed into the standardized normal distribution (Z), with mean 0 and variance 1

$Z \sim N(0,1)$



- Need to transform X units into Z units by subtracting the mean of X and dividing by its standard deviation

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

University of Essex

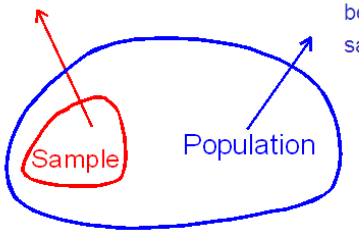
Online

Research Methods and Professional Practice – Statistics

- Making statements about a population by examining sample results

Sample statistics → Population parameters

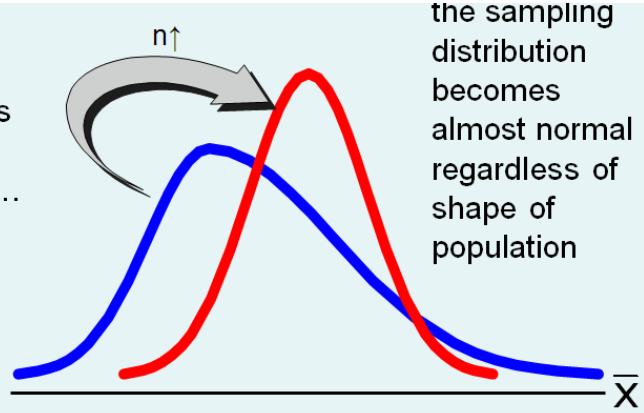
(known) **Inference** (unknown, but can be estimated from sample evidence)



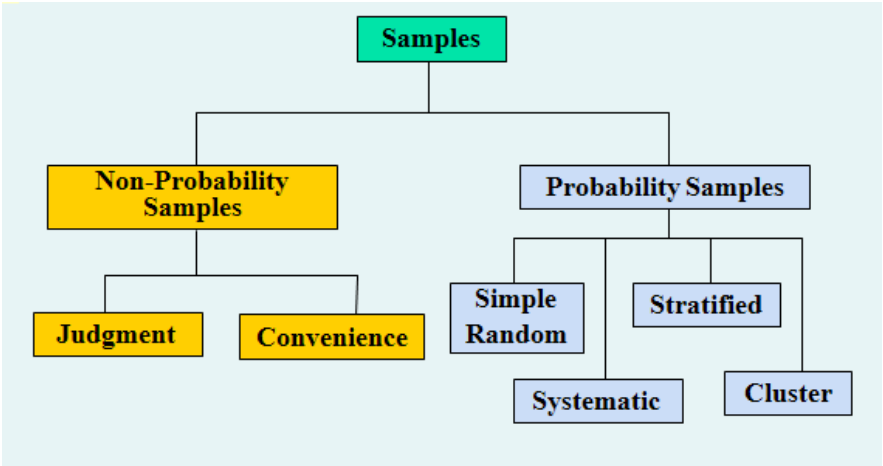
SO, WHY USING SAMPLES?

- **Less time** consuming than a census
- **Less costly** to administer than a census
- It is possible to obtain statistical results of **a sufficiently high precision** based on samples

As the sample size gets large enough…



the sampling distribution becomes almost normal regardless of shape of population



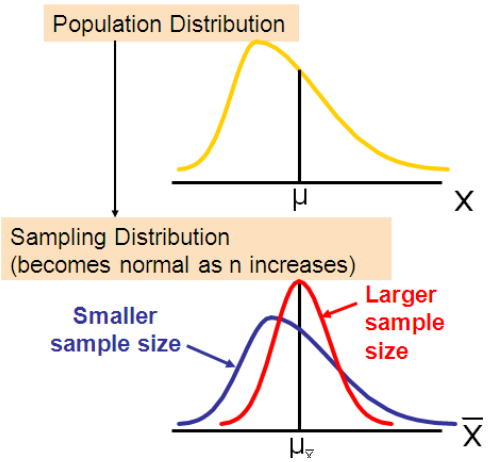
Sampling distribution properties:

Central Tendency

$$\mu_{\bar{x}} = \mu$$

Variation

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$



University of Essex

Online

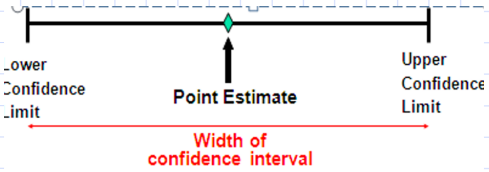
Research Methods and Professional Practice – Statistics

Estimation Process

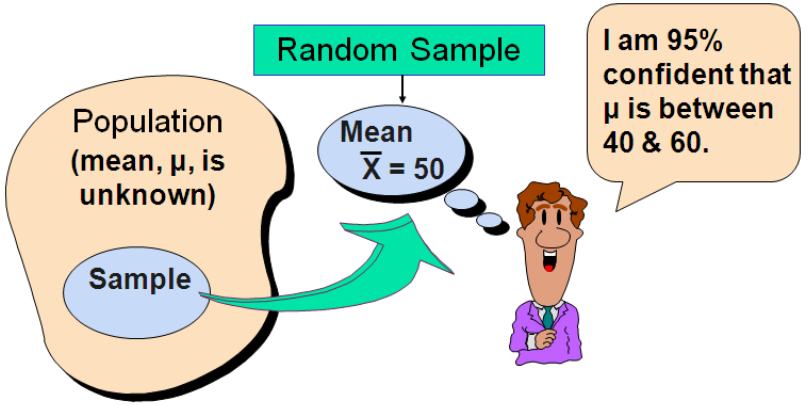
In which way we do estimate? Which kind of estimates we can do?

ESSENTIALLY TWO

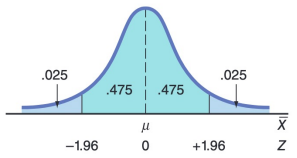
- (1) **point estimate** → a single number
- (2) **confidence interval** → it provides additional information about variability

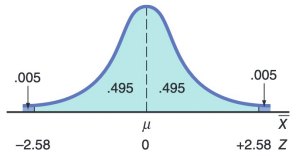


An interval estimate provides more information about a population characteristic than does a point estimate



There is a different critical value for each level of confidence, $1 - \alpha$. A level of confidence of 95% leads to a Z value of 1.96 (see Figure 8.3). 99% confidence corresponds to an α value of 0.01. The Z value is approximately 2.58 because the upper-tail area is 0.005 and the cumulative area less than $Z = 2.58$ is 0.995 (see Figure 8.4).





University of Essex

Online

Research Methods and Professional Practice – Statistics

Statistical decision making is like a courtroom

- You are the **judge**
- Your **data** is on trial
- The *assumption of innocence* is called the **null hypothesis** (often written $\mathbf{H_0}$)
- Only reject the assumption of innocence (i.e. convict your data of guilt) if there is *evidence beyond reasonable doubt* (**sufficient evidence to reject the null hypothesis**)

RES7003: PG Cert in Research Practice

BIRMINGHAM CITY University

- A hypothesis is a claim (assumption) about a population parameter:

  - population mean

    Example: The mean monthly cell phone bill of this city is μ = $42

  - population proportion

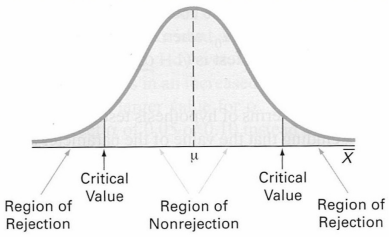
    Example: The proportion of adults in this city with cell phones is p = .68

Hypothesis testing

So the key point here is:

What level of evidence against the null hypothesis should we consider enough to reject this hypothesis, Ho?

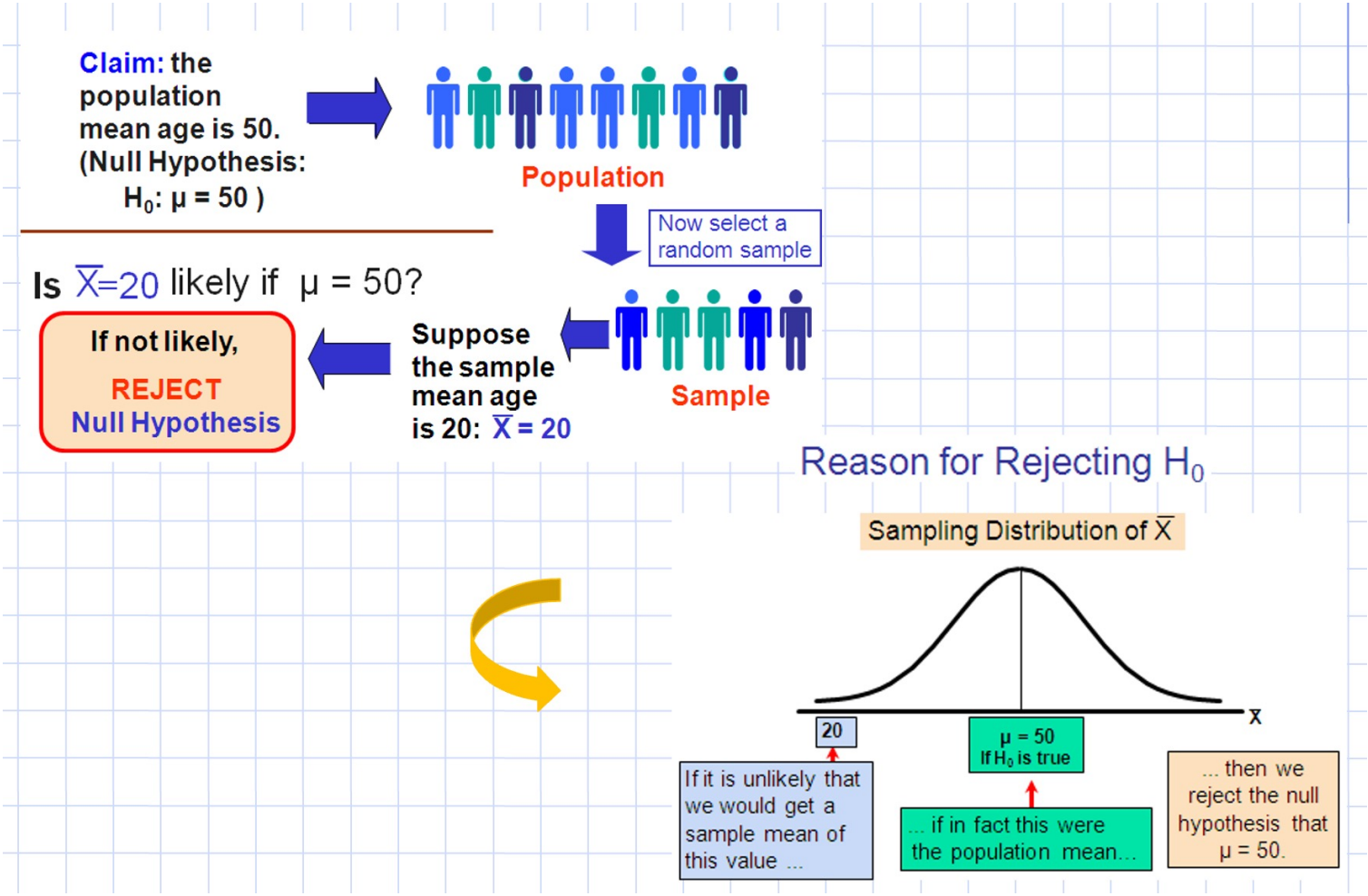
Guess what? Answers lie in our Gaussian....



University of Essex

Online

Research Methods and Professional Practice – Statistics



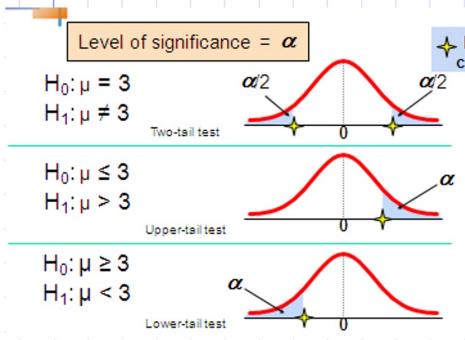
University of Essex

Online

Research Methods and Professional Practice – Statistics

In the hypothesis testing we will use α as a critical value to test if we have to accept or reject it

α **here represents the critical value**



- **Defines the unlikely values of the sample statistic if the null hypothesis is true**
  - Defines rejection region of the sampling distribution
- Is designated by α, (level of significance)
  - Typical values are .01, .05, or .10
- Is selected by the researcher at the beginning
- Provides the critical value(s) of the test

The p-value is also called the **observed level of significance**, and **it is the smallest value of α for which H_0 can be rejected**

- Compare the p-value with α
  - If p-value $<$ α, reject H_0
  - If p-value $\geq$ α, do not reject H_0
- Remember
  - If the p-value is low then H_0 must go

- **Type I Error**
  - Reject a true null hypothesis
  - Considered a serious type of error
  - The probability of a Type I Error is α
    - Called level of significance of the test
    - Set by researcher in advance
- **Type II Error**
  - Failure to reject false null hypothesis
  - The probability of a Type II Error is β

University of Essex

Online

Research Methods and Professional Practice – Statistics

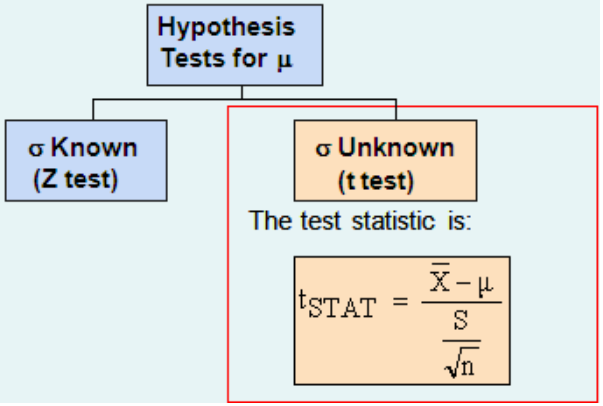
If the population standard deviation is unknown, you instead use the sample standard deviation S.

Because of this change, you use the t distribution instead of the Z distribution to test the null hypothesis about the mean.

When using the t distribution you must assume the population you are sampling from follows a normal distribution.

All other steps, concepts, and conclusions are the same.

Convert sample statistic ($\bar{X}$) to a t_{STAT} test statistic



Student's t Distribution

- The t is a family of distributions
- The $t_{\alpha/2}$ value depends on degrees of freedom (d.f.)
  - Number of observations that are free to vary after sample mean has been calculated

d.f. = n - 1

Idea: Number of observations that are free to vary after sample mean has been calculated

Example: Suppose the mean of 3 numbers is 8.0

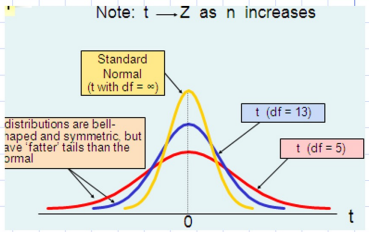
Let X_1 = 7
Let X_2 = 8
What is X_3?

→ If the mean of these three values is 8.0, then X_3 must be 9 (i.e., X_3 is not free to vary)

Here, n = 3, so degrees of freedom = n – 1 = 3 – 1 = 2

(2 values can be any numbers, but the third is not free to vary for a given mean)

Note: t → Z as n increases



Upper Tail Area			
df	.25	.10	.05
1	1.000	3.078	6.314
2	0.817	1.886	2.920
3	0.765	1.638	2.353

The body of the table contains t values, not probabilities

Let: n = 3
df = n - 1 = 2
α = 0.10
α/2 = 0.05

α

- The Excel output below does this:

t Test for the Hypothesis of the Mean

Data		
Null Hypothesis μ=	$	168.00
Level of Significance		0.05
Sample Size		25
Sample Mean	$	172.50
Sample Standard Deviation	$	15.40

Intermediate Calculations			
Standard Error of the Mean	$	3.08	=B8/SQRT(B6)
Degrees of Freedom		24	=B6-1
t test statistic		**1.46**	=(B7-B4)/B11

Two-Tail Test		
Lower Critical Value	-2.0639	=-TINV(B5,B12)
Upper Critical Value	2.0639	=TINV(B5,B12)
p-value	0.157	=TDIST(ABS(B13),B12,2)
Do Not Reject Null Hypothesis		=IF(B18<B5, "Reject null hypothesis", "Do not reject null hypothesis")

p-value > α
So do not reject H_0

Research Methods and Professional Practice – Statistics

In hypothesis testing, you reject the null hypothesis when the sample evidence suggests that it is far more likely that the alternative hypothesis is true. However, failure to reject the null hypothesis is not proof that it is true. You can never prove that the null hypothesis is correct because the decision is based only on the sample information, not on the entire population. Therefore, if you fail to reject the null hypothesis, you can only conclude that there is insufficient evidence to warrant its rejection. The following key points summarize the null and alternative hypotheses:

- The null hypothesis, H_0, represents the current belief in a situation.
- The alternative hypothesis, H_1, is the opposite of the null hypothesis and represents a research claim or specific inference you would like to prove.
- If you reject the null hypothesis, you have statistical proof that the alternative hypothesis is correct.
- If you do not reject the null hypothesis, you have failed to prove the alternative hypothesis. The failure to prove the alternative hypothesis, however, does not mean that you have proven the null hypothesis.
- The null hypothesis, H_0, always refers to a specified value of the population parameter (such as μ), not a sample statistic (such as $\overline{X}$).
- The statement of the null hypothesis always contains an equal sign regarding the specified value of the population parameter (e.g., $H_0 : \mu = 368$ grams).
- The statement of the alternative hypothesis never contains an equal sign regarding the specified value of the population parameter (e.g., $H_1 : \mu \neq 368$ grams).

University of Essex

Online

Research Methods and Professional Practice – Statistics

Potential Pitfalls and Ethical Considerations

- Use randomly collected data to reduce selection biases
- Do not use human subjects without informed consent
- Choose the level of significance, α, and the type of test (one-tail or two-tail) before data collection
- Do not employ "data snooping" to choose between one-tail and two-tail test, or to determine the level of significance
- Do not practice "data cleansing" to hide observations that do not support a stated hypothesis
- Report all pertinent findings including both statistical significance and practical importance

eurostat
Compact guides

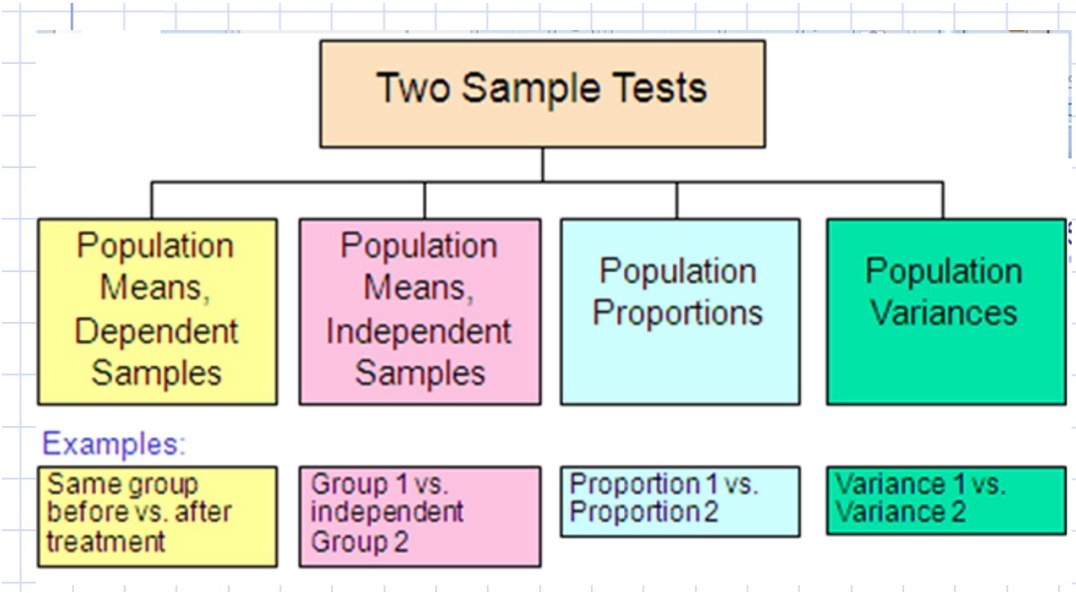
Find it at: https://ec.europa.eu/eurostat/documents/3217494/5719241/KS-31-09-254-EN.PDF.pdf/e82af841-615f-47d4-8c6d-647705de97a7?t=1414775067000

University of Essex

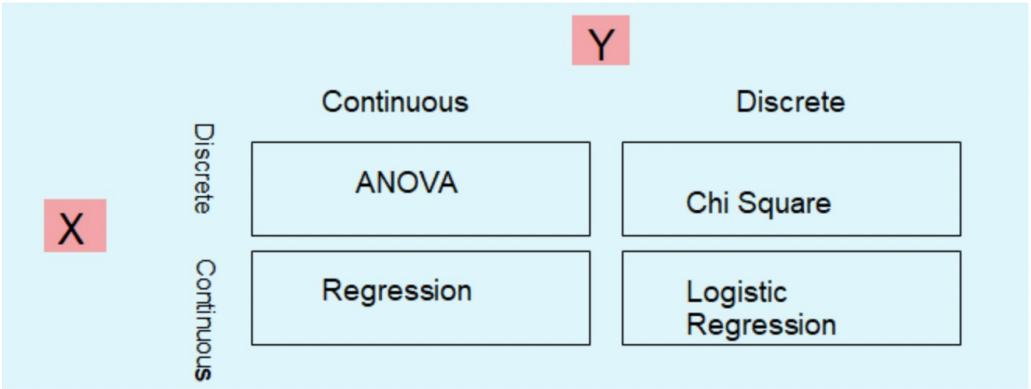
Online

There is much, much more that you can do with statistics…

Research Methods and Professional Practice – Statistics



ALSO:
Non-paramentric tests
as:
χ^2 test,
Wilcoxon rank test, and a lot of others…



University of Essex

Online

Research Methods and Professional Practice - Recap U 8-12

8. Inferential Statistics	8	Hypothesis Testing worksheet, **submit a brief outline of your Research Proposal**, Exercises
9. Analysing Qualitative Data	9	Discussion forum, Charts Worksheet and Analysis,
10. Research Writing	10	Submit your statistical worksheets **Research Proposal Presentation**
11. Professional Development – Your e-Portfolio.	11	e-Portfolio Preparation,
12. Project Management and Managing Risk	12	Self Test Quiz, **End of Module Assignment** - Individual e-Portfolio

University of Essex

Online

Assignment Brief

The research proposal presentation fulfils the learning outcomes where you will evaluate critically existing literature, research design and methodology for your chosen topic, including data analysis processes and so produce and evaluate critically a research proposal for your chosen topic. This can be based on the literature review topic you chose in Unit 1 or the topic of your capstone project (MSc students).

Research Proposal Presentation

- Content:
  - Project Title.
  - Significance/Contribution to the discipline/Research Problem.
  - Research Question.
  - Aims and Objectives.
  - Key literature related to the project.
  - Methodology/Development strategy/Research Design.
  - Ethical considerations and risk assessment (as part of your ethical approval application).
  - Description of artefact(s) that will be created (if applicable).
  - Timeline of proposed activities.

- For this assignment, you are required to record a **15-minute presentation**, along with a transcript of the audio file. (**See the Instructions**)

University of Essex

Online

Questions?

There is no word count for the slide presentation, and there is no fixed word count for the oral presentation/transcript. However, note that an average speech rate for a comfortable and clearly orally paced presentation is about 100-150 words per minute. Given that your presentation should be about 15 minutes, you can work against a minimum of 1500 to a maximum of 2250 words in your oral presentation/transcript.