Quantitative Methods

Human Sciences, 2020–21

Elias Nosrati

Lecture 6: 19 November 2020

Today

Today

- Statistical models and the problem of inference.

Today

- Statistical models and the problem of inference.
- Introduction to statistical inference.

Today

- Statistical models and the problem of inference.
- Introduction to statistical inference.
- Introduction to linear regression models.

Statistical models

Statistical models

- A model is an abstraction of reality.

Statistical models

- A model is an abstraction of reality.
- Judged by its usefulness, in the context of a particular purpose.

Statistical models

- A model is an abstraction of reality.
- Judged by its usefulness, in the context of a particular purpose.
- Model might not be "true" or entirely realistic, but does it provide us with useful information about some quantity of interest?

Data generating process

You are given a data set called `lifexp`:

```r
head(lifexp)

## [1] 69 74 78 73 70 75

tail(lifexp)

## [1] 73 65 68 69 80 72

summary(lifexp)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   59.00   69.00   72.00   71.93   74.00   84.00
```

Data generating process

You are given a data set called `lifexp`:

```
head(lifexp)

## [1] 69 74 78 73 70 75

tail(lifexp)

## [1] 73 65 68 69 80 72

summary(lifexp)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   59.00   69.00   72.00   71.93   74.00   84.00
```

What is the underlying data generating process? What are the key features of the underlying distribution?

Constructing a model

Constructing a model

- Any model has to make some assumptions.

Constructing a model

- Any model has to make some assumptions.
- In the `lifexp` example, assume Normal distribution:

$$Y \sim \mathcal{N}(\mu, \sigma^2).$$

Constructing a model

- Any model has to make some assumptions.
- In the `lifexp` example, assume Normal distribution:

$$Y \sim \mathcal{N}(\mu, \sigma^2).$$

- Quantity of interest:

Constructing a model

- Any model has to make some assumptions.
- In the `lifexp` example, assume Normal distribution:

$$Y \sim \mathcal{N}(\mu, \sigma^2).$$

- Quantity of interest: $\mu = \mathbb{E}(Y)$.

Constructing a model

- Any model has to make some assumptions.
- In the `lifexp` example, assume Normal distribution:

$$Y \sim \mathcal{N}(\mu, \sigma^2).$$

- Quantity of interest: $\mu = \mathbb{E}(Y)$.
- Decompose the model into two components: $M = \{A, \mu\}$.

Constructing a model

- Any model has to make some assumptions.
- In the `lifexp` example, assume Normal distribution:

$$Y \sim \mathcal{N}(\mu, \sigma^2).$$

- Quantity of interest: $\mu = \mathbb{E}(Y)$.
- Decompose the model into two components: $M = \{A, \mu\}$.
- A is the *assumed* part (the data generating process is Normal) and μ is to be *estimated* using the data we have.

Constructing a model

- Any model has to make some assumptions.
- In the `lifexp` example, assume Normal distribution:

$$Y \sim \mathcal{N}(\mu, \sigma^2).$$

- Quantity of interest: $\mu = \mathbb{E}(Y)$.
- Decompose the model into two components: $M = \{A, \mu\}$.
- A is the *assumed* part (the data generating process is Normal) and μ is to be *estimated* using the data we have.
- But how do we estimate μ?

Inverse probability and the problem of inference

Inverse probability and the problem of inference

- *Inference* is the process of using facts we have (the data) to learn about facts we do not have (the underlying data generating process): involves uncertainty.

Inverse probability and the problem of inference

- *Inference* is the process of using facts we have (the data) to learn about facts we do not have (the underlying data generating process): involves uncertainty.
- Probability as a model of uncertainty.

Inverse probability and the problem of inference

- *Inference* is the process of using facts we have (the data) to learn about facts we do not have (the underlying data generating process): involves uncertainty.
- Probability as a model of uncertainty.
- Probabilities are traditionally expressed as $\mathbb{P}(\text{data} \mid \text{model})$.

Inverse probability and the problem of inference

- *Inference* is the process of using facts we have (the data) to learn about facts we do not have (the underlying data generating process): involves uncertainty.
- Probability as a model of uncertainty.
- Probabilities are traditionally expressed as $\mathbb{P}(\text{data} \mid \text{model})$.
- For the purposes of inference, we are interested in the *inverse probability* $\mathbb{P}(\text{model} \mid \text{data})$.

Inverse probability and the problem of inference

- *Inference* is the process of using facts we have (the data) to learn about facts we do not have (the underlying data generating process): involves uncertainty.
- Probability as a model of uncertainty.
- Probabilities are traditionally expressed as $\mathbb{P}(\text{data} \mid \text{model})$.
- For the purposes of inference, we are interested in the *inverse probability* $\mathbb{P}(\text{model} \mid \text{data})$.
- Uncertainty lies with the model (the process that generates the data) rather than the (already observed) data.

Inverse probability and the problem of inference

- *Inference* is the process of using facts we have (the data) to learn about facts we do not have (the underlying data generating process): involves uncertainty.
- Probability as a model of uncertainty.
- Probabilities are traditionally expressed as $\mathbb{P}(\text{data} \mid \text{model})$.
- For the purposes of inference, we are interested in the *inverse probability* $\mathbb{P}(\text{model} \mid \text{data})$.
- Uncertainty lies with the model (the process that generates the data) rather than the (already observed) data.
- How to go from $\mathbb{P}(\text{data} \mid \text{model})$ to $\mathbb{P}(\text{model} \mid \text{data})$?

Inverse probability and the problem of inference

- *Inference* is the process of using facts we have (the data) to learn about facts we do not have (the underlying data generating process): involves uncertainty.
- Probability as a model of uncertainty.
- Probabilities are traditionally expressed as $\mathbb{P}(\text{data} \mid \text{model})$.
- For the purposes of inference, we are interested in the *inverse probability* $\mathbb{P}(\text{model} \mid \text{data})$.
- Uncertainty lies with the model (the process that generates the data) rather than the (already observed) data.
- How to go from $\mathbb{P}(\text{data} \mid \text{model})$ to $\mathbb{P}(\text{model} \mid \text{data})$?
- How to go from $\mathbb{P}(\text{known} \mid \text{unknown})$ to $\mathbb{P}(\text{unknown} \mid \text{known})$?

Inverse probability and the problem of inference

- *Inference* is the process of using facts we have (the data) to learn about facts we do not have (the underlying data generating process): involves uncertainty.
- Probability as a model of uncertainty.
- Probabilities are traditionally expressed as $\mathbb{P}(\text{data} \mid \text{model})$.
- For the purposes of inference, we are interested in the *inverse probability* $\mathbb{P}(\text{model} \mid \text{data})$.
- Uncertainty lies with the model (the process that generates the data) rather than the (already observed) data.
- How to go from $\mathbb{P}(\text{data} \mid \text{model})$ to $\mathbb{P}(\text{model} \mid \text{data})$?
- How to go from $\mathbb{P}(\text{known} \mid \text{unknown})$ to $\mathbb{P}(\text{unknown} \mid \text{known})$?
- How to go from $\mathbb{P}(\texttt{lifexp} \mid \mu)$ to $\mathbb{P}(\mu \mid \texttt{lifexp})$?

Building a theory of inference

Building a theory of inference

- Let y be the realisation of Y.

Building a theory of inference

- Let y be the realisation of Y.
- Let θ be any parameter of interest.

Building a theory of inference

- Let y be the realisation of Y.
- Let θ be any parameter of interest.
- Inverse probability — use Bayes' Rule:

Building a theory of inference

- Let y be the realisation of Y.
- Let θ be any parameter of interest.
- Inverse probability — use Bayes' Rule:

$$\mathbb{P}(\theta \mid y) = \frac{\mathbb{P}(\theta, y)}{\mathbb{P}(y)}$$

Building a theory of inference

- Let y be the realisation of Y.
- Let θ be any parameter of interest.
- Inverse probability — use Bayes' Rule:

$$\mathbb{P}(\theta \mid y) = \frac{\mathbb{P}(\theta, y)}{\mathbb{P}(y)} = \frac{\mathbb{P}(\theta)\mathbb{P}(y \mid \theta)}{\mathbb{P}(y)}.$$

Building a theory of inference

- Let y be the realisation of Y.
- Let θ be any parameter of interest.
- Inverse probability — use Bayes' Rule:

$$\mathbb{P}(\theta \mid y) = \frac{\mathbb{P}(\theta, y)}{\mathbb{P}(y)} = \frac{\mathbb{P}(\theta)\mathbb{P}(y \mid \theta)}{\mathbb{P}(y)}.$$

- Let $\mathbb{P}(\theta)/\mathbb{P}(y)$ be an unknown constant function $k(y)$.

Building a theory of inference

- Let y be the realisation of Y.
- Let θ be any parameter of interest.
- Inverse probability — use Bayes' Rule:

$$\mathbb{P}(\theta \mid y) = \frac{\mathbb{P}(\theta, y)}{\mathbb{P}(y)} = \frac{\mathbb{P}(\theta)\mathbb{P}(y \mid \theta)}{\mathbb{P}(y)}.$$

- Let $\mathbb{P}(\theta)/\mathbb{P}(y)$ be an unknown constant function $k(y)$.
- Define the *likelihood* function:

$$L(\theta \mid y) = k(y)\mathbb{P}(y \mid \theta).$$

Building a theory of inference

- Let y be the realisation of Y.
- Let θ be any parameter of interest.
- Inverse probability — use Bayes' Rule:

$$\mathbb{P}(\theta \mid y) = \frac{\mathbb{P}(\theta, y)}{\mathbb{P}(y)} = \frac{\mathbb{P}(\theta)\mathbb{P}(y \mid \theta)}{\mathbb{P}(y)}.$$

- Let $\mathbb{P}(\theta)/\mathbb{P}(y)$ be an unknown constant function $k(y)$.
- Define the *likelihood* function:

$$L(\theta \mid y) = k(y)\mathbb{P}(y \mid \theta).$$

- The numerical value of L is interpretable as the *likelihood* of a model having generated the observed data.

Building a theory of inference

- Let y be the realisation of Y.
- Let θ be any parameter of interest.
- Inverse probability — use Bayes' Rule:

$$\mathbb{P}(\theta \mid y) = \frac{\mathbb{P}(\theta, y)}{\mathbb{P}(y)} = \frac{\mathbb{P}(\theta)\mathbb{P}(y \mid \theta)}{\mathbb{P}(y)}.$$

- Let $\mathbb{P}(\theta)/\mathbb{P}(y)$ be an unknown constant function $k(y)$.
- Define the *likelihood* function:

$$L(\theta \mid y) = k(y)\mathbb{P}(y \mid \theta).$$

- The numerical value of L is interpretable as the *likelihood* of a model having generated the observed data.
- This numerical value is not equal to, but is *proportional* to a probability density:

$$L(\theta \mid y) \propto \mathbb{P}(y \mid \theta).$$

Building a theory of inference

- Let y be the realisation of Y.
- Let θ be any parameter of interest.
- Inverse probability — use Bayes' Rule:

$$\mathbb{P}(\theta \mid y) = \frac{\mathbb{P}(\theta, y)}{\mathbb{P}(y)} = \frac{\mathbb{P}(\theta)\mathbb{P}(y \mid \theta)}{\mathbb{P}(y)}.$$

- Let $\mathbb{P}(\theta)/\mathbb{P}(y)$ be an unknown constant function $k(y)$.
- Define the *likelihood* function:

$$L(\theta \mid y) = k(y)\mathbb{P}(y \mid \theta).$$

- The numerical value of L is interpretable as the *likelihood* of a model having generated the observed data.
- This numerical value is not equal to, but is *proportional* to a probability density:

$$L(\theta \mid y) \propto \mathbb{P}(y \mid \theta).$$

- A *relative* (not absolute) measure of uncertainty.

Likelihood theory of inference

Likelihood theory of inference

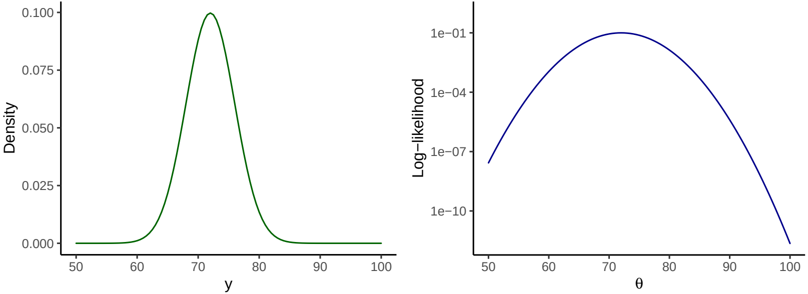
- A normal probability density $f(y \mid \theta)$ holds θ fixed and is a function of the data, defined on the *data scale*.

Likelihood theory of inference

- A normal probability density $f(y \mid \theta)$ holds θ fixed and is a function of the data, defined on the *data scale*.
- The likelihood function holds the data fixed and is a function of θ, defined on the *parameter scale*.

Likelihood theory of inference

- A normal probability density $f(y \mid \theta)$ holds θ fixed and is a function of the data, defined on the *data scale*.
- The likelihood function holds the data fixed and is a function of θ, defined on the *parameter scale*.



Likelihood theory of inference (cont.)

Likelihood theory of inference (cont.)

- For mathematical and computational reason, it is common to use the *log*-likelihood function.

Likelihood theory of inference (cont.)

- For mathematical and computational reason, it is common to use the *log*-likelihood function.
- Key question: given what we observe, which model is most likely to have generated our data?

Likelihood theory of inference (cont.)

- For mathematical and computational reason, it is common to use the *log*-likelihood function.
- Key question: given what we observe, which model is most likely to have generated our data?
- The most likely value of θ is where L is at its maximum.

Likelihood theory of inference (cont.)

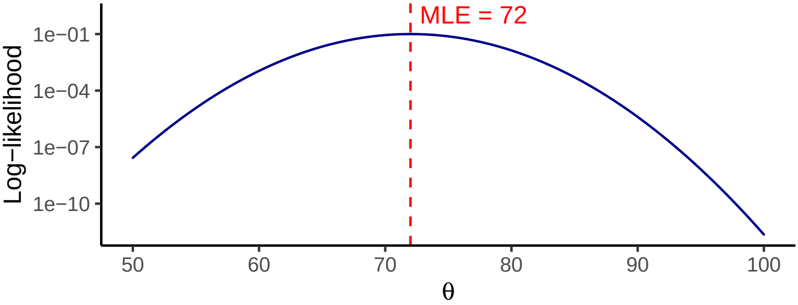
- For mathematical and computational reason, it is common to use the *log*-likelihood function.
- Key question: given what we observe, which model is most likely to have generated our data?
- The most likely value of θ is where L is at its maximum.
- Maximum likelihood estimation: *maximise* the likelihood.

Likelihood theory of inference (cont.)

- For mathematical and computational reason, it is common to use the *log*-likelihood function.
- Key question: given what we observe, which model is most likely to have generated our data?
- The most likely value of θ is where L is at its maximum.
- Maximum likelihood estimation: *maximise* the likelihood.
- The result: a maximum likelihood estimate (MLE) $\hat{\theta}$.

Likelihood theory of inference (cont.)

- For mathematical and computational reason, it is common to use the *log*-likelihood function.
- Key question: given what we observe, which model is most likely to have generated our data?
- The most likely value of θ is where L is at its maximum.
- Maximum likelihood estimation: *maximise* the likelihood.
- The result: a maximum likelihood estimate (MLE) $\hat{\theta}$.



Bayesian theory of inference

Bayesian theory of inference

- Back to the inverse probability:

$$\mathbb{P}(\theta \mid y) \propto \mathbb{P}(\theta)\mathbb{P}(y \mid \theta).$$

Bayesian theory of inference

- Back to the inverse probability:

$$\mathbb{P}(\theta \mid y) \propto \mathbb{P}(\theta)\mathbb{P}(y \mid \theta).$$

- $\mathbb{P}(\theta \mid y)$: posterior density.

Bayesian theory of inference

- Back to the inverse probability:

$$\mathbb{P}(\theta \mid y) \propto \mathbb{P}(\theta)\mathbb{P}(y \mid \theta).$$

- $\mathbb{P}(\theta \mid y)$: posterior density.
- $\mathbb{P}(y \mid \theta)$: traditional probability ($\propto$ likelihood).

Bayesian theory of inference

- Back to the inverse probability:

$$\mathbb{P}(\theta \mid y) \propto \mathbb{P}(\theta)\mathbb{P}(y \mid \theta).$$

- $\mathbb{P}(\theta \mid y)$: posterior density.
- $\mathbb{P}(y \mid \theta)$: traditional probability ($\propto$ likelihood).
- $\mathbb{P}(\theta)$: prior density.

Bayesian theory of inference

- Back to the inverse probability:

$$\mathbb{P}(\theta \mid y) \propto \mathbb{P}(\theta)\mathbb{P}(y \mid \theta).$$

- $\mathbb{P}(\theta \mid y)$: posterior density.
- $\mathbb{P}(y \mid \theta)$: traditional probability ($\propto$ likelihood).
- $\mathbb{P}(\theta)$: prior density.
- Bayesian inference differs from likelihood inference in its use of *prior* information about θ:

$$\text{posterior} \propto \text{prior} \times \text{likelihood}.$$

Bayesian theory of inference

- Back to the inverse probability:

$$\mathbb{P}(\theta \mid y) \propto \mathbb{P}(\theta)\mathbb{P}(y \mid \theta).$$

- $\mathbb{P}(\theta \mid y)$: posterior density.
- $\mathbb{P}(y \mid \theta)$: traditional probability ($\propto$ likelihood).
- $\mathbb{P}(\theta)$: prior density.
- Bayesian inference differs from likelihood inference in its use of *prior* information about θ:

$$\text{posterior} \propto \text{prior} \times \text{likelihood}.$$

- A way of integrating evidence from outside the data into the model.

Bayesian theory of inference

- Back to the inverse probability:

$$\mathbb{P}(\theta \mid y) \propto \mathbb{P}(\theta)\mathbb{P}(y \mid \theta).$$

- $\mathbb{P}(\theta \mid y)$: posterior density.
- $\mathbb{P}(y \mid \theta)$: traditional probability ($\propto$ likelihood).
- $\mathbb{P}(\theta)$: prior density.
- Bayesian inference differs from likelihood inference in its use of *prior* information about θ:

$$\text{posterior} \propto \text{prior} \times \text{likelihood}.$$

- A way of integrating evidence from outside the data into the model.
- Practical difference: minor unless prior is actually important.

Election example

Election example

- Candidate A and candidate B.

Election example

- Candidate A and candidate B.
- Outcome of experiment is Bernoulli random variable Y with parameter $p = \mathbb{P}(A \text{ wins})$.

Election example

- Candidate A and candidate B.
- Outcome of experiment is Bernoulli random variable Y with parameter $p = \mathbb{P}(A \text{ wins})$.
- We use polling data calculate posterior density $\mathbb{P}(p \mid y)$.

Election example

- Candidate A and candidate B.
- Outcome of experiment is Bernoulli random variable Y with parameter $p = \mathbb{P}(A \text{ wins})$.
- We use polling data calculate posterior density $\mathbb{P}(p \mid y)$.
- Likelihood: $L(p \mid y) \propto \mathbb{P}(y \mid p)$.

Election example

- Candidate A and candidate B.
- Outcome of experiment is Bernoulli random variable Y with parameter $p = \mathbb{P}(A \text{ wins})$.
- We use polling data calculate posterior density $\mathbb{P}(p \mid y)$.
- Likelihood: $L(p \mid y) \propto \mathbb{P}(y \mid p)$.
- Bayesian: use prior information about p (e.g., previous poll data or evidence of rigged election).

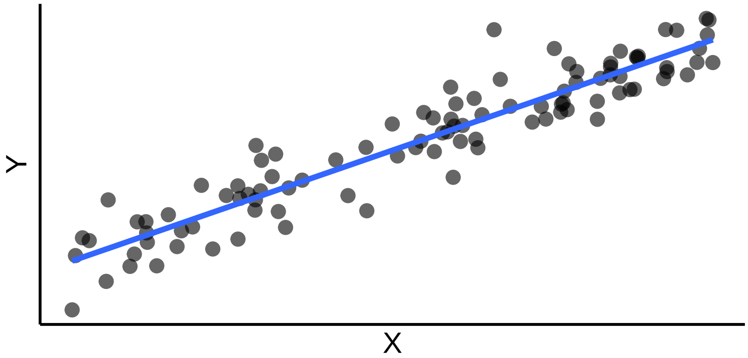
Election example

- Candidate A and candidate B.
- Outcome of experiment is Bernoulli random variable Y with parameter $p = \mathbb{P}(A \text{ wins})$.
- We use polling data calculate posterior density $\mathbb{P}(p \mid y)$.
- Likelihood: $L(p \mid y) \propto \mathbb{P}(y \mid p)$.
- Bayesian: use prior information about p (e.g., previous poll data or evidence of rigged election).
- Challenge: what is the exact value of the prior?

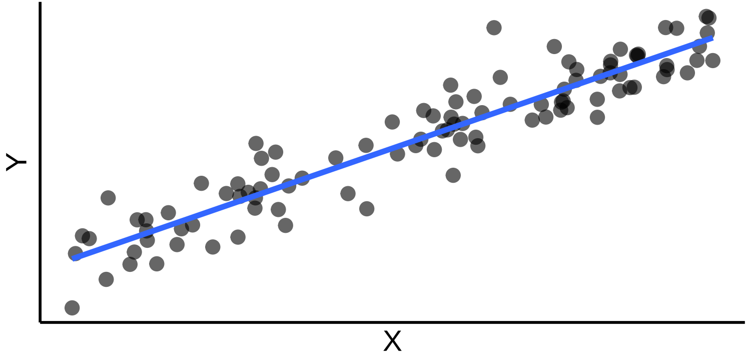
Election example

- Candidate A and candidate B.
- Outcome of experiment is Bernoulli random variable Y with parameter $p = \mathbb{P}(A \text{ wins})$.
- We use polling data calculate posterior density $\mathbb{P}(p \mid y)$.
- Likelihood: $L(p \mid y) \propto \mathbb{P}(y \mid p)$.
- Bayesian: use prior information about p (e.g., previous poll data or evidence of rigged election).
- Challenge: what is the exact value of the prior?
- Typical solution: try a range of possible values.

A non-theory of inference: ordinary least squares

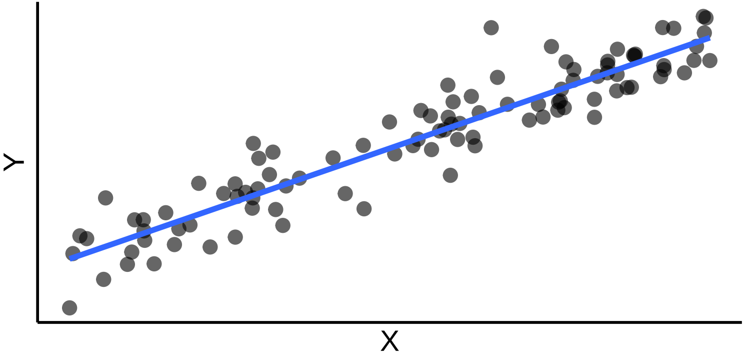


A non-theory of inference: ordinary least squares



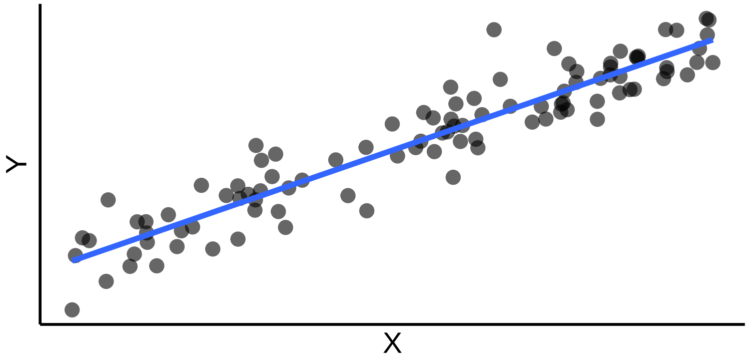
- A common approach: minimise ("*least*") the squared deviations ("*squares*") of the observed data from the estimated parameter value.

A non-theory of inference: ordinary least squares



- A common approach: minimise ("*least*") the squared deviations ("*squares*") of the observed data from the estimated parameter value.
- Classical example: how to fit a line to a cloud of points.

A non-theory of inference: ordinary least squares



- A common approach: minimise ("*least*") the squared deviations ("*squares*") of the observed data from the estimated parameter value.
- Classical example: how to fit a line to a cloud of points.
- The corresponding parameter estimate (slope of line) is known as the *ordinary least squares* (OLS) estimate.

A non-theory of inference: ordinary least squares (cont.)

A non-theory of inference: ordinary least squares (cont.)

- Not a theory of inference . . .

A non-theory of inference: ordinary least squares (cont.)

- Not a theory of inference ...
- ... but, under certain assumptions, coincides with maximum likelihood.

A non-theory of inference: ordinary least squares (cont.)

- Not a theory of inference ...
- ... but, under certain assumptions, coincides with maximum likelihood.
- Bonus: easy to implement.

Regression model notation

Regression model notation

- Notation: $Y \sim \mathcal{N}(y \mid \mu, \sigma^2)$.

Regression model notation

- Notation: $Y \sim \mathcal{N}(y \mid \mu, \sigma^2)$.
- Equivalently: $y = \mu + \epsilon$.

Regression model notation

- Notation: $Y \sim \mathcal{N}(y \mid \mu, \sigma^2)$.
- Equivalently: $y = \mu + \epsilon$.
- y: outcome (dependent) variable.

Regression model notation

- Notation: $Y \sim \mathcal{N}(y \mid \mu, \sigma^2)$.
- Equivalently: $y = \mu + \epsilon$.
- y: outcome (dependent) variable.
- μ: intercept (mean).

Regression model notation

- Notation: $Y \sim \mathcal{N}(y \mid \mu, \sigma^2)$.
- Equivalently: $y = \mu + \epsilon$.
- y: outcome (dependent) variable.
- μ: intercept (mean).
- ϵ: error (disturbance) term, with $\epsilon \sim \mathcal{N}(0, \sigma^2)$.

OLS regression in **R**

```
lm(lifexp ~ 1)
```

OLS regression in R

```r
lm(lifexp ~ 1)
```

```
##
## Call:
## lm(formula = lifexp ~ 1)
##
## Coefficients:
## (Intercept)
##       71.93
```

OLS regression in R

```r
lm(lifexp ~ 1)
```

```
##
## Call:
## lm(formula = lifexp ~ 1)
##
## Coefficients:
## (Intercept)
##       71.93
```

```r
mean(lifexp)
```

```
## [1] 71.932
```