

1. Introduction to Bayesian reasoning

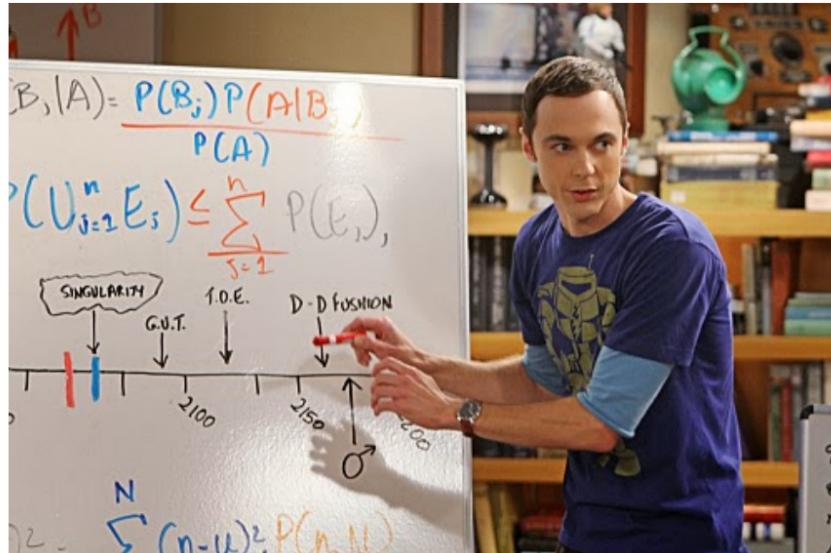
Gianluca Baio

Department of Statistical Science | University College London

- ✉ g.baio@ucl.ac.uk
- 🌐 <https://gianluca.statistica.it/>
- 🌐 <https://egon.stats.ucl.ac.uk/research/statistics-health-economics/>
- 🌐 <https://github.com/giabaio>
- 🌐 <https://github.com/StatisticsHealthEconomics>
- 🐦 [@gianlubaio](https://twitter.com/gianlubaio)

STAT0019 - Bayesian Methods in Health Economics, UCL

Objective of this course



- Introduction to **Bayesian analysis**
 - MCMC methods
 - Using R and BUGS
- Apply Bayesian analysis to **health economic evaluations**
 - Cost-effectiveness analysis
 - Probabilistic sensitivity analysis
 - Advanced modelling
- Emphasis on **practical examples**
 - BUGS analysis
 - R/BUGS and BCEA
 - Problem-specific vs standardised analysis

Relevant resources

The course [website](#) contains all the relevant information

- [Reading list](#)
- [Course description & assessment](#)
- [Full timetable](#)
- [Full syllabus](#)
- [Useful tips of the computer specification \(for the practicals\)](#)

All the lecture slides are also available from the main page (see top menu under "Slides")

The material for the computer practicals is also available from the main page (see top menu under "Practicals")

The relevant slides and practical material will be made available **before** the scheduled lecture.

Annotated solutions to the practicals will also be made available **after** the sessions.

Relevant resources

The course **website** contains all the relevant information

- **Reading list**
- **Course description & assessment**
- **Full timetable**
- **Full syllabus**
- **Useful tips of the computer specification (for the practicals)**

All the lecture slides are also available from the main page (see top menu under "Slides")

The material for the computer practicals is also available from the main page (see top menu under "Practicals")

The relevant slides and practical material will be made available **before** the scheduled lecture.

Annotated solutions to the practicals will also be made available **after** the sessions.

- Some  resources:
 - **This** is a very comprehensive introduction.
 - **This** is also a very good introduction, particularly around many of the more modern features of R (e.g. the **tidyverse** package/approach).
- NICE Decision Support Unit website: <http://nicedsu.org.uk/>
- Moodle page (UCL-registered): <https://moodle.ucl.ac.uk/course/view.php?id=8596>

The BUGS Book

A Practical Introduction to Bayesian Analysis

Authors/Affiliations

David Lunn, MRC Biostatistics Unit, Cambridge, UK
 Chris Jackson, MRC Biostatistics Unit, Cambridge, UK, MRC Biostatistics Unit, Cambridge, UK
 Nicky Best, Imperial College, London, UK, Imperial College London, London, UK
 Andrew Thomas, MRC Biostatistics Unit, Cambridge, UK, MRC Biostatistics Unit, Cambridge, UK
 David Spiegelhalter, University of Cambridge, UK, University of Cambridge, UK

In recent years, Bayesian methods have become the most widely used statistical methods for data analysis and modelling. The BUGS software has become the most popular software for Bayesian analysis worldwide. Authored by the team that originally developed this software, *The BUGS Book* provides a practical introduction to this program and its use. The text presents complete coverage of all the functionalities of BUGS, including prediction, missing data, model criticism, and prior sensitivity. It also features a large number of worked examples, covering a wide range of applications from various disciplines, and exercises, solutions, code and data on a supplementary website.

Key Features

- Provides an accessible introduction to Bayesian analysis using the BUGS software
- Covers all the functionalities of BUGS, including prediction, missing data, model criticism, and prior sensitivity
- Features a large number of worked examples and applications from a wide range of disciplines
- Includes detailed exercises and solutions on the supporting website
- Authored by the team that developed the BUGS software.

Selected Contents

Introduction: Probability and Parameters. Monte Carlo Simulations using BUGS. Introduction to Bayesian Inference. Introduction to Markov Chain Monte Carlo Methods. Prior Distributions. Regression Models. Categorical Data. Model Checking and Comparison. Issues in Modeling. Hierarchical Models. Specialized Models. Different Implementations of BUGS. Appendices. Bibliography. Index.

SAVE 20% when you order online and enter Promo Code EZL20
FREE standard shipping when you order online.

Catalog no. C8490
 October 2012, 399 pp.
 ISBN: 978-1-58488-849-9
 \$52.95 / £25.99

www.crcpress.com

CRC Press
 Taylor & Francis Group

Bayesian Methods in Health Economics

Author/Affiliation

Gianluca Baio, Department of Statistical Science, University College London, UK

Health economics is concerned with the study of the cost-effectiveness of health care interventions. This book provides an overview of Bayesian methods for the analysis of health economic data. After an introduction to the basic economic concepts and methods of evaluation, it presents Bayesian statistics using accessible mathematics. The next chapters describe the theory and practice of cost-effectiveness analysis from a statistical viewpoint, and Bayesian computation, notably MCMC. The final chapter presents three detailed case studies covering cost-effectiveness analyses using individual data from clinical trials, evidence synthesis and hierarchical models and Markov models. The text uses WinBUGS and JAGS with datasets and code available online.

Key Features

- Provides an overview of Bayesian methods for cost-effectiveness analysis, and includes all necessary background on economics and Bayesian statistics
- Presents three detailed case studies of the cost-effectiveness analysis of health care interventions
- Includes several worked examples to guide through the process of health economic evaluation
- Contains extensive coverage of the practice of making Bayesian analysis integrating software such as JAGS and R, specifically for the application of health economic analysis
- Systematically describes the methodological issues related to the application of Bayesian inference and decision process in health economics
- Designed as a reference for students and practitioners working in the field of health economic evaluations and medical statistics

The book is linked to [code](#) with which to replicate the examples, and an associated R package ([BCEA](#)) can be used in real applications to produce systematic health economic evaluations of Bayesian models.

Selected Contents

Introduction to Health Economic Evaluation. Introduction to Bayesian Inference. Statistical Cost-Effectiveness Analysis. Bayesian Analysis in Practice. Health Economic Evaluation in Practice.

SAVE 20% when you order online and enter Promo Code EZL20
FREE standard shipping when you order online.

Catalog no. K14236
 November 2012, 243 pp.
 ISBN: 978-1-4398-9555-9
 \$93.95 / £59.99

www.crcpress.com

CRC Press
 Taylor & Francis Group

Disclaimer...



A screenshot of a Twitter post from user @ManuelaJoore. The post contains a quote from Gianluca Baio about Bayesian statisticians. Below the quote is a reply from Gianluca Baio expressing his readiness for a session on open source models & methods. The post includes standard Twitter interaction metrics like likes, replies, and a link to read more replies.

Manuela Joore @ManuelaJoore

Best opening sentence [#ISPOREurope](#) from Gianluca Baio: “statisticians should rule the world and Bayesian statisticians should rule all statisticians”

Gianluca Baio @gianlubaio

Ready for our session on open source models & methods!

4:52 PM · Nov 4, 2019

16 likes, 1 reply, 1 link to Tweet

[Read 2 replies](#)

...Just so you know what you're about to get into... 😊

- Sampling variability
 - Probability calculus vs Statistics
- Deductive inference
 - "Standard" statistical methods
 - Confidence intervals & testing
- Inductive inference
 - Bayesian reasoning
 - Basic ideas
 - Forming "priors"

References

❑ *The BUGS Book*, chapters 1, 2, 5

 Library

 Book website

❑ *Bayesian Methods in Health Economics*, chapters 2, 4

 Library

 Book website (CRC)

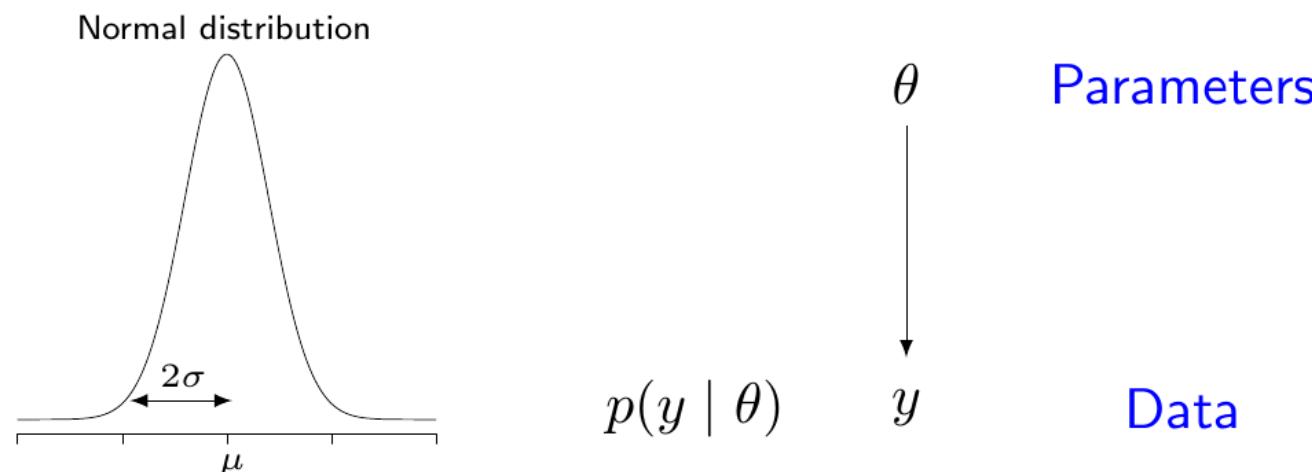
Book website

Code

🌐 <http://www.statistica.it/gianluca/teaching/intro-stats/>

What is statistics all about?

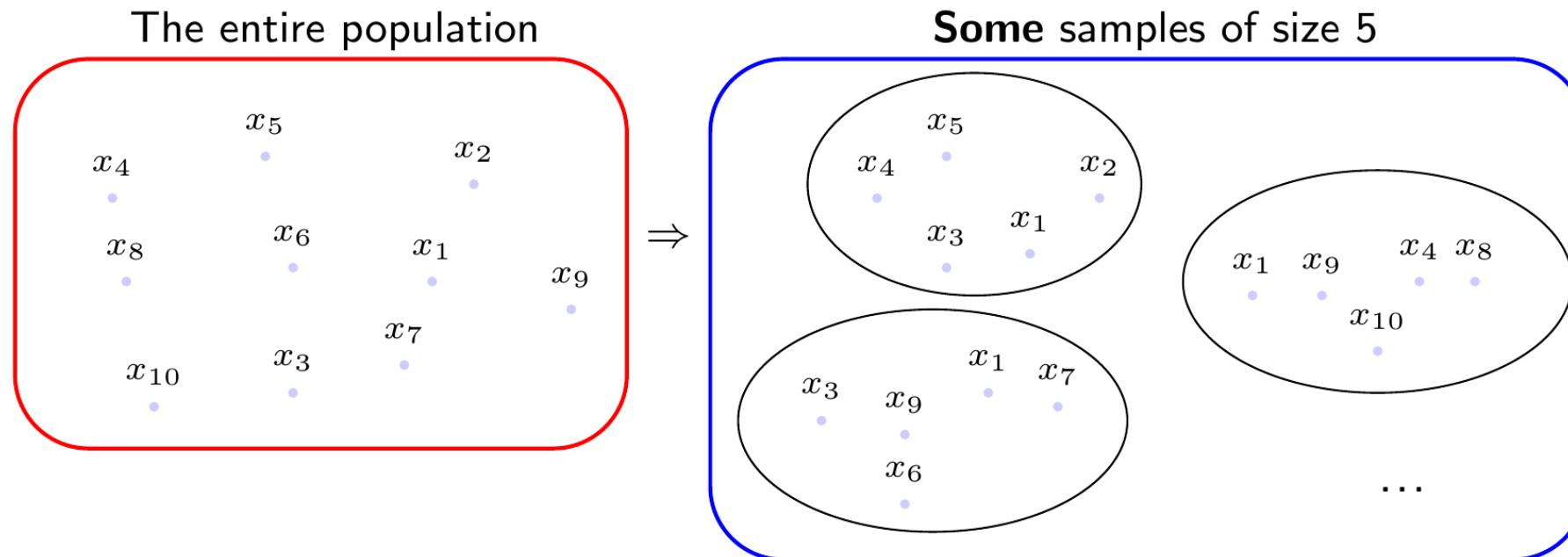
- Typically, we observe some data and we want to use them to learn about some unobservable feature of the general population in which we are interested
- To do this, we use statistical models to describe the probabilistic mechanism by which (we assume!) that the data have arisen



NB: Roman letters (y or x) typically indicate **observable data**, while Greek letters ($\theta, \mu, \sigma, \dots$) indicate **population parameters**

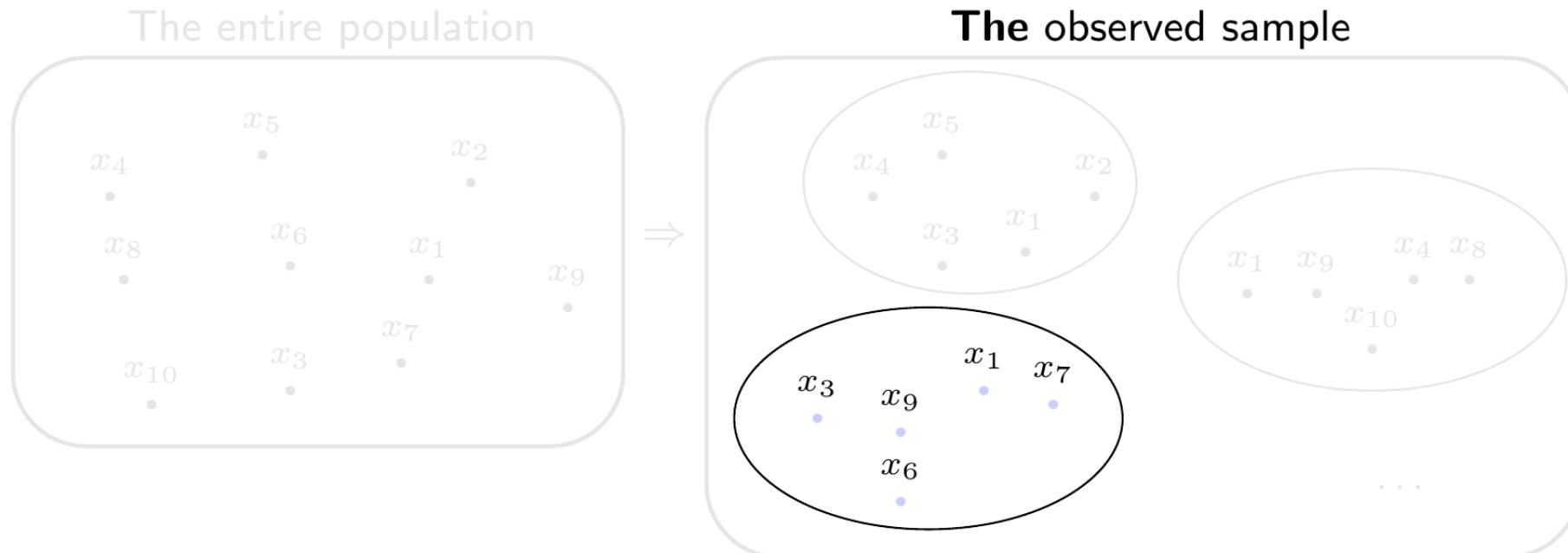
Sampling variability

Probability calculus



- Population size $N = 10$
- “True” population Mean μ
- “True” Standard deviation σ
- Sample size $n = 5$
- Sample Mean \bar{x}
- Sample Standard deviation s_x

Sampling variability



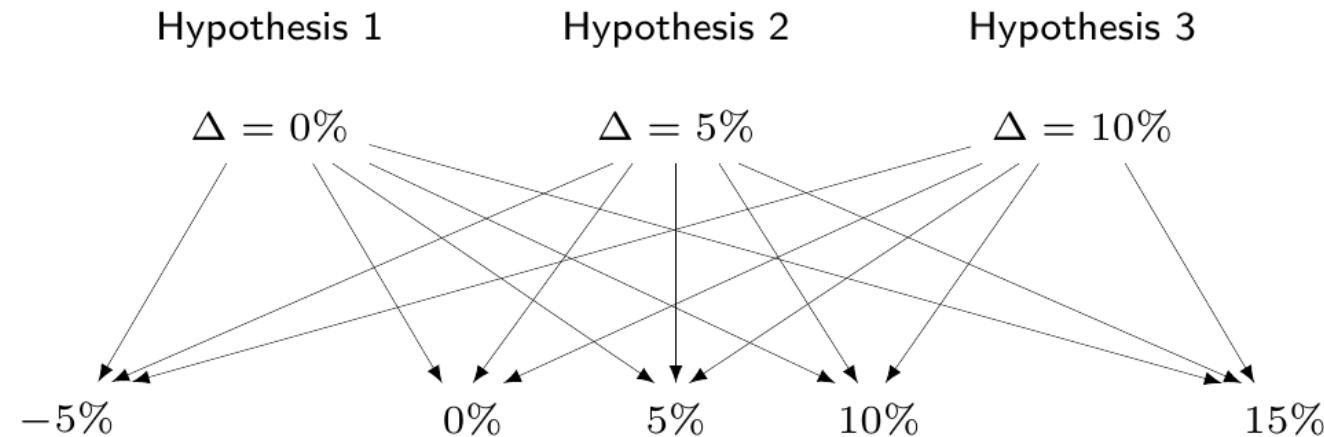
- Population size $N = 10$
 - “True” population Mean μ
 - “True” Standard deviation σ
- ⇐
- Sample size $n = 5$
 - Sample Mean \bar{x}
 - Sample Standard deviation s_x

In reality we observe **only one** such sample (out of the many possible – in fact there are 252 different ways of picking **at random** 5 units out of a population of size 10!) and we want to use the information contained in **that sample** to **infer** about the **population parameters** (e.g. the true mean and standard deviation)

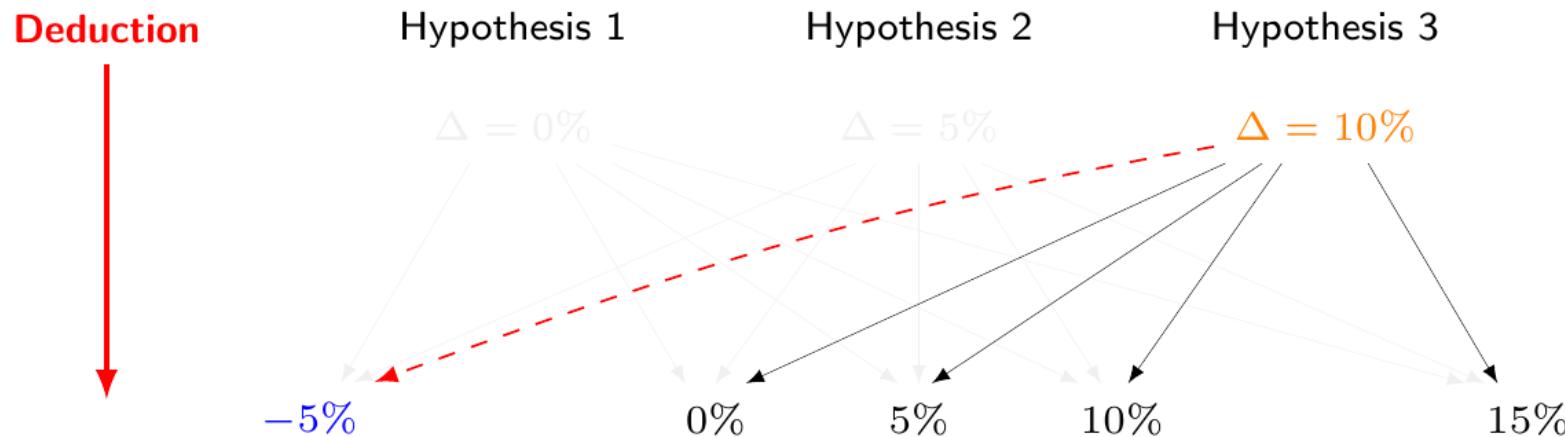
The Sherlock conundrum



Deductive vs inductive inference

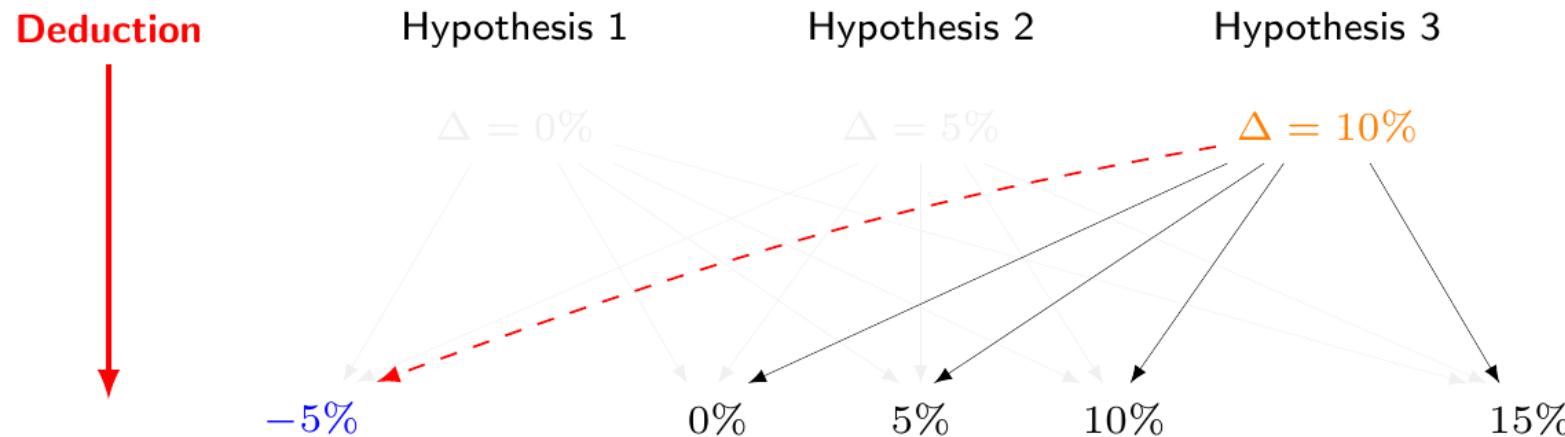


Deductive vs inductive inference



- Standard (frequentist) procedures fix the working hypotheses and, by **deduction**, make inference on the observed data:
 - If my hypothesis is true, what is the probability of randomly selecting the data that I actually observed? If small, then *deduce* weak support of the evidence to the hypothesis

Deductive vs inductive inference



- Standard (frequentist) procedures fix the working hypotheses and, by deduction, make inference on the observed data:
 - If my hypothesis is true, what is the probability of randomly selecting the data that I actually observed? If small, then *deduce* weak support of the evidence to the hypothesis
 - Assess $\Pr(\text{Observed data} \mid \text{Hypothesis})$
 - Directly relevant for standard frequentist summaries, eg p-values, Confidence Intervals, etc
 - NB: Comparison with data that could have been observed, but haven't!

Adapted from  Goodman (1999)

⚠ See <http://www.statistica.it/gianluca/teaching/intro-stats/interval-estimation.html>!

Drug to cure headaches - "true" probability of success: $\pi = 40/73 \approx 0.55$

⚠ See <http://www.statistica.it/gianluca/teaching/intro-stats/interval-estimation.html>!

Drug to cure headaches - "true" probability of success: $\pi = 40/73 \approx 0.55$

- You get to see data for, say, $n = 10$ individuals, under the "true" **data generating process** (DGP):
 $\mathbf{y} = (y_1, \dots, y_{10}) = (0, 0, 1, 1, 0, 1, 0, 1, 0, 1)$
- Can make estimates to infer from sample to population

– Sample mean: $\bar{y} = \hat{\pi} = \sum_{i=1}^n \frac{y_i}{n} = \frac{5}{10} = 0.5$ Standard error: $\text{se}(\hat{\pi}) = \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}} = 0.16$

⚠ See <http://www.statistica.it/gianluca/teaching/intro-stats/interval-estimation.html>!

Drug to cure headaches - "true" probability of success: $\pi = 40/73 \approx 0.55$

- You get to see data for, say, $n = 10$ individuals, under the "true" **data generating process** (DGP):
 $\mathbf{y} = (y_1, \dots, y_{10}) = (0, 0, 1, 1, 0, 1, 0, 1, 0, 1)$
- Can make estimates to infer from sample to population

– Sample mean: $\bar{y} = \hat{\pi} = \sum_{i=1}^n \frac{y_i}{n} = \frac{5}{10} = 0.5$ Standard error: $\text{se}(\hat{\pi}) = \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}} = 0.16$

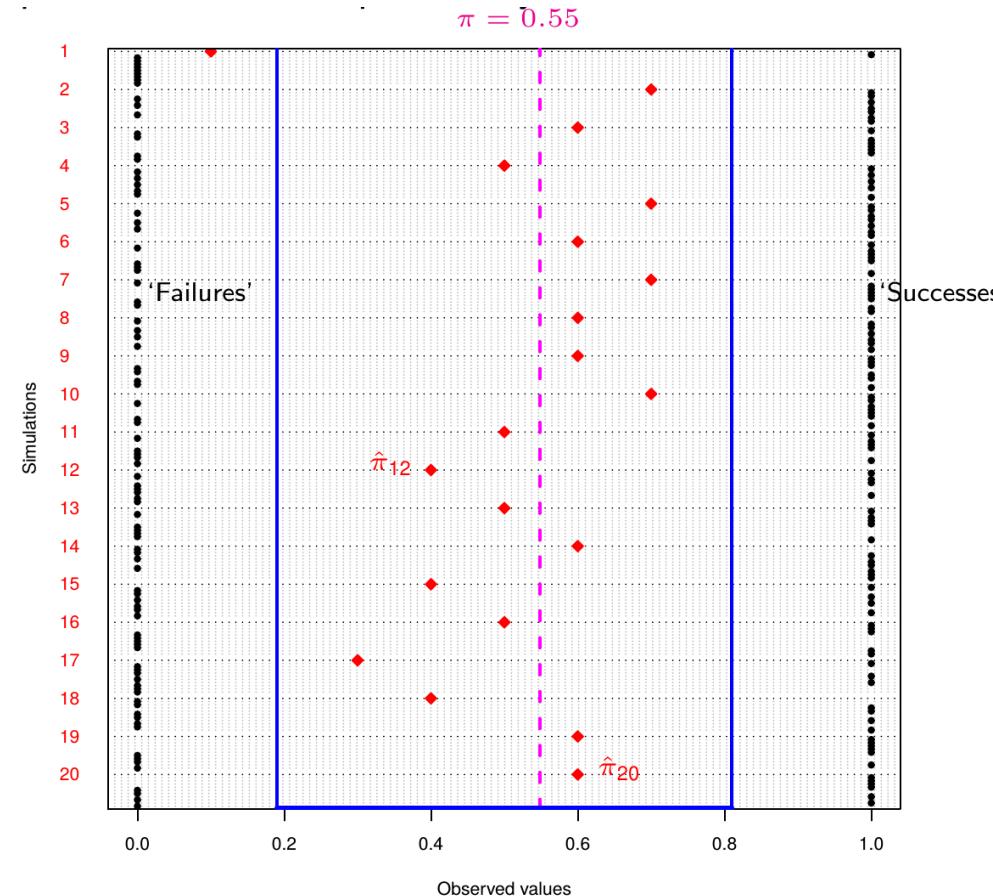
- Can compute the interval estimate (using some approximation/theoretical results...)

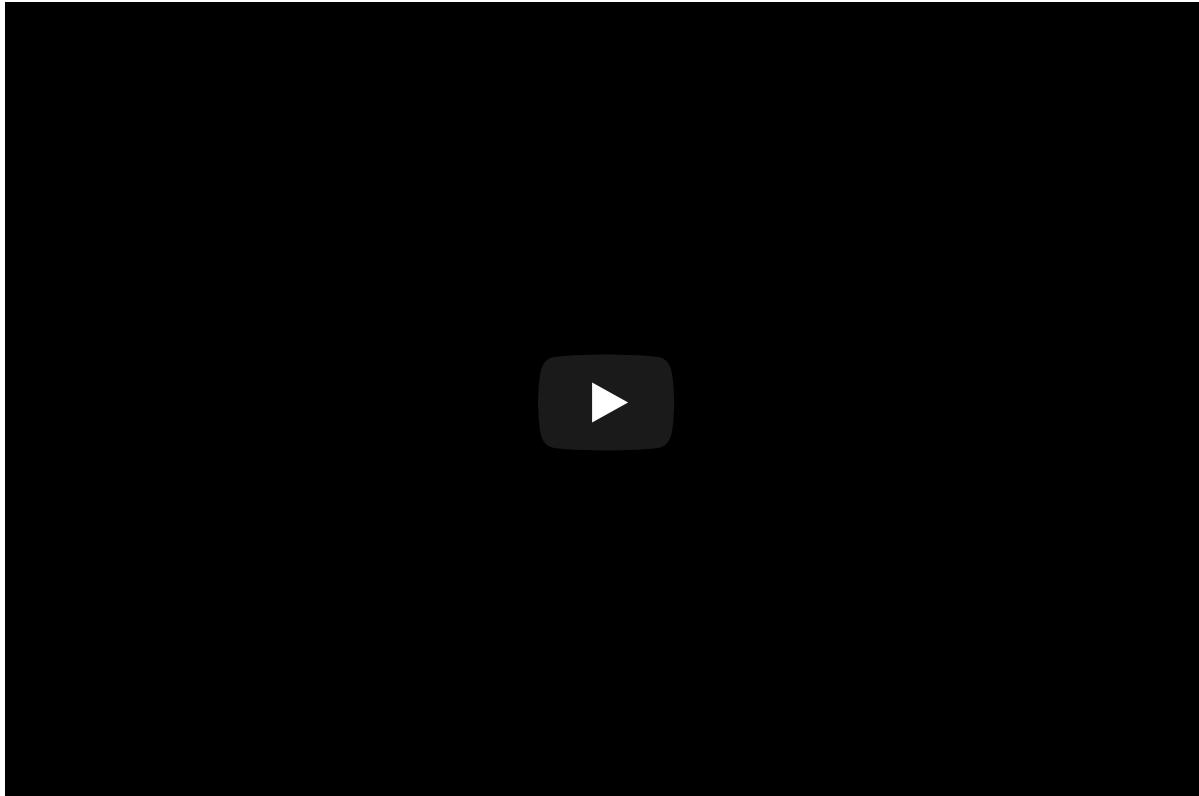
$$95\% \text{ CI} \approx [\hat{\pi} - 2\text{se}(\hat{\pi}); \hat{\pi} + 2\text{se}(\hat{\pi})] = [0.5 - 0.32; 0.5 + 0.32] = [0.19; 0.81]$$

- Assuming the observed sample is representative of the DGP and using the sample estimates, if we were able to replicate the experiment over and over again under the same conditions, 95% of the times, the estimate for the "true" probability of success will be included in the interval $[0.19; 0.81]$
- That is how you interpret a 95% Confidence Interval!

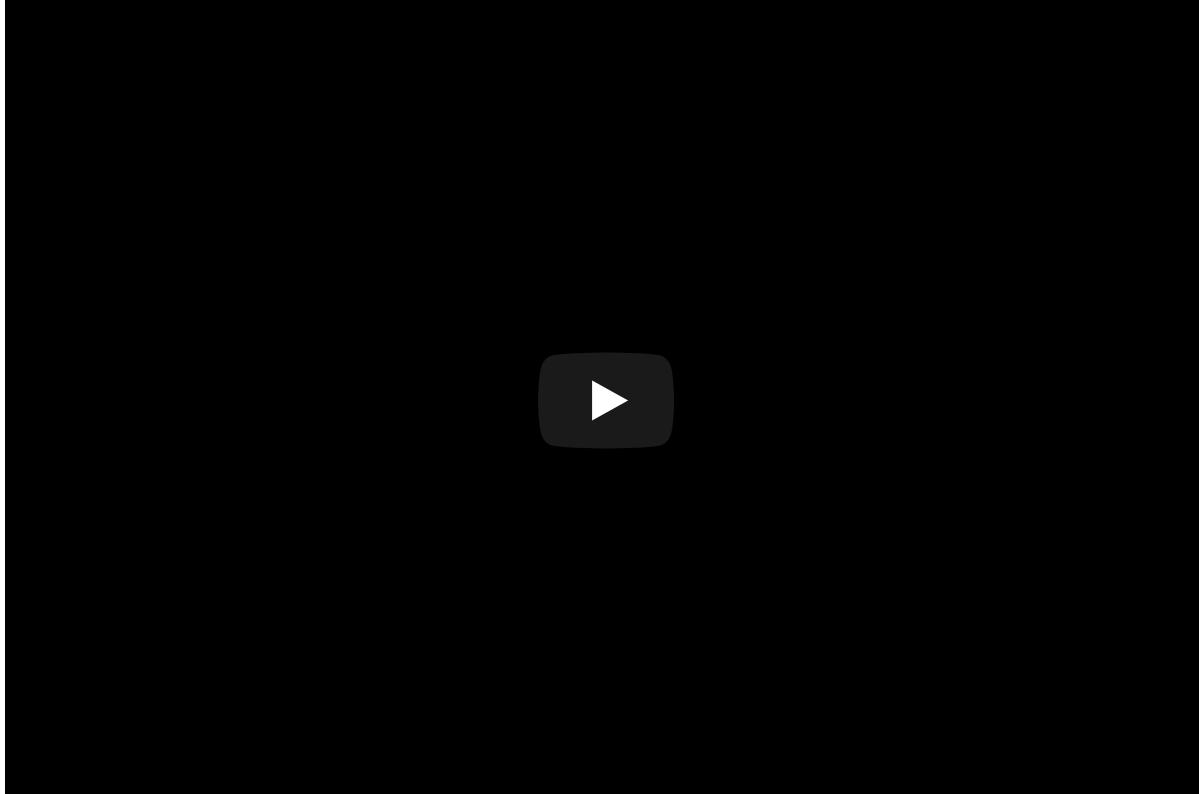
Confidence intervals

- Simulate n_{sim} (e.g. = 20) studies sampling data from a DGP assuming a "true" $\pi = 0.5$ (although in fact, $\pi = 0.55!$) and $n = 10$
- For each, estimate the probability $\hat{\pi}$





Sample size calculations



Designing a study

- Designing a study is just as important as analysing it
 - If we don't have "enough" information in the data, we won't be able to detect an underlying signal
- Related to "hypothesis" testing

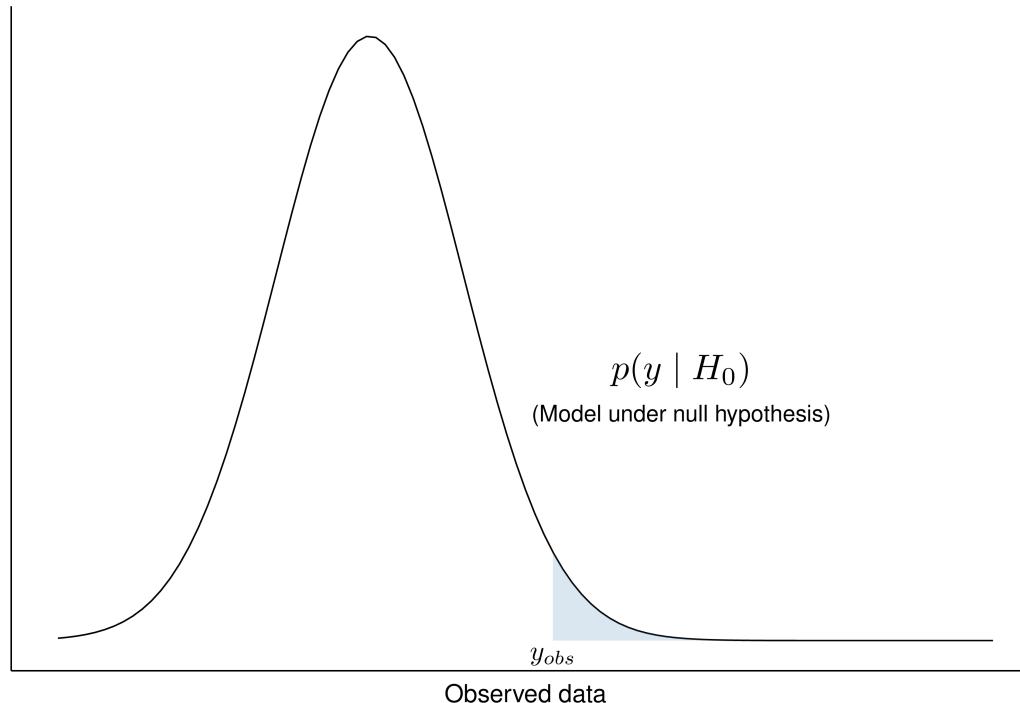
		“Null” hypothesis H_0	
		<i>True</i>	<i>False</i>
Decision on “Null”	<i>Reject</i>	Type I error α (False positive)	Correct inference (True positive)
	<i>Fail to reject</i>	Correct inference (True negative)	Type II error β (False negative)

- 1 Set the Type I error to some low level (typically: $\alpha = 0.05$)
- 2 Set the Type II error to some set level (typically: $\beta = 0.10$ or $\beta = 0.20$)
- 3 Define the "clinically relevant outcome" (eg difference in treatment effects), δ
- 4 Set an estimate of variability in the underlying population
- 5 Use assumptions about sampling variability and determine minimum number of observations to be able to detect δ

Originally devised to guide quality control of processes

Analysing a study

- Interpretation: Under the null hypothesis (ie IF it is true), what is the probability of observing something as extreme or even more extreme as the observed data?



- If $p < 0.01$ then **strong evidence against the null hypothesis**
- If $0.01 < p < 0.05$ then **fairly strong evidence against the null hypothesis**
- If $p > 0.05$ then **little or no evidence against the null hypothesis**



Two sides of the same coin?

- Often, hypothesis testing and p-values are seen as the same thing. **They are not!**

Two sides of the same coin?

- Often, hypothesis testing and p-values are seen as the same thing. **They are not!**
- **Hypothesis testing** (HT)
 - Considers formally two competing hypotheses – a "null" and an "alternative" (**NB**: that determines the treatment effect)
 - **Sets** the probabilities of error α and β
 - Aims at "rejecting" the null – so it has a binary outcome (yes/no)
- **Significance testing** (ST, p-values)
 - Concerned with the sampling distribution of the data under the null hypothesis
 - Measures the strength of the evidence for/against the null, but has no formal involvement of alternative explanations for the observed data

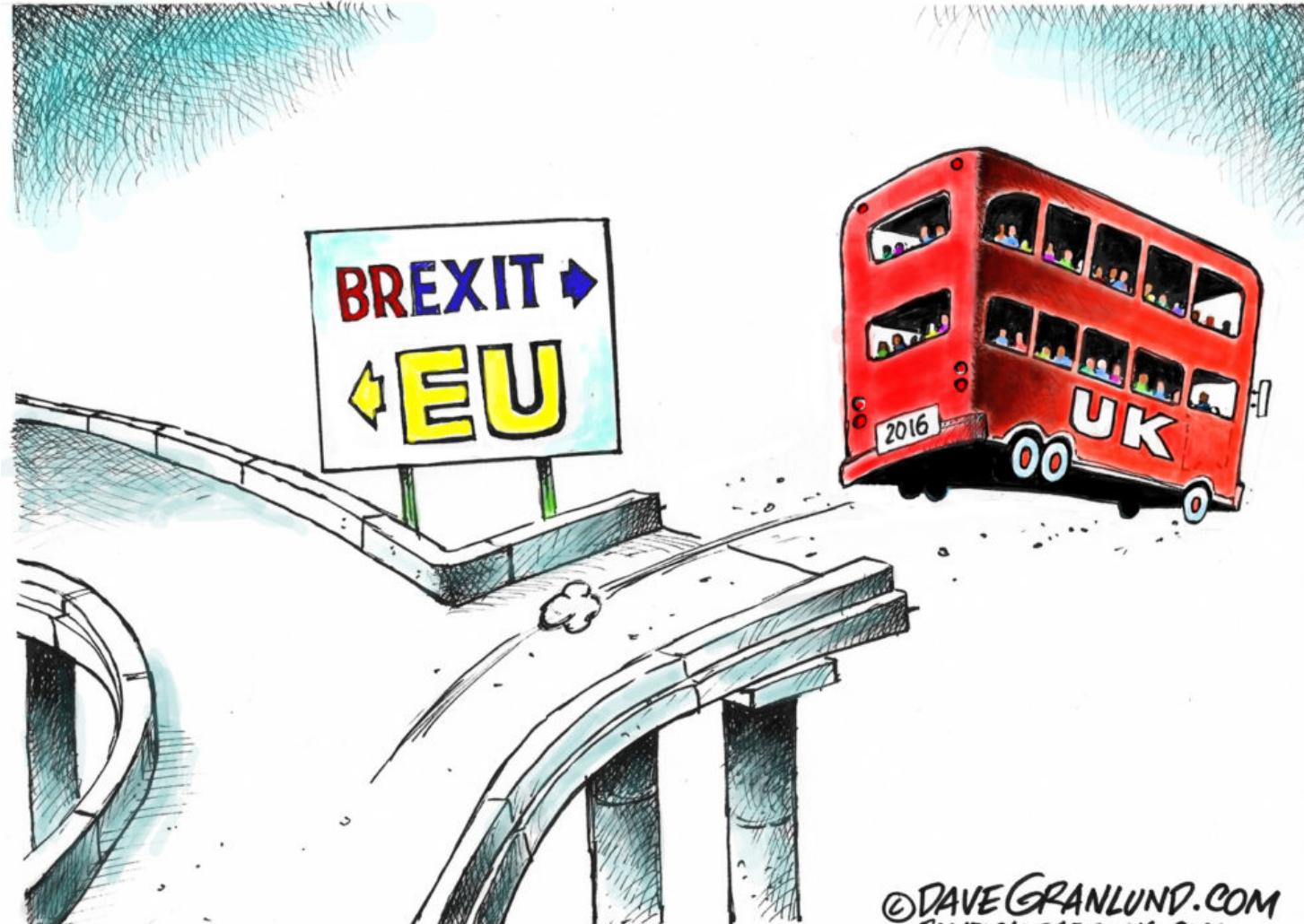
Two sides of the same coin?

- Often, hypothesis testing and p-values are seen as the same thing. **They are not!**
- **Hypothesis testing** (HT)
 - Considers formally two competing hypotheses – a "null" and an "alternative" (**NB**: that determines the treatment effect)
 - **Sets** the probabilities of error α and β
 - Aims at "rejecting" the null – so it has a binary outcome (yes/no)
- **Significance testing** (ST, p-values)
 - Concerned with the sampling distribution of the data under the null hypothesis
 - Measures the strength of the evidence for/against the null, but has no formal involvement of alternative explanations for the observed data
- $p \neq \alpha$ even if often the **threshold** is set at 0.05 for both!
 - α is **set** by the researcher
 - p is **computed** from the data (as extreme or more extreme than those observed)

Two sides of the same coin?

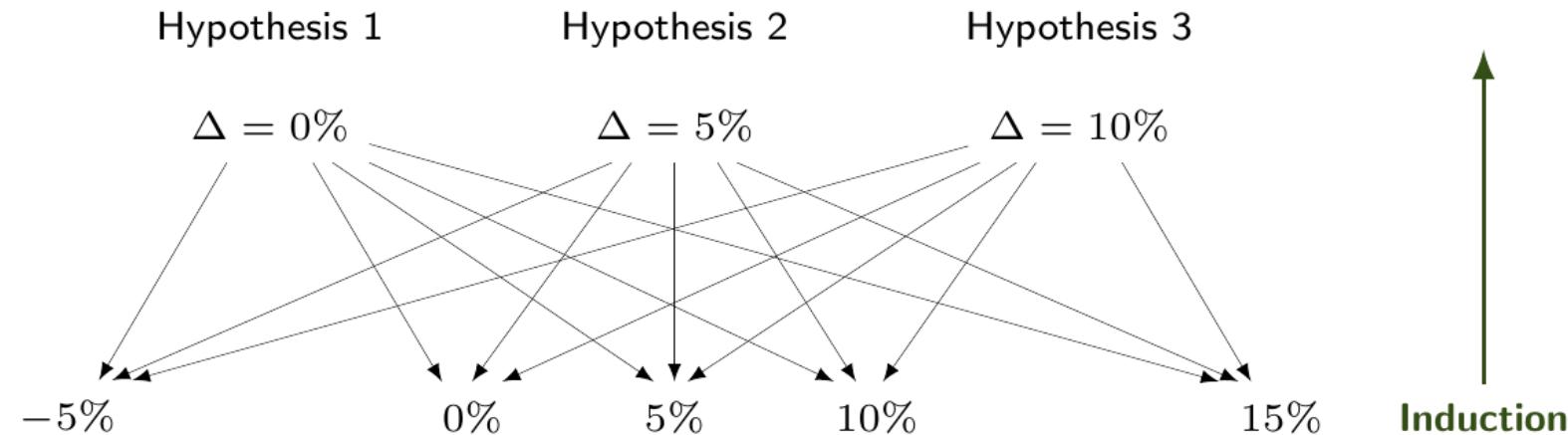
- Often, hypothesis testing and p-values are seen as the same thing. **They are not!**
- **Hypothesis testing** (HT)
 - Considers formally two competing hypotheses – a "null" and an "alternative" (**NB**: that determines the treatment effect)
 - **Sets** the probabilities of error α and β
 - Aims at "rejecting" the null – so it has a binary outcome (yes/no)
- **Significance testing** (ST, p-values)
 - Concerned with the sampling distribution of the data under the null hypothesis
 - Measures the strength of the evidence for/against the null, but has no formal involvement of alternative explanations for the observed data
- $p \neq \alpha$ even if often the **threshold** is set at 0.05 for both!
 - α is **set** by the researcher
 - p is **computed** from the data (as extreme or more extreme than those observed)
- **NB:** Confusingly, experimental studies are **designed** under a HT setting, but **analysed** under a ST setting!
- Increasing recognition of pitfalls in science ([here](#) and [here](#))

Is there another way?...



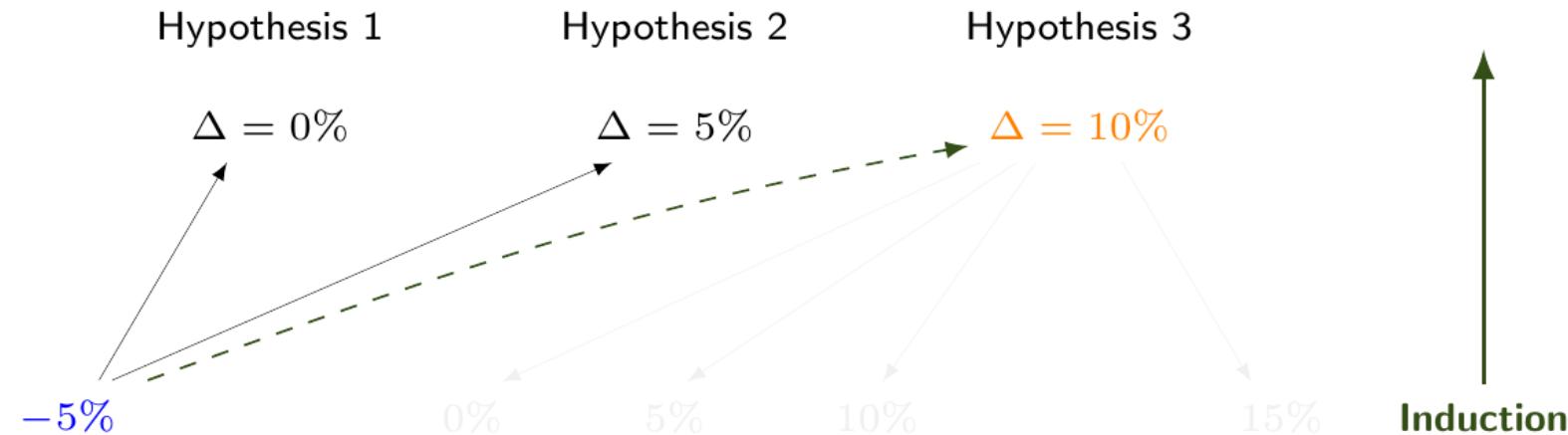
© DAVE GRANLUND.COM
POLITICALCARTOONS.COM

Deductive vs **inductive** inference



- The **Bayesian** philosophy proceeds fixing the value of the observed data and, **by induction**, makes inference on unobservable hypotheses
 - What is the probability of my hypothesis, given the data I observed? If less than the probability of other competing hypotheses, then weak support of the evidence to the hypothesis

Deductive vs **inductive** inference



- The **Bayesian** philosophy proceeds fixing the value of the observed data and, **by induction**, makes inference on unobservable hypotheses
 - What is the probability of my hypothesis, given the data I observed? If less than the probability of other competing hypotheses, then weak support of the evidence to the hypothesis
 - Assess $\Pr(\text{Hypothesis} \mid \text{Observed data})$
 - Can express in terms of an **interval estimate**: $\Pr(a \leq \text{parameter} \leq b \mid \text{Data})$
 - **NB:** Unobserved data have no role in the inference!

How did it all start?

In 1763, Reverend Thomas Bayes of Tunbridge Wells wrote

P R O B L E M.

Given the number of times in which an unknown event has happened and failed: *Required* the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named.

In modern language, given $r \sim \text{Binomial}(\theta, n)$, what is $\Pr(\theta_1 < \theta < \theta_2 \mid r, n)$?

Some historical references

 <http://www.bayesian.org/resources/bayes.html>

 S. Bertsch McGrayne (2011). *The Theory That Would Not Die*

see [Lecture 2](#)

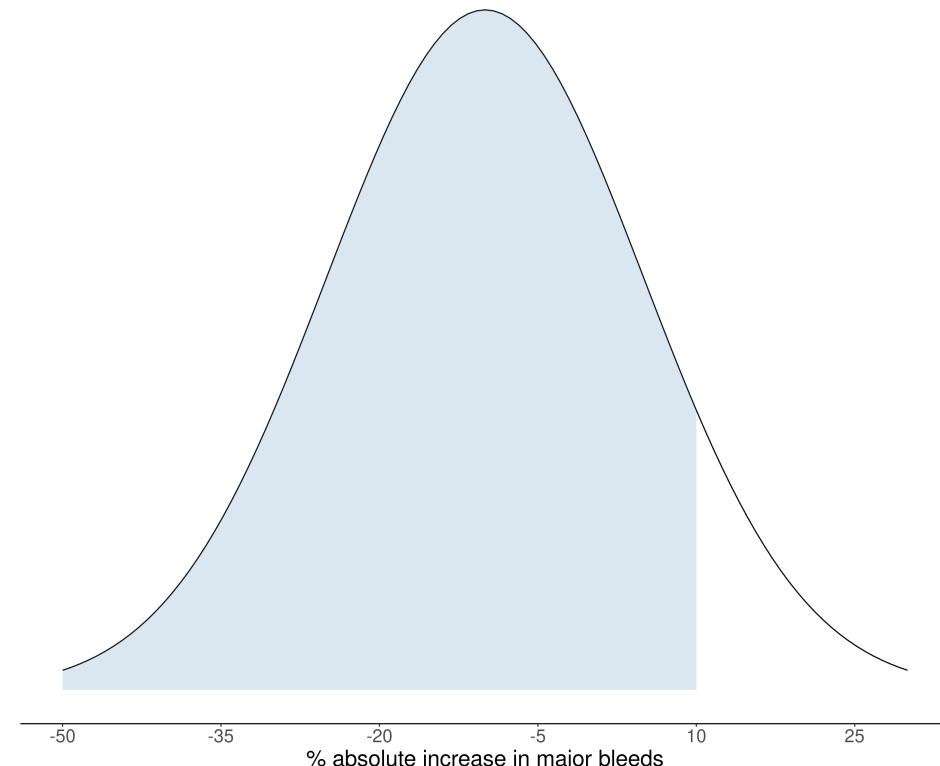
 S. Fienberg (2006). *When did Bayesian inference become Bayesian?*

Basic ideas

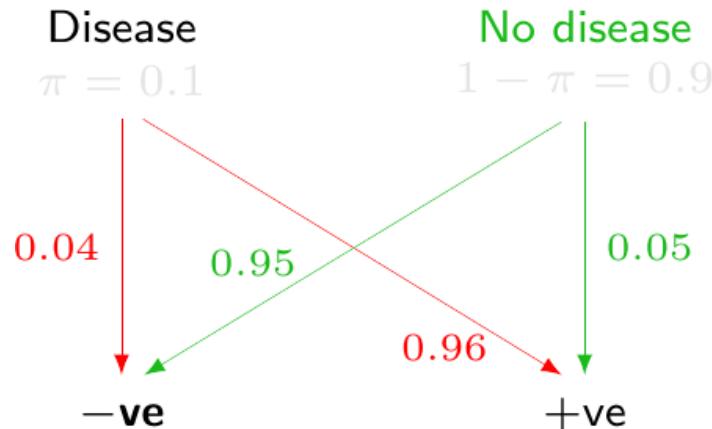
Direct expression of uncertainty about unknown parameters

"There is an 89% probability that the absolute increase in major bleeds is less than 10 percent with low-dose PLT transfusions"

( [Tinmouth et al, Transfusion, 2004](#))

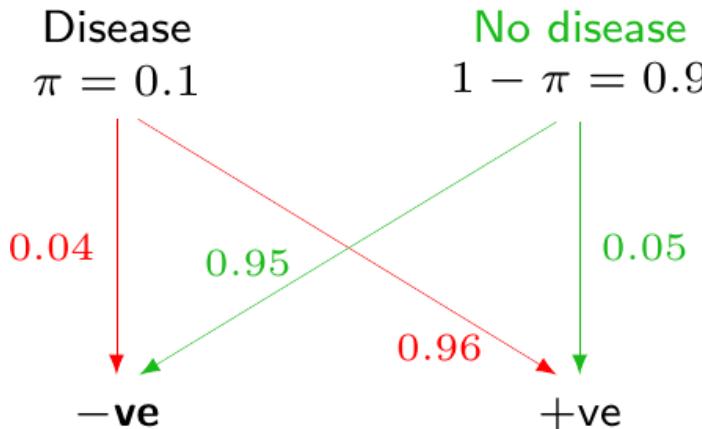


Basic ideas



- Suppose a patient is tested for HIV. The test comes up negative (-ve)
- Given the assumptions/model, this indicates **fairly strong** evidence against the hypothesis that the true status is "Disease", so basically $p = 0.04$

Basic ideas



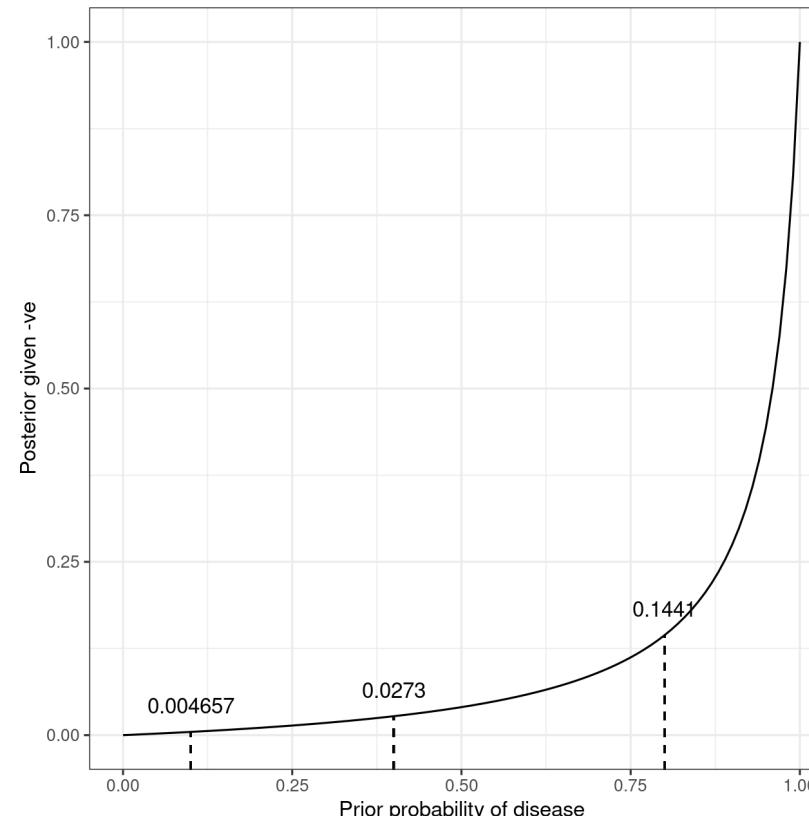
- Suppose a patient is tested for HIV. The test comes up negative (-ve)
- Given the assumptions/model, this indicates **fairly strong** evidence against the hypothesis that the true status is "Disease", so basically $p = 0.04$
- But: how **prevalent** is the disease in the population?
 - We can model our prior knowledge about this and combine this information with the evidence from the data (using **Bayes theorem**)

$$\Pr(\text{Disease} \mid \text{-ve}) = \frac{\Pr(\text{Disease})\Pr(\text{-ve} \mid \text{Disease})}{\Pr(\text{-ve})}$$

- Update uncertainty given the evidence provided by the data

Prior vs posterior

- The evidence **from the data alone** tells us that the observed result is extremely unlikely under the hypothesis of "Disease"
- This is strongly dependent on the **context**, as provided by the prior knowledge/epistemic uncertainty, though!



Basic ideas

- A Bayesian model specifies a **full probability distribution** to describe uncertainty
- This applies to
 - **Data**, which are subject to **sampling variability**
 - **Parameters** (or hypotheses), typically unobservable and thus subject to **epistemic uncertainty**
 - And even future, yet unobserved realisations of the observable variables (data)

Basic ideas

- A Bayesian model specifies a **full probability distribution** to describe uncertainty
- This applies to
 - **Data**, which are subject to **sampling variability**
 - **Parameters** (or hypotheses), typically unobservable and thus subject to **epistemic uncertainty**
 - And even future, yet unobserved realisations of the observable variables (data)
- Probability is the only language in the Bayesian framework to assess any form of imperfect information or knowledge
 - No need to distinguish between probability and confidence
 - Before even seeing the data, we need to identify a suitable probability distribution to describe the overall uncertainty about the data y and the parameters θ

Basic ideas

- A Bayesian model specifies a **full probability distribution** to describe uncertainty
- This applies to
 - **Data**, which are subject to **sampling variability**
 - **Parameters** (or hypotheses), typically unobservable and thus subject to **epistemic uncertainty**
 - And even future, yet unobserved realisations of the observable variables (data)
- Probability is the only language in the Bayesian framework to assess any form of imperfect information or knowledge
 - No need to distinguish between probability and confidence
 - Before even seeing the data, we need to identify a suitable probability distribution to describe the overall uncertainty about the data \mathbf{y} and the parameters $\boldsymbol{\theta}$

$$p(\mathbf{y}, \boldsymbol{\theta}) = p(\boldsymbol{\theta})p(\mathbf{y} \mid \boldsymbol{\theta}) = p(\mathbf{y})p(\boldsymbol{\theta} \mid \mathbf{y})$$

(see also [Lecture 5](#)) from which we derive Bayes Theorem

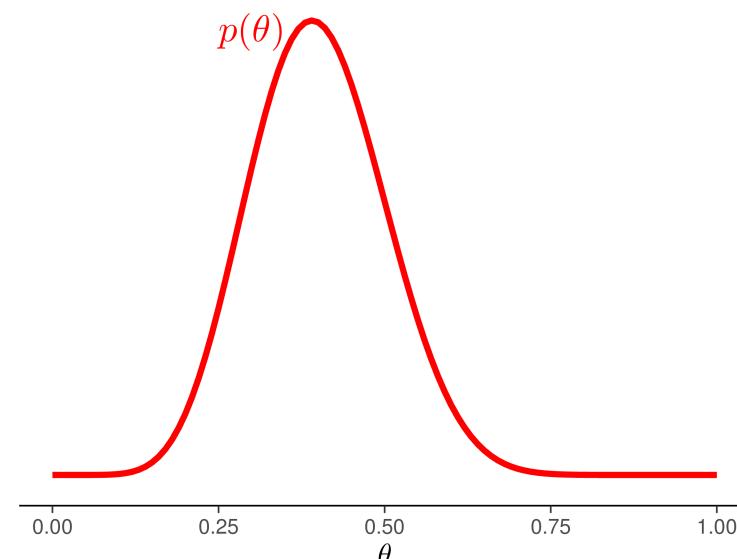
$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{p(\boldsymbol{\theta})p(\mathbf{y} \mid \boldsymbol{\theta})}{p(\mathbf{y})}$$

- Express beliefs in form of a probability distribution

(Super) silly example: drug

Existing knowledge

- Population registries
- Observational studies
- Small/pilot RCTs
- Expert opinion



Encode the assumption that a drug has a response rate between 20% and 60%

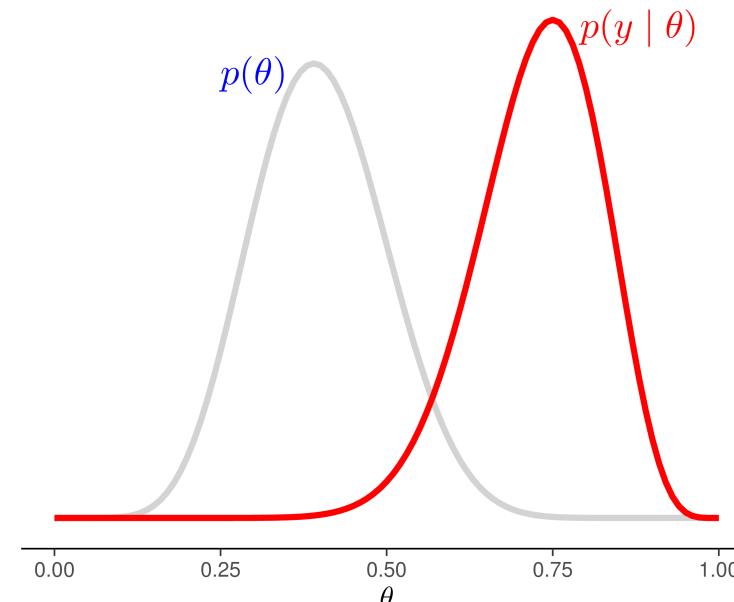
(Super) silly example: drug

Existing knowledge

- Population registries
- Observational studies
- Small/pilot RCTs
- Expert opinion

Current data

- Large(r) scale RCT
- Observational study
- Relevant summaries



Observe a study with 150 responders out of 200 patients given the drug

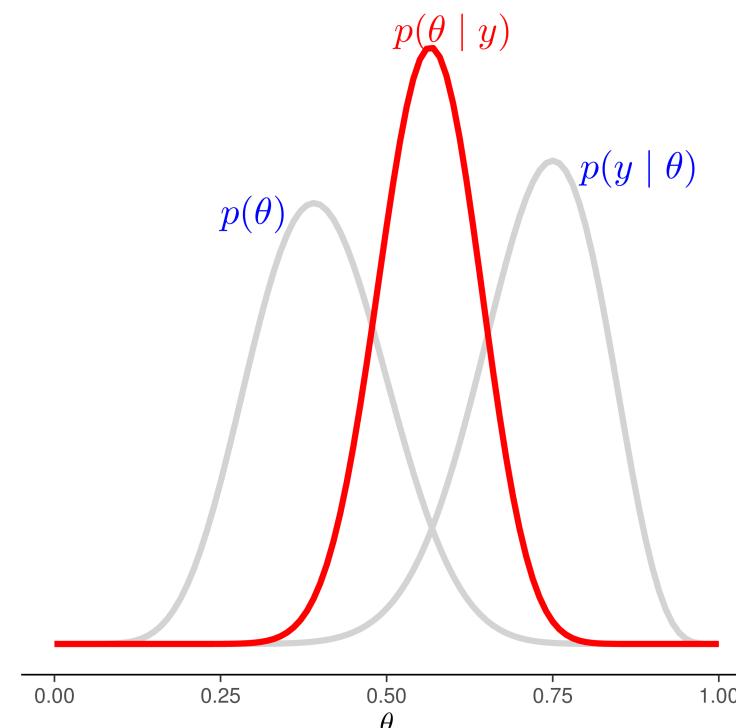
(Super) silly example: drug

Existing knowledge

- Population registries
- Observational studies
- Small/pilot RCTs
- Expert opinion

Current data

- Large(r) scale RCT
- Observational study
- Relevant summaries



Update knowledge to describe revised "state of science"

*But how can I form a prior? I know **nothing** about this parameter!...*



*But how can I form a prior? I know **nothing** about this parameter!...*



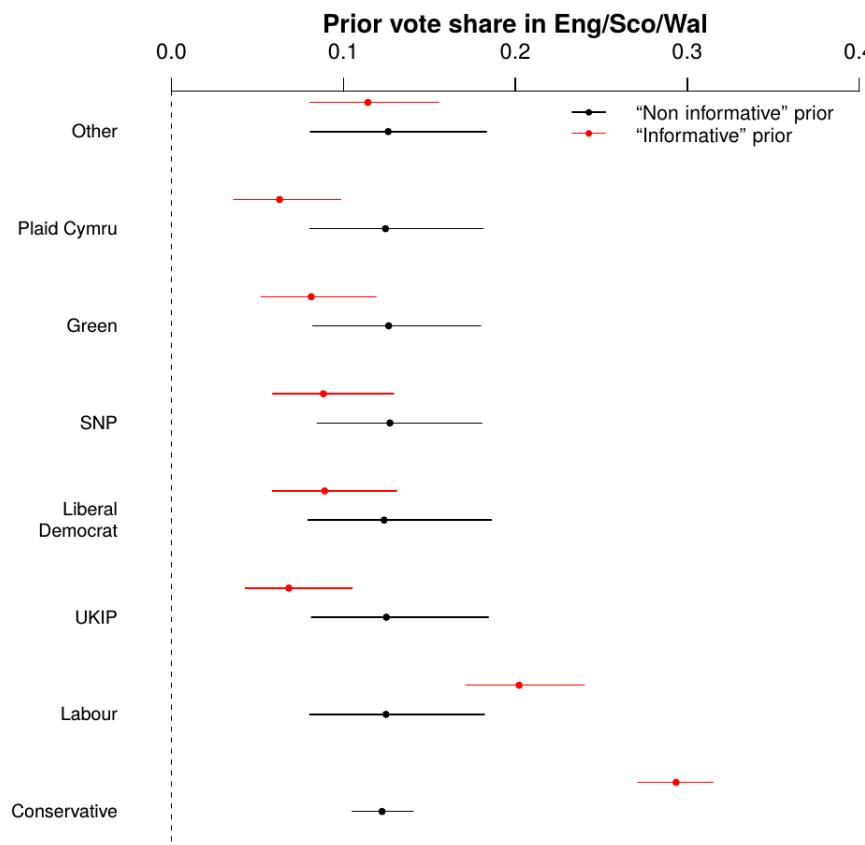
*But how can I form a prior? I know **nothing** about this parameter!...*



*But how can I form a prior? I know **nothing** about this parameter!...*



- Predicting the output of the 2017 UK General Election using poll data (see [here](#) and subsequent posts)
 - Data: number of people out of the N_i respondents in poll i intending to vote for party p (multinomial counts)
 - Objective of estimation:** $(\pi_1, \dots, \pi_P) =$ population vote share for each party
 - Can model $\pi_p = (\phi_p / \sum \phi_p)$ and $\log(\phi_p) = \alpha_p + \beta_p X_p$



→ [Next lecture](#)