Course introduction

Michael Noonan

December 29, 2020

Biol 520C: Statistical modelling for biological data

Table of contents

Course Overview

About Me

Name: Michael Noonan

About Me

Name: Michael Noonan

Office: SCI 379

About Me

Name: Michael Noonan

Office: SCI 379

Email: michael.noonan [at] ubc.ca (use subject heading BIOL520C in all email communication)

About Me

Name: Michael Noonan

Office: SCI 379

Email: michael.noonan [at] ubc.ca (use subject heading BIOL520C in all email communication)

Office Hours: Thurs 13h-14h; Fri 10h-12h, or by appointment arranged via email.

About Me

Name: Michael Noonan

Office: SCI 379

Email: michael.noonan [at] ubc.ca (use subject heading BIOL520C in all email communication)

Office Hours: Thurs 13h-14h; Fri 10h-12h, or by appointment arranged via email.

Course Website: https://noonanm.github.io/Biol520C/index.html

What this course is about

What this course is about

- Focus is on model-based inference

What this course is about

- Focus is on model-based inference (i.e., combining data with models to generate mechanistic descriptions of biological patterns).

What this course is about

- Focus is on model-based inference (i.e., combining data with models to generate mechanistic descriptions of biological patterns).

- Building from simple linear regression, you'll learn regression methods for handling the most routinely encountered features in biological data

What this course is about

- Focus is on model-based inference (i.e., combining data with models to generate mechanistic descriptions of biological patterns).
- Building from simple linear regression, you'll learn regression methods for handling the most routinely encountered features in biological data (hierarchical data structures, non-Gaussian error distributions, non-linearity, autocorrelation, etc...).

What this course is about

- Focus is on model-based inference (i.e., combining data with models to generate mechanistic descriptions of biological patterns).
- Building from simple linear regression, you'll learn regression methods for handling the most routinely encountered features in biological data (hierarchical data structures, non-Gaussian error distributions, non-linearity, autocorrelation, etc...).
- Emphasis on statistical best practices.

What this course is about

- Focus is on model-based inference (i.e., combining data with models to generate mechanistic descriptions of biological patterns).
- Building from simple linear regression, you'll learn regression methods for handling the most routinely encountered features in biological data (hierarchical data structures, non-Gaussian error distributions, non-linearity, autocorrelation, etc...).
- Emphasis on statistical best practices.
- How to use open source software (R) to apply these analyses.

What this course is not about

What this course is not about

- Basic statistics (concepts like means, medians, variances, probability distributions, regression should be familiar to you).

What this course is not about

- Basic statistics (concepts like means, medians, variances, probability distributions, regression should be familiar to you).

- Math

What this course is not about

- Basic statistics (concepts like means, medians, variances, probability distributions, regression should be familiar to you).

- Math you should be familiar with basic calculus (derivatives and integrals) and linear algebra (operations on vectors and matrices).

What this course is not about

- Basic statistics (concepts like means, medians, variances, probability distributions, regression should be familiar to you).
- Math you should be familiar with basic calculus (derivatives and integrals) and linear algebra (operations on vectors and matrices).
- Computer programming (we will be using `R`, but the course is not focused on 'how to code').

What this course is not about

- Basic statistics (concepts like means, medians, variances, probability distributions, regression should be familiar to you).
- Math you should be familiar with basic calculus (derivatives and integrals) and linear algebra (operations on vectors and matrices).
- Computer programming (we will be using `R`, but the course is not focused on 'how to code').
- Methods for handling *ad hoc*, corner cases.

How the course is structured

How the course is structured

- For each topic, there will be a core lecture and an associated practical assignment.

How the course is structured

- For each topic, there will be a core lecture and an associated practical assignment.

- Lectures will cover the core concepts of the course. Lecture slides will be posted on the course website the evening prior to the lecture. You are encouraged to take notes, and to ask questions in the lectures. All lectures will be recorded and made available to you.

How the course is structured

- For each topic, there will be a core lecture and an associated practical assignment.

- Lectures will cover the core concepts of the course. Lecture slides will be posted on the course website the evening prior to the lecture. You are encouraged to take notes, and to ask questions in the lectures. All lectures will be recorded and made available to you.

- The practicals use structured tutorials to guide you on the use of the open-source software program `R` for applying the methods learned in the lectures to data. The lectures and practicals are designed to be *complementary* and not all the material in the practicals will be covered in the lectures and vice versa.

Course Evaluation

Course Evaluation

Practicals (10)	20%	Due on $\sim$weekly basis
Participation from practicals	10%	Exact sched. in course outline

Course Evaluation

Practicals (10)	20%	Due on ~weekly basis
Participation from practicals	10%	Exact sched. in course outline
Paper #1	25%	Week 7
Presentation #1	10%	Week 7

Course Evaluation

THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

Practicals (10)	20%	Due on ∼weekly basis
Participation from practicals	10%	Exact sched. in course outline
Paper #1	25%	Week 7
Presentation #1	10%	Week 7
Paper #2	25%	Week 14
Presentation #1	10%	Weeks 13 & 14

Course Evaluation

Practicals (10)	20%	Due on ∼weekly basis
Participation from practicals	10%	Exact sched. in course outline
Paper #1	25%	Week 7
Presentation #1	10%	Week 7
Paper #2	25%	Week 14
Presentation #1	10%	Weeks 13 & 14
Total	100%	

Practicals and participation

Practicals and participation

Beginning the week of January 10, 2020, you will be asked to complete practical assignments on an ∼weekly basis. There will be a total of 10 practicals to be completed throughout the course.

Practicals and participation

Beginning the week of January 10, 2020, you will be asked to complete practical assignments on an $\sim$weekly basis. There will be a total of 10 practicals to be completed throughout the course.

The course web page on github will host the practicals, and the various datasets associated with each practical. Lectures will be given on Mondays and Tuesdays. After the Tuesday lecture, we will have covered all of the material that is needed to complete the week's practical assignment material which is due before the start of the following Tuesday lecture (to be submitted online).

Practicals and participation

Beginning the week of January 10, 2020, you will be asked to complete practical assignments on an ∼weekly basis. There will be a total of 10 practicals to be completed throughout the course.

The course web page on github will host the practicals, and the various datasets associated with each practical. Lectures will be given on Mondays and Tuesdays. After the Tuesday lecture, we will have covered all of the material that is needed to complete the week's practical assignment material which is due before the start of the following Tuesday lecture (to be submitted online).

Grading: Each practical assignment is worth a total of 3% of your total grade. Of this, 1% is given for submitting the tutorial on time, irrespective of whether or not the answers are correct (participation). The remaining 2% comes from the answers provided. Late practicals will be accepted, but will only be worth a maximum of 2%.

Papers 1 and 2

Papers 1 and 2

You will be required to apply the modelling tools covered in the lectures on two dataset and write two short papers, each comprised of 6 sections:

You will be required to apply the modelling tools covered in the lectures on two dataset and write two short papers, each comprised of 6 sections:

- **Introduction:** Provide a brief description of the study system from which the data come and an outline of what questions you intend on exploring with the data. **(12.5%)**

Papers 1 and 2

You will be required to apply the modelling tools covered in the lectures on two dataset and write two short papers, each comprised of 6 sections:

- **Introduction:** Provide a brief description of the study system from which the data come and an outline of what questions you intend on exploring with the data. **(12.5%)**
- **Methods:** Describe how the data were collected, what variables are included, and what analyses were applied. **(20%)**

Papers 1 and 2

THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

You will be required to apply the modelling tools covered in the lectures on two dataset and write two short papers, each comprised of 6 sections:

- **Introduction:** Provide a brief description of the study system from which the data come and an outline of what questions you intend on exploring with the data. **(12.5%)**
- **Methods:** Describe how the data were collected, what variables are included, and what analyses were applied. **(20%)**
- **Results:** Length: Describe your statistical findings. Tables and figures should be used throughout. **(20%)**

Papers 1 and 2

You will be required to apply the modelling tools covered in the lectures on two dataset and write two short papers, each comprised of 6 sections:

- **Introduction:** Provide a brief description of the study system from which the data come and an outline of what questions you intend on exploring with the data. **(12.5%)**
- **Methods:** Describe how the data were collected, what variables are included, and what analyses were applied. **(20%)**
- **Results:** Length: Describe your statistical findings. Tables and figures should be used throughout. **(20%)**
- **Discussion:** Provide a brief summary of your findings and place them in a biological context. **(12.5%)**

Papers 1 and 2

You will be required to apply the modelling tools covered in the lectures on two dataset and write two short papers, each comprised of 6 sections:

- **Introduction:** Provide a brief description of the study system from which the data come and an outline of what questions you intend on exploring with the data. **(12.5%)**
- **Methods:** Describe how the data were collected, what variables are included, and what analyses were applied. **(20%)**
- **Results:** Length: Describe your statistical findings. Tables and figures should be used throughout. **(20%)**
- **Discussion:** Provide a brief summary of your findings and place them in a biological context. **(12.5%)**
- **References:** Include references to all necessary literature and statistical packages employed. **(5%)**

Papers 1 and 2

You will be required to apply the modelling tools covered in the lectures on two dataset and write two short papers, each comprised of 6 sections:

- **Introduction:** Provide a brief description of the study system from which the data come and an outline of what questions you intend on exploring with the data. **(12.5%)**
- **Methods:** Describe how the data were collected, what variables are included, and what analyses were applied. **(20%)**
- **Results:** Length: Describe your statistical findings. Tables and figures should be used throughout. **(20%)**
- **Discussion:** Provide a brief summary of your findings and place them in a biological context. **(12.5%)**
- **References:** Include references to all necessary literature and statistical packages employed. **(5%)**
- **Appendix:** The appendix material should include an R markdown document that details every step of the analyses. **(30%)**

Papers 1 and 2 cont.

Datasets: To complete these assignments, you will have access to a number of pre-selected datasets. You can opt to use your own data to complete these assignments if you prefer, and are encouraged to do so, but you must seek instructor approval. If you intend on using your own data, it is recommended that you discuss this with me as early as possible.

Papers 1 and 2 cont.

Datasets: To complete these assignments, you will have access to a number of pre-selected datasets. You can opt to use your own data to complete these assignments if you prefer, and are encouraged to do so, but you must seek instructor approval. If you intend on using your own data, it is recommended that you discuss this with me as early as possible.

Late Assignments: You are to submit Paper 1 by the end of the day on Feb. 27th, and paper 2 by the end of the day on Apr 17th. Late papers will have 10% deducted per day that they are overdue, and will receive a grade of zero if more than 10 days late without a valid excuse.

Presentations 1 and 2

Presentations 1 and 2

Prior to submitting their papers to the instructor, students will be required to give a 10-minute presentation to the class.

Presentations 1 and 2

Prior to submitting their papers to the instructor, students will be required to give a 10-minute presentation to the class.

The presentation should include all of the sections that are included in the paper, however the appendix detailing the `R` code that was used should be integrated into the methods section of the presentation.

Prior to submitting their papers to the instructor, students will be required to give a 10-minute presentation to the class.

The presentation should include all of the sections that are included in the paper, however the appendix detailing the `R` code that was used should be integrated into the methods section of the presentation.

Grading: Grading will based on a rubric provided in advance.

Optional Material

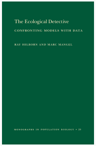
Optional Material

There is no textbook for this course, but if you are interested in expanding your knowledge beyond what is covered, the following are recommended:

Optional Material

There is no textbook for this course, but if you are interested in expanding your knowledge beyond what is covered, the following are recommended:

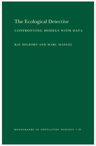
- Hilborn R, Mangel M. The ecological detective: confronting models with data. 1997. Princeton University Press. $\sim$ \$15



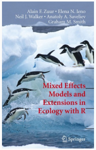
Optional Material

There is no textbook for this course, but if you are interested in expanding your knowledge beyond what is covered, the following are recommended:

- Hilborn R, Mangel M. The ecological detective: confronting models with data. 1997. Princeton University Press. ∼ $15



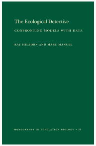
- Zuur, A et al. (2009). Mixed effects models and extensions in ecology with R. Springer. ∼ $90



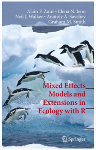
Optional Material

There is no textbook for this course, but if you are interested in expanding your knowledge beyond what is covered, the following are recommended:

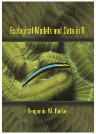
- Hilborn R, Mangel M. The ecological detective: confronting models with data. 1997. Princeton University Press. $\sim$ \$15



- Zuur, A et al. (2009). Mixed effects models and extensions in ecology with R. Springer. $\sim$ \$90



- Bolker, B. M. (2008). Ecological models and data in R. Princeton University Press. $\sim$ \$50



Lecture outline

Week/Dates	Lecture Topics
1 – Jan 10-16	Course introduction; Regression refresher
2 – Jan 17-23	Probability theory; Likelihood; Maximum likelihood
3 – Jan 24-30	Mult. linear regression; Param inter.; Interpreting residuals
4 – Jan 31-Feb 6	Mixed effects models; Model Selection; Information criterion
5 – Feb 7-13	Model Selection; Model averaging; Independent Project Work
6 – Feb 14-20	Mid-term break, no lectures
7 – Feb 21-27	Student presentations (**10%**) & Term paper 1 due (**25%**)
8 – Feb 28-Mar 6	Heteroskedasticity; Temporal autocorrelation
9 – Mar 7-13	Spatial Autocorrelation; Phylogenetic inertia
10 – Mar 14-20	Logistic and Poisson regression
11 – Mar 21-27	Non-linear modelling; Deterministic functions; Stochastic simulation and power analysis
12 – Mar 28-Apr3	Course Overview; Independent Project Work
13 – Apr 4-10	No lecture on Monday; Student presentations (10%)
14 – Apr 11-17	Student presentations (**10%**) & Term paper 1 due (**25%**)

Design- vs. Model-based Inference

The Scientific Method

Science is a process of learning about nature. As scientists, we weigh competing ideas about how the world works (hypotheses) against observations (data).

The Scientific Method

Science is a process of learning about nature. As scientists, we weigh competing ideas about how the world works (hypotheses) against observations (data). But our descriptions of the world are almost always incomplete, our observations have error, and important data is often missing...

The Scientific Method

Science is a process of learning about nature. As scientists, we weigh competing ideas about how the world works (hypotheses) against observations (data). But our descriptions of the world are almost always incomplete, our observations have error, and important data is often missing... So, how do we accurately compare what we observe with what we hypothesize without bias?

The Scientific Method

Science is a process of learning about nature. As scientists, we weigh competing ideas about how the world works (hypotheses) against observations (data). But our descriptions of the world are almost always incomplete, our observations have error, and important data is often missing... So, how do we accurately compare what we observe with what we hypothesize without bias?

Statistics

Science is a process of learning about nature. As scientists, we weigh competing ideas about how the world works (hypotheses) against observations (data). But our descriptions of the world are almost always incomplete, our observations have error, and important data is often missing... So, how do we accurately compare what we observe with what we hypothesize without bias?

Statistics

The process of making scientific inference can be split into two broad categories:

Science is a process of learning about nature. As scientists, we weigh competing ideas about how the world works (hypotheses) against observations (data). But our descriptions of the world are almost always incomplete, our observations have error, and important data is often missing... So, how do we accurately compare what we observe with what we hypothesize without bias?

Statistics

The process of making scientific inference can be split into two broad categories: design-based

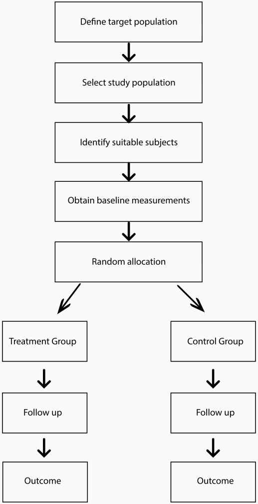
The Scientific Method

Science is a process of learning about nature. As scientists, we weigh competing ideas about how the world works (hypotheses) against observations (data). But our descriptions of the world are almost always incomplete, our observations have error, and important data is often missing... So, how do we accurately compare what we observe with what we hypothesize without bias?

Statistics

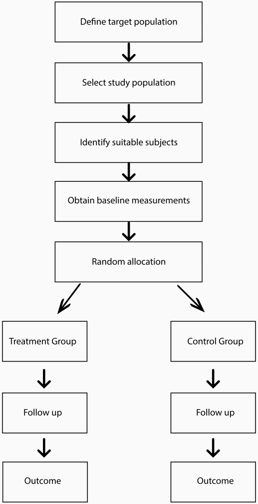
The process of making scientific inference can be split into two broad categories: design-based and model-based

Design-based inference



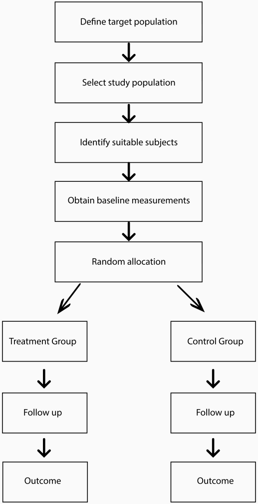
Design-based inference

- In design-based inference, most of the focus is on experimental design



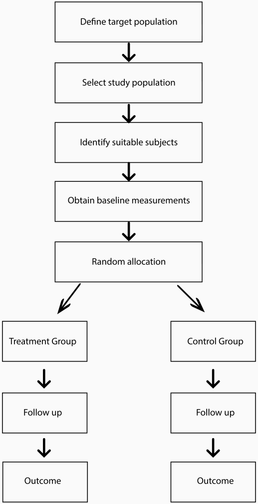
Design-based inference

- In design-based inference, most of the focus is on experimental design
- You assume your sample is a random sample of a target population



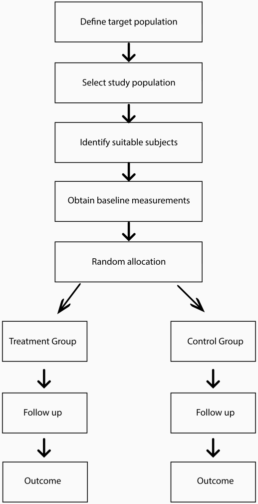
Design-based inference

- In design-based inference, most of the focus is on experimental design
- You assume your sample is a random sample of a target population
- Inference is focused towards a target population



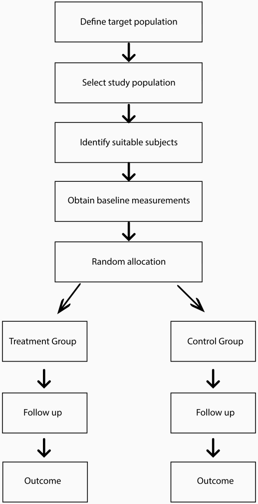
Design-based inference

- In design-based inference, most of the focus is on experimental design
- You assume your sample is a random sample of a target population
- Inference is focused towards a target population
- Goal is to be able to demonstrate that 'x causes y'

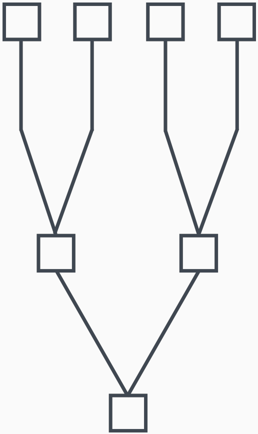


Design-based inference

- In design-based inference, most of the focus is on experimental design
- You assume your sample is a random sample of a target population
- Inference is focused towards a target population
- Goal is to be able to demonstrate that 'x causes y'
- Data are typically analysed by comparing means and variances across groups (e.g., ANOVAs, t-tests, etc...)

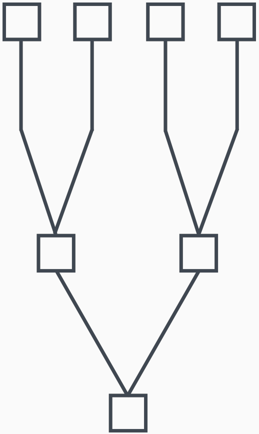


Worklow of design-based inference



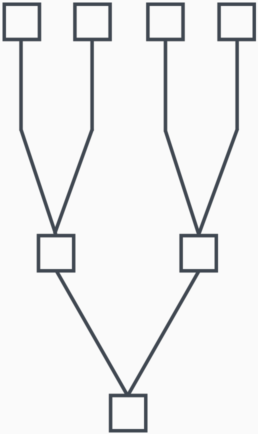
Worklow of design-based inference

- Devise a hypotheses



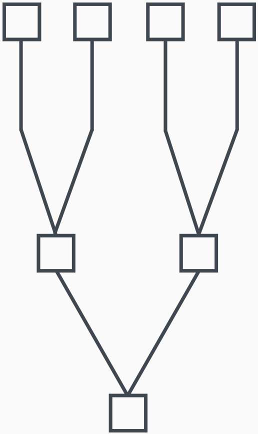
Worklow of design-based inference

- Devise a hypotheses
- Devise an experiment with outcomes that will clearly accept or reject the hypothesis



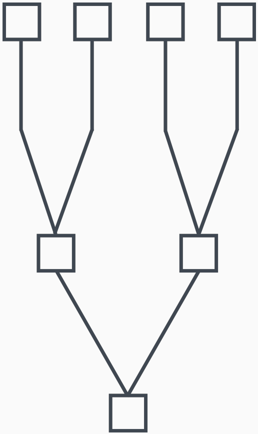
Worklow of design-based inference

- Devise a hypotheses
- Devise an experiment with outcomes that will clearly accept or reject the hypothesis
- Carry out the experiment so as to get a clean result



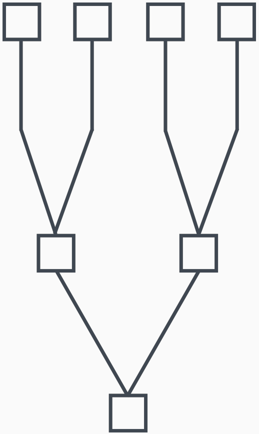
Worklow of design-based inference

- Devise a hypotheses
- Devise an experiment with outcomes that will clearly accept or reject the hypothesis
- Carry out the experiment so as to get a clean result
- Recycle the procedure to refine the remaining possibilities (Platt, 1964)



Worklow of design-based inference

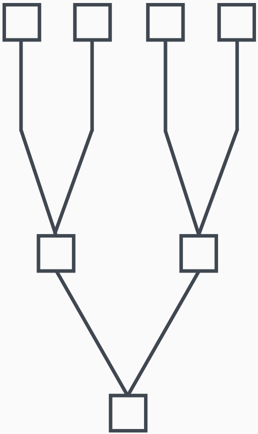
- Devise a hypotheses
- Devise an experiment with outcomes that will clearly accept or reject the hypothesis
- Carry out the experiment so as to get a clean result
- Recycle the procedure to refine the remaining possibilities (Platt, 1964)



Core of design-based inference is confronting single hypotheses with data

Limitations of design-based inference

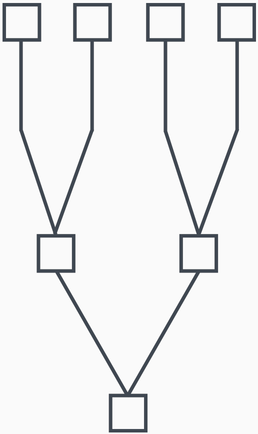
Platt's decision tree is based on:



Limitations of design-based inference

Platt's decision tree is based on:

i) Clear, distinct hypotheses

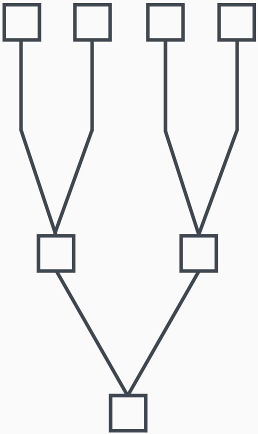


Limitations of design-based inference

Platt's decision tree is based on:

i) Clear, distinct hypotheses

ii) Unambiguous outcomes



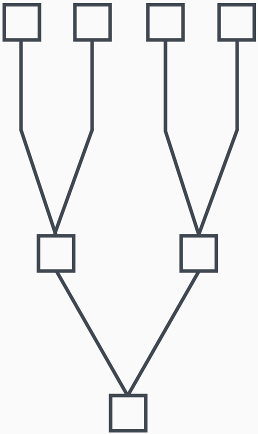
Limitations of design-based inference

Platt's decision tree is based on:

i) Clear, distinct hypotheses

ii) Unambiguous outcomes

iii) A relationship between statistical significance and biological relevance



Limitations of design-based inference

Many biological processes have long time-scales.

Long time-scales

Long time-scales

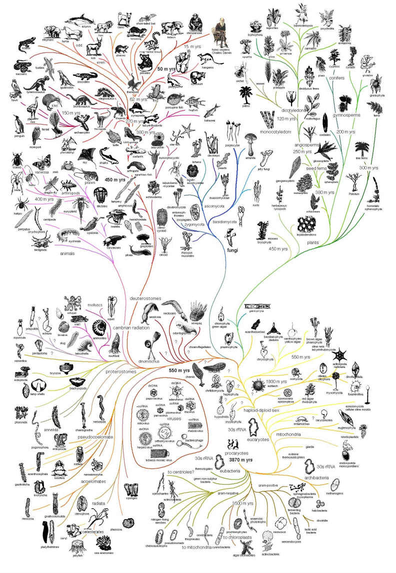
Humpback whales (*Megaptera novaeangliae*) can live for 50+ years. Source: David Valencia

Long time-scales


Humpback whales (*Megaptera novaeangliae*) can live for 50+ years. Source: David Valencia


Bristlecone pines (*Pinus longaeva*) live for thousands of years. Source: wired.com

Long time-scales



Source: Chris King

Limitations of design-based inference

Many biological processes have long time-scales.

Limitations of design-based inference

Many biological processes have long time-scales.

Many biological systems have very poor reproducibility.

Poor reproducibility

Source: Tom and Pat Leeson

Limitations of design-based inference

Many biological processes have long time-scales.

Many biological systems have very poor reproducibility.

How can you design a controlled experiment in a wild population?

Limitations of design-based inference

What if you're interested in species conservation?

What if you're interested in species conservation?

Source: Wikipedia

In 1988 the wild pop. of black footed ferrets (*Mustela nigripes*) was down to 18 ind.

Limitations of design-based inference

What if you're interested in species conservation?

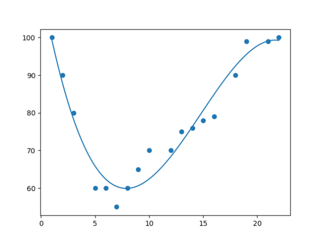
Source: Wikipedia

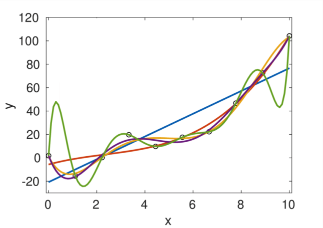
In 1988 the wild pop. of black footed ferrets (*Mustela nigripes*) was down to 18 ind.

What do you do if a power analysis says you need 20 animals?

Model-based inference

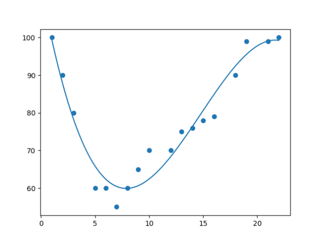
THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

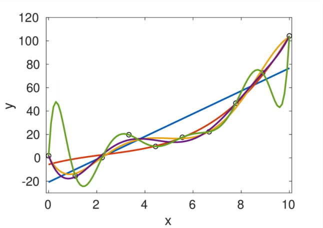




Model-based inference

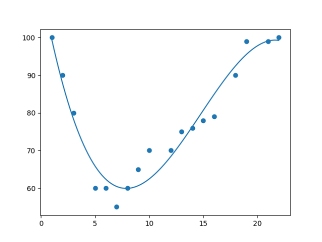
- In model-based inference, most of the focus is on identifying an unknown deterministic model

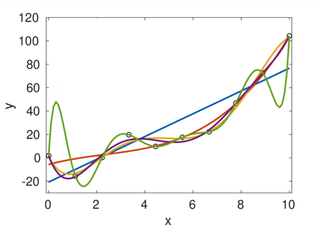




Model-based inference

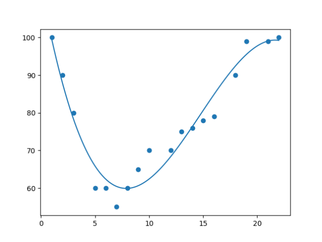
- In model-based inference, most of the focus is on identifying an unknown deterministic model
- Inference can be extrapolated beyond the target population

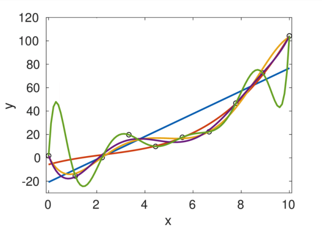




Model-based inference

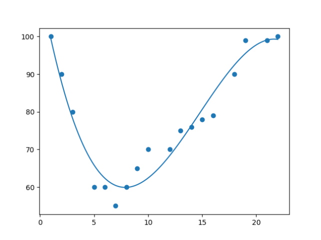
- In model-based inference, most of the focus is on identifying an unknown deterministic model
- Inference can be extrapolated beyond the target population
- Goal is to provide a theoretical framework for why ‘x causes y’

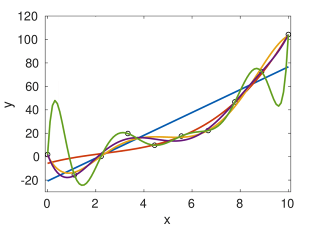




Model-based inference

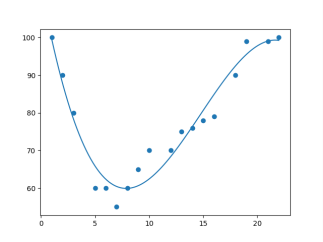
- In model-based inference, most of the focus is on identifying an unknown deterministic model
- Inference can be extrapolated beyond the target population
- Goal is to provide a theoretical framework for why ‘x causes y’
- In model-based inference you make distributional assumptions to make your response a random variable

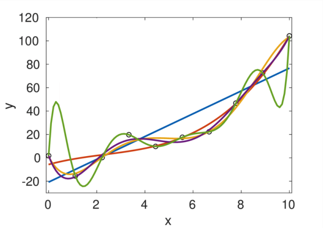




Model-based inference

- In model-based inference, most of the focus is on identifying an unknown deterministic model
- Inference can be extrapolated beyond the target population
- Goal is to provide a theoretical framework for why 'x causes y'
- In model-based inference you make distributional assumptions to make your response a random variable
- Data are typically analysed by fitting a model to data and interpreting the parameter estimates





What is modelling?

Hypothesis vs. model

Hypothesis vs. model

Hypothesis:

Hypothesis vs. model

Hypothesis: An idea, supposition, or otherwise unproven theory used as the basis for further investigation.

Hypothesis vs. model

Hypothesis: An idea, supposition, or otherwise unproven theory used as the basis for further investigation.

Model:

Hypothesis vs. model

Hypothesis: An idea, supposition, or otherwise unproven theory used as the basis for further investigation.

Model: A generalised description of some phenomenon.

Hypothesis vs. model

Hypothesis: An idea, supposition, or otherwise unproven theory used as the basis for further investigation.

Model: A generalised description of some phenomenon.

$$\text{Model} \neq \text{Hypothesis}$$

One hypothesis, many models

One hypothesis, many models

A single hypothesis can be represented by multiple models.

One hypothesis, many models

A single hypothesis can be represented by multiple models.

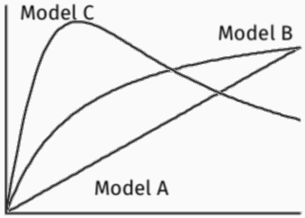
Hypothesis: Body mass M increases with age L

A single hypothesis can be represented by multiple models.

Hypothesis: Body mass M increases with age L

Models:

- $M = aL$ Model A: Body mass is proportional to age
- $M = \frac{AL}{1+bL}$ Model B: Body mass saturates as age increases
- $M = aLe^{-bL}$ Model C: Body mass increases and then decreases as age increases



Source: Hillborn and Mangel 1997

General points about modelling

General points about modelling

The equation of a model is a very specific expression of the hypothesis.

General points about modelling

The equation of a model is a very specific expression of the hypothesis. In other words, models help us clarify verbal descriptions of nature and mechanisms.

General points about modelling

The equation of a model is a very specific expression of the hypothesis. In other words, models help us clarify verbal descriptions of nature and mechanisms.

Models help us understand which parameters and processes are important, and which ones are not.

General points about modelling

The equation of a model is a very specific expression of the hypothesis. In other words, models help us clarify verbal descriptions of nature and mechanisms.

Models help us understand which parameters and processes are important, and which ones are not.

No model is completely correct.

No model is completely correct

No model is completely correct

The goal of modelling is not to provide a perfect description of the world

No model is completely correct

The goal of modelling is not to provide a perfect description of the world, but to distill a process down to the most important components

No model is completely correct

The goal of modelling is not to provide a perfect description of the world, but to distill a process down to the most important components

Complicated models with lots of parameters usually provide better fits

No model is completely correct

The goal of modelling is not to provide a perfect description of the world, but to distill a process down to the most important components

Complicated models with lots of parameters usually provide better fits, but if the model is as complicated as nature itself why bother with modelling? Just go for a walk in the woods and be happy.

No model is completely correct

The goal of modelling is not to provide a perfect description of the world, but to distill a process down to the most important components

Complicated models with lots of parameters usually provide better fits, but if the model is as complicated as nature itself why bother with modelling? Just go for a walk in the woods and be happy.

Don't fall in love with a model, the important thing is the system

Model complexity

Model complexity

Complex models provide more numerical precision

Model complexity

Complex models provide more numerical precision, but simple models are more interpretable.

Model complexity

Complex models provide more numerical precision, but simple models are more interpretable.

Simple models risk leaving out important parameters

Model complexity

Complex models provide more numerical precision, but simple models are more interpretable.

Simple models risk leaving out important parameters, complex models need lots of data for good parameter estimation

Model complexity

Complex models provide more numerical precision, but simple models are more interpretable.

Simple models risk leaving out important parameters, complex models need lots of data for good parameter estimation

How complex should a model be?

Model complexity

Complex models provide more numerical precision, but simple models are more interpretable.

Simple models risk leaving out important parameters, complex models need lots of data for good parameter estimation

How complex should a model be?

Short answer: Let the data tell you.

Model complexity

Complex models provide more numerical precision, but simple models are more interpretable.

Simple models risk leaving out important parameters, complex models need lots of data for good parameter estimation

How complex should a model be?

Short answer: Let the data tell you.

Long answer: There are methods for this that we'll cover in later lectures.

Components of a model

Model Components

Model Components

Models are comprised of two main components:

Model Components

Models are comprised of two main components:

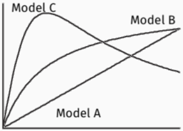
Deterministic part: Describes the shape of the relationship (i.e., your hypothesis).

Model Components

Models are comprised of two main components:

Deterministic part: Describes the shape of the relationship (i.e., your hypothesis).

- Model A: $M = aL$
- Model B: $M = \frac{AL}{1+bL}$
- Model C: $M = aLe^{-bL}$

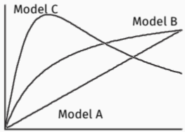


Model Components

Models are comprised of two main components:

Deterministic part: Describes the shape of the relationship (i.e., your hypothesis).

- Model A: $M = aL$
- Model B: $M = \frac{AL}{1+bL}$
- Model C: $M = aLe^{-bL}$



Stochastic part: Describes the randomness of the process (i.e., captures the noise in a system).

Deterministic vs. Stochastic

Deterministic models

Deterministic models

- No components are uncertain

Deterministic vs. Stochastic

Deterministic models

- No components are uncertain

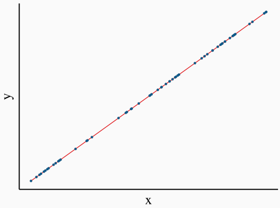
- Outcome is always the same

Deterministic models

- No components are uncertain

- Outcome is always the same

- $y_i = \beta_0 + \beta_1 x_i$



Deterministic vs. Stochastic

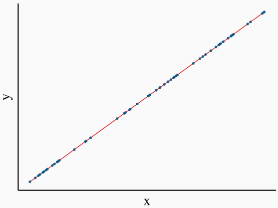
THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

Deterministic models

- No components are uncertain

- Outcome is always the same

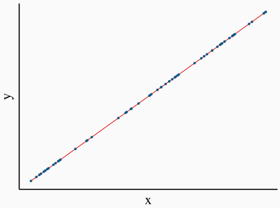
- $y_i = \beta_0 + \beta_1 x_i$



Stochastic models

Deterministic models

- No components are uncertain
- Outcome is always the same
- $y_i = \beta_0 + \beta_1 x_i$

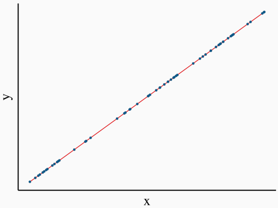


Stochastic models

- Some components are uncertain and characterised by probability distributions

Deterministic models

- No components are uncertain
- Outcome is always the same
- $y_i = \beta_0 + \beta_1 x_i$

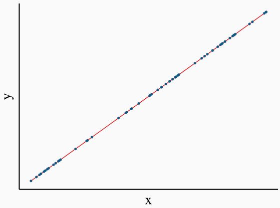


Stochastic models

- Some components are uncertain and characterised by probability distributions
- Outcome is variable

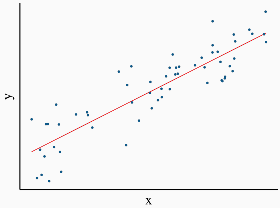
Deterministic models

- No components are uncertain
- Outcome is always the same
- $y_i = \beta_0 + \beta_1 x_i$



Stochastic models

- Some components are uncertain and characterised by probability distributions
- Outcome is variable
- $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$



References

Platt, J.R. (1964). Strong inference. *science*, 146, 347–353.