Statistical Frameworks

Today's agenda:
Discuss statistical frameworks,
Discuss the modelling process,
Start working with code

What are the two main modelling frameworks?

What are the two main modelling frameworks?

Frequentist vs Bayesian

What are the two main modelling frameworks?

Frequentist vs Bayesian

How do they differ?

Frequentist vs Bayesian stats

Aspect	Frequentist Statistics	Bayesian Statistics
Interpretation of probability	Long-run frequency of events	Degree of belief / uncertainty
Parameters	Fixed but unknown	Random variables with distributions
Data	Random (due to sampling)	Fixed, updates beliefs about parameters
Prior information	Not used	Required (can be informative or weak/noninformative)
Main output	Point estimates, confidence intervals, p-values	Posterior distributions, credible intervals, Bayes factors
Confidence vs. credibility	95% CI: in repeated samples, 95% of intervals contain the true value	95% credible interval: given data + prior, 95% chance parameter lies in the interval
Inference process	Hypothesis tests, maximum likelihood, resampling	Bayes' theorem: Prior × Likelihood → Posterior
Computation	Often analytic formulas, asymptotics, bootstraps	Often simulation (e.g., MCMC, variational inference)
Common critique	Ignores prior knowledge, interpretation of p-values can be misleading	Can be subjective (depends on choice of prior), computationally intensive
Example question	"If we repeated this study many times, how often would we see results like this?"	"Given the data, what is the probability the hypothesis is true?"

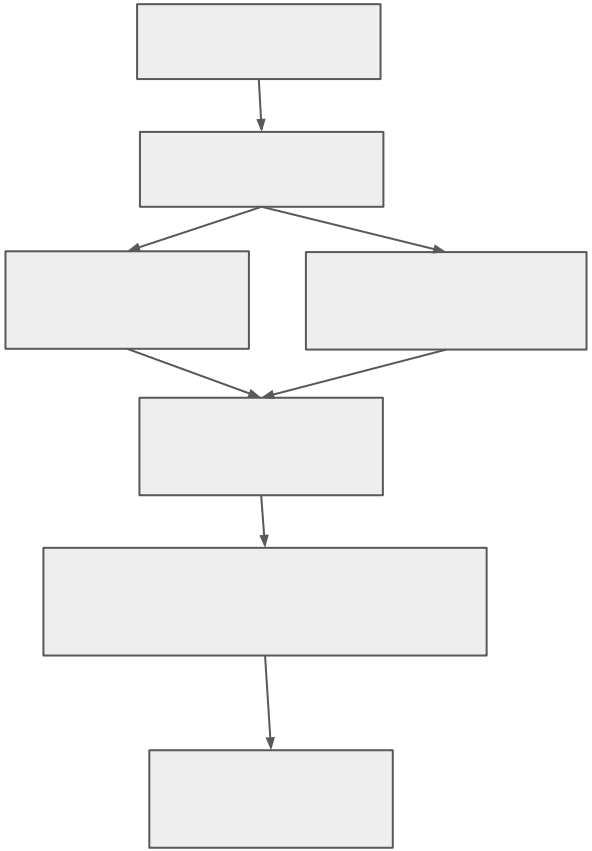
Frequentist vs Bayesian stats

Aspect	Frequentist Statistics	Bayesian Statistics
Interpretation of probability	Long-run frequency of events	Degree of belief / uncertainty
Parameters	Fixed but unknown	Random variables with distributions
Data	Random (due to sampling)	Fixed, updates beliefs about parameters
Prior information	Not used	Required (can be informative or weak/noninformative)
Main output	Point estimates, confidence intervals, p-values	Posterior distributions, credible intervals, Bayes factors
Confidence vs. credibility	95% CI: in repeated samples, 95% of intervals contain the true value	95% credible interval: given data + prior, 95% chance parameter lies in the interval
Inference process	Hypothesis tests, maximum likelihood, resampling	Bayes' theorem: Prior × Likelihood → Posterior
Computation	Often analytic formulas, asymptotics, bootstraps	Often simulation (e.g., MCMC, variational inference)
Common critique	Ignores prior knowledge, interpretation of p-values can be misleading	Can be subjective (depends on choice of prior), computationally intensive
Example question	"If we repeated this study many times, how often would we see results like this?"	"Given the data, what is the probability the hypothesis is true?"

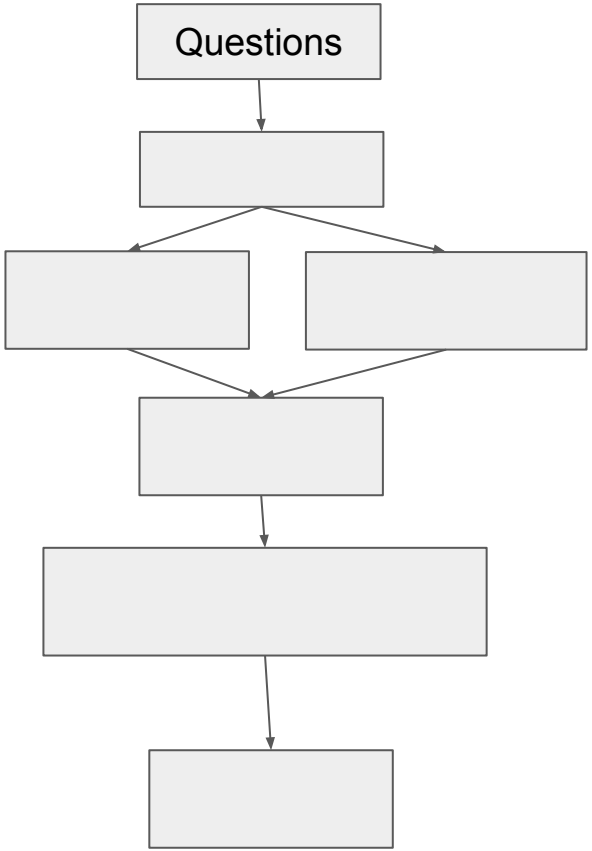
Questions so far?

(Note: this stuff will become more clear as we go)

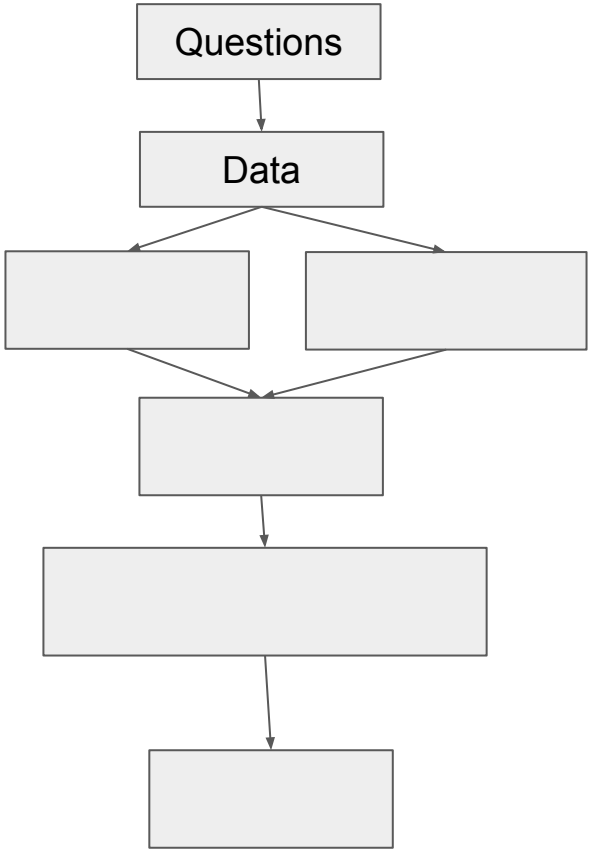
What is the modelling process?



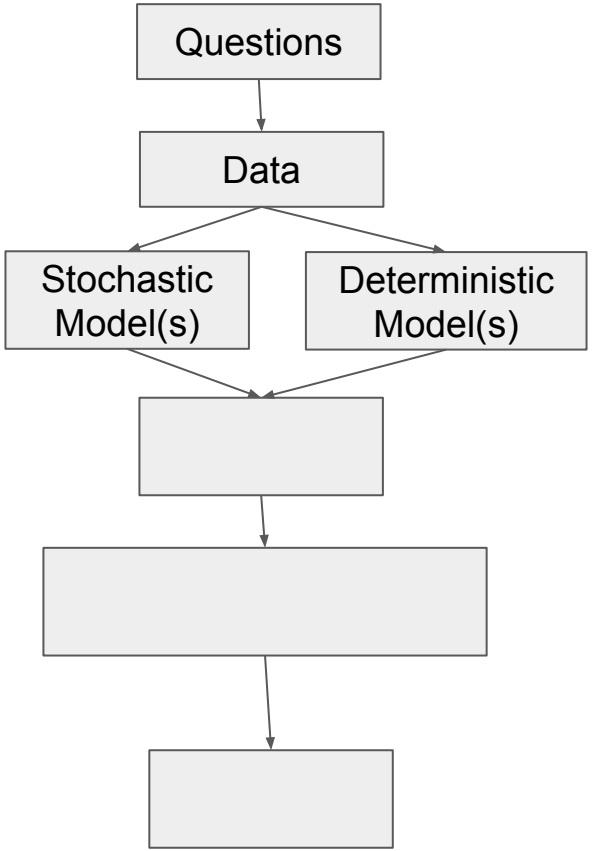
What is the modelling process?



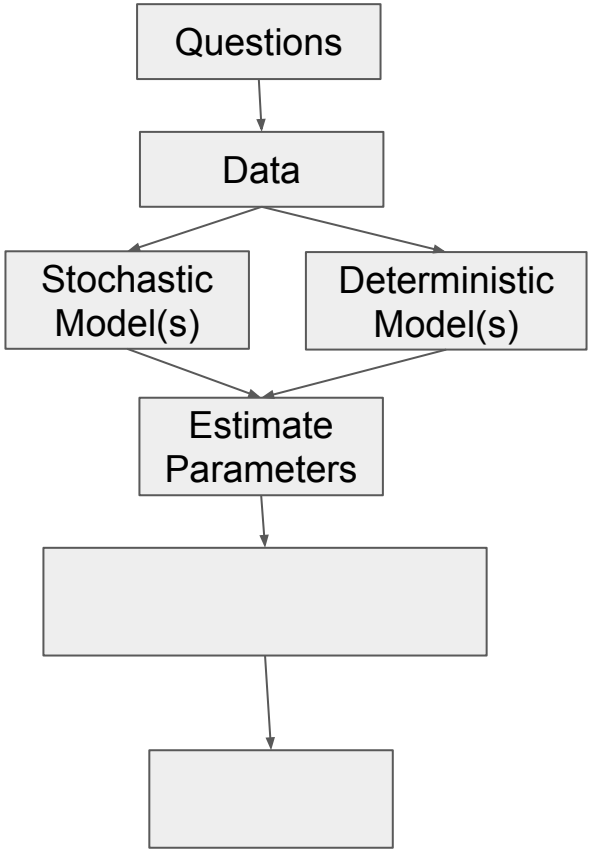
What is the modelling process?



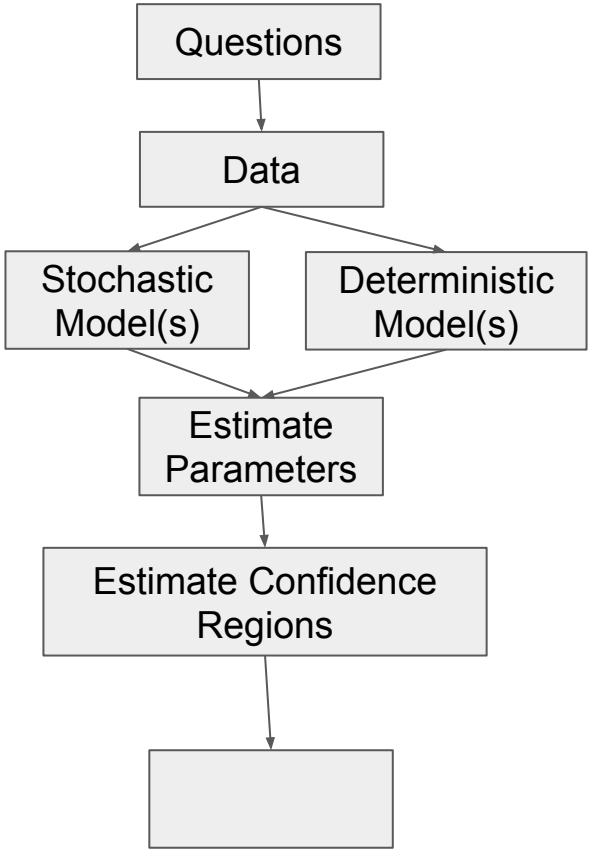
What is the modelling process?



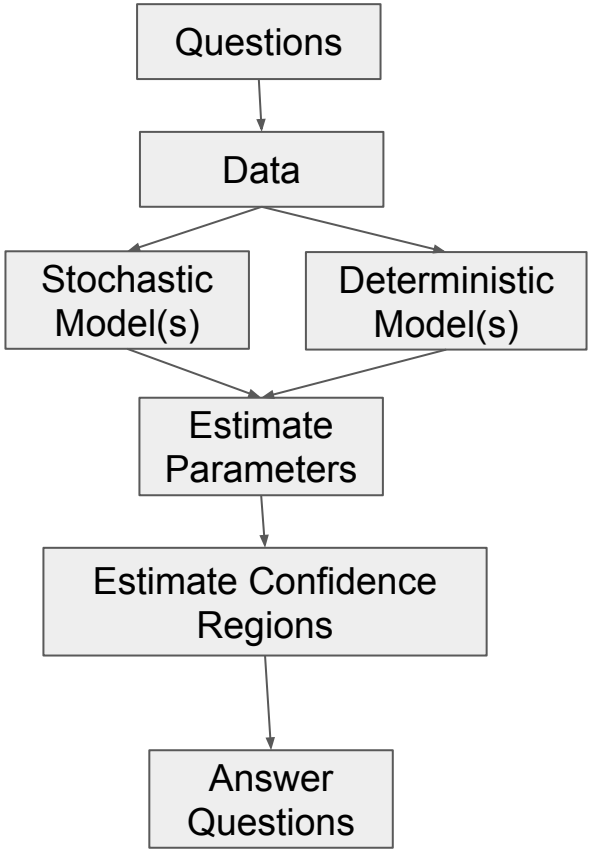
What is the modelling process?



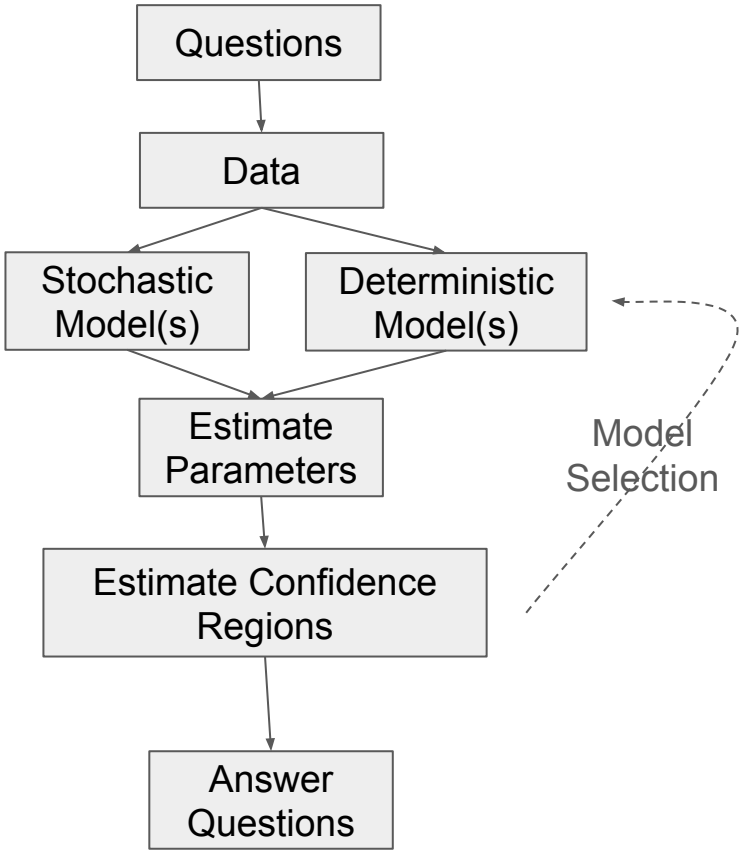
What is the modelling process?



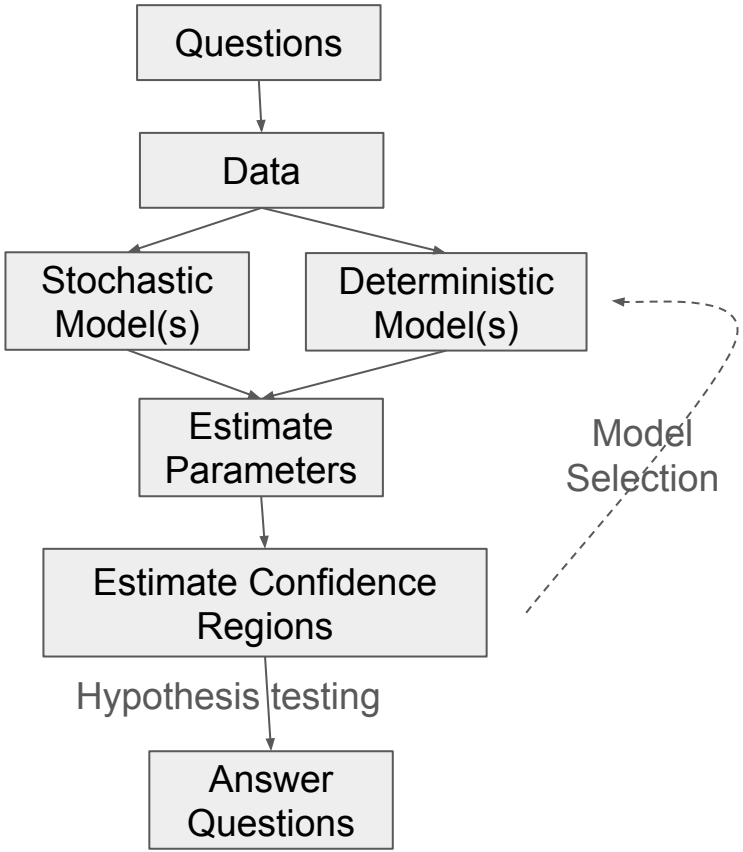
What is the modelling process?



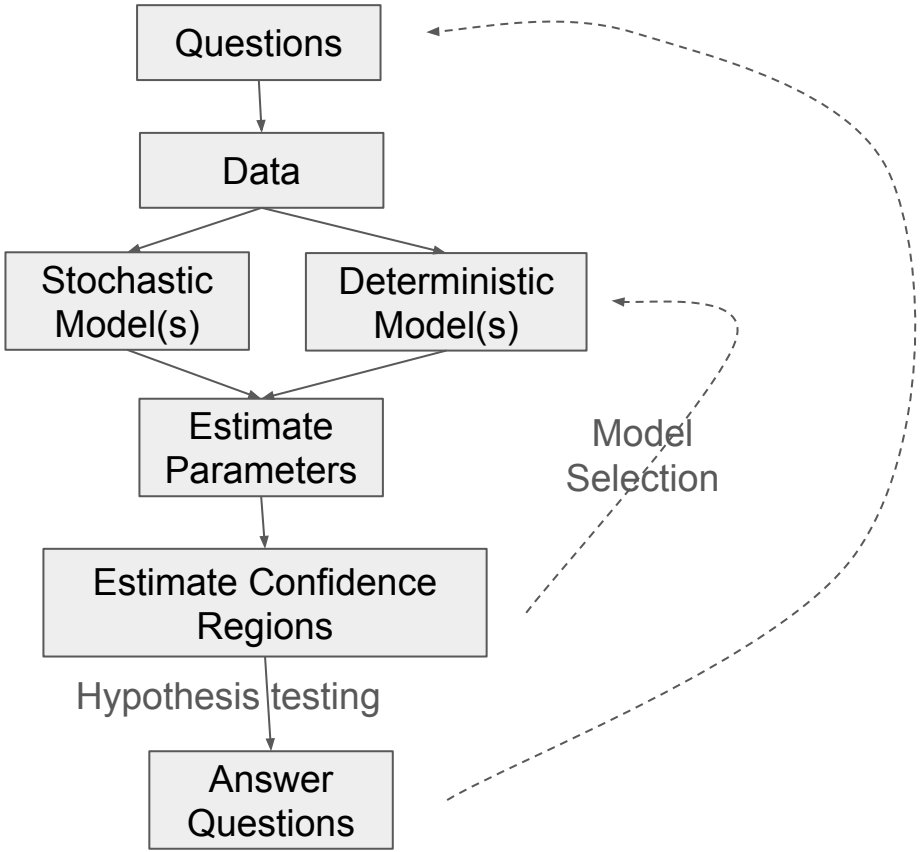
What is the modelling process?



What is the modelling process?



What is the modelling process?



Questions so far?

(Again, this will become more clear as we go along)

Our goal for today is starting to play with R

Work through the examples in Section 1.7 of the book

- Ask questions as you go!
- You'll probably see slight differences with R vs RStudio
- Recommend making an .R file in the project you created with Github

Next time:

Before class:

- read 2.1 - 2.3
- Think about types of data you might want to look at

During class:

- Discuss 2.1 - 2.3
- Work on loading in data