Statistical Inference

R Brown Bag Series #5

Gary R Seamans

The MITRE Corporation

April 5, 2018

Agenda

- Overview
- Discussion
- Questions

Overview

Statistical inference can be defined: *"...as the process of generating conclusions about a population from a noisy sample."* (Caffo 2016)

Concerns

- Representative sample
- Contamination
- Bias
- Randomness

Concerns Example

The China Study is a well known study that claims, among other things, that animal protein is the primary cause of many illnesses, including cancer and heart disease.

Critics claim that the China study exhibits pretty much everything that can done wrong in a study. Without taking sides it is an interesting example. The back and forth between Dr. Campbell and his critics can be illuminating.

Starting Points:

China Study (Campbell, Campbell, and Rudnicki 1968)

Rebuttal (*What Dr. Campbell Won't Tell You About the China Study*)

Goals

- Estimate and Quantify
- Is it a benchmark value?
- Mechanistic relationship
- Policy impact
- Probability of occurrence

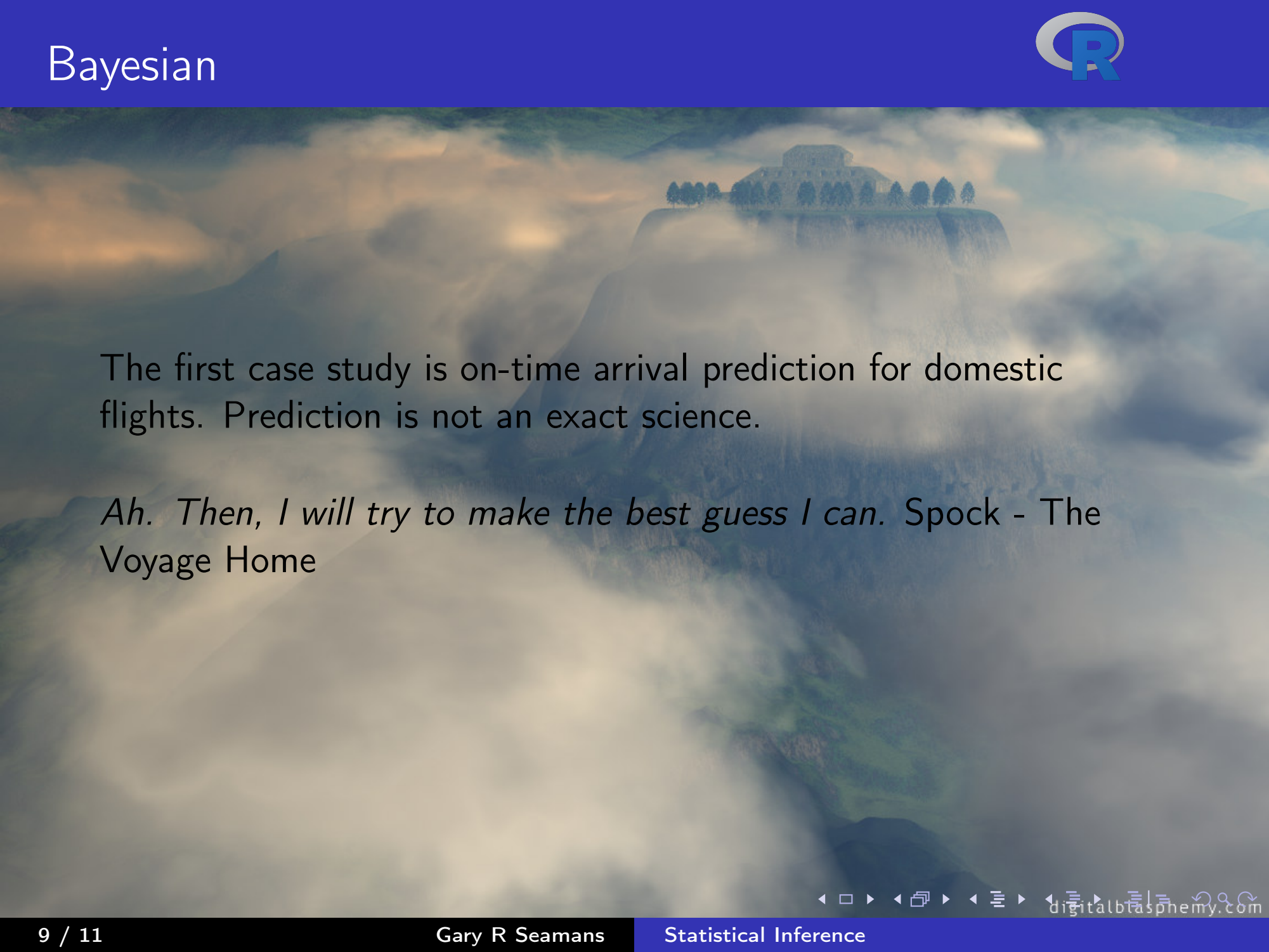
Tools

- Randomization
- Random sampling
- Sampling models
- Hypothesis testing
- Confidence intervals
- Study design
- Nonparametric bootstrapping
- Permutation

Styles

- Frequency probability and inference
- Bayesian probability and inference

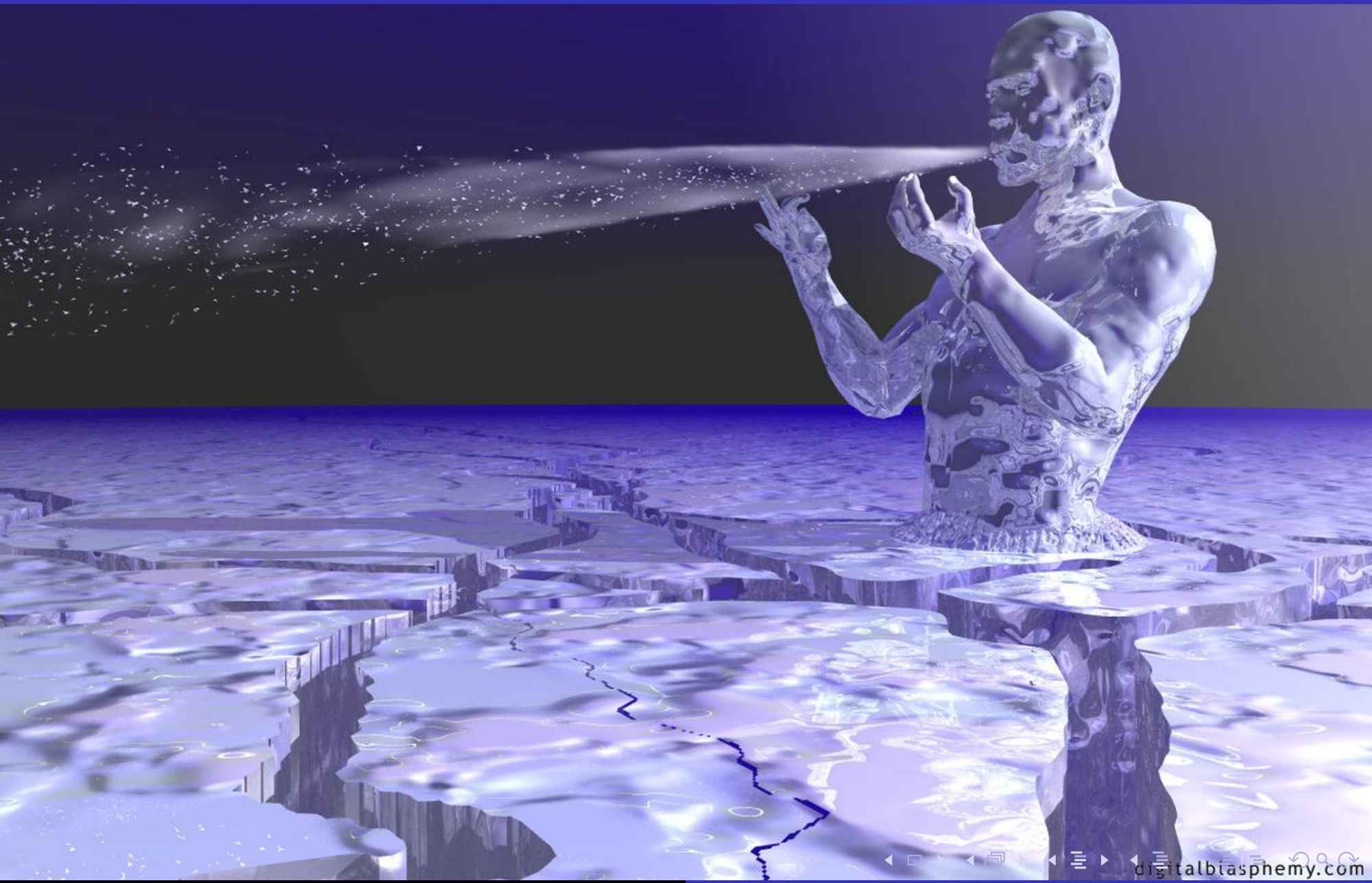
Bayesian



The first case study is on-time arrival prediction for domestic flights. Prediction is not an exact science.

Ah. Then, I will try to make the best guess I can. Spock - The Voyage Home

Questions



References I


Caffo, Brian (May 2016). *Statistical Inference for Data Science*. Leanpub.


Campbell, T Colin, Thomas M Campbell, and Stefan Rudnicki (1968). *The China Study*. Inter-University Consortium for Political Research.


What Dr. Campbell Won't Tell You About the China Study. http://www.cholesterol-and-health.com/China-Study.html.