

The second midterm covers all the course material but with more emphasis on weeks 5-8.

The Course Outline for Weeks 5-8 is

5. **Study Design.** Eddy & Smykla 211-226: The Last Frontier: Understanding the Human Mind. Kahn & Roseman 373-390: Advertising as an Engineering Science.
6. **Models.** Zipfs Law. Gerow, Miquelle & Aramilev 105-118: Monitoring Tiger Prey Abundance in the Russian Far East. Matis & Kiffe 119-134: Predicting the African Bee Invasion. Madigan 135-148: Statistics and the War on Spam. Brookmeyer 197-210: Modeling an Outbreak of Anthrax. De Veaux & Edelstein 307-322: Reducing Junk Mail Using Data Mining Techniques.
7. **Sampling.** Randomized Response Technique. Hill 69-88: Evaluating School Choice Programs. Heilig, Hill & Karon 227-242: Leveraging Chance in HIV Research.
8. **Rare Events.** Lotteries. Average and actual returns. Assessment of Coincidences.

Here is a reminder of the important ideas from those readings and lectures:

Eddy & Smykla 211-226: The Last Frontier: Understanding the Human Mind.

General advice for design of an experiment: “Use repetition, control what you can, and randomize the rest.”

Systematic vs Random Variation
Averaging to reduce Noise

Kahn & Roseman 373-390: Advertising as an Engineering Science.

Experiment Strategy: assign “treatments” to “experimental units”.
Covariate Strategy: adjusting outcomes for nuisance variables.
Bar charts for result display

Zipfs Law.

Use of models in comparisons

Gerow, Miquelle & Aramilev 105-118: Monitoring Tiger Prey Abundance in the Russian Far East.

Using substitute measures for efficient assessment
Logarithms to produce constant percentage variability
Confidence Interval for estimating a population parameter

Matis & Kiffe 119-134: Predicting the African Bee Invasion.

Graphical Display for communicating a complex process
Distance = Rate * Time (They used Time/100 km) Alternative measures
Gamma distribution as a model for right skewed data.
Logistic Growth Curve (S curve of typical population growth)
Birth-Death rate curves leading to equilibrium of the population

Madigan 135-148: Statistics and the War on Spam.

More complex signal detection leads to better sorting of spam
Practical need to simplify

Use of single word contingency tables instead of multiple word ones.
Odds ratio: chance of yes/chance of no.

Brookmeyer 197-210: Modeling an Outbreak of Anthrax.

The observed data was “censored” by an intervention (drugs)

The observed data had the short incubation times

Model needed to help substitute for lack of data

De Veaux & Edelstein 307-322: Reducing Junk Mail Using Data Mining Techniques.

Mass mailing with low response rate: how to select best targets?

Regression models for prediction

Logistic regression for prediction of yes/no response

Trial and Error in model building – and split sample testing.

Randomized Response Technique.

Use of probabilities to protect privacy

Hill 69-88: Evaluating School Choice Programs

Experiments vs Observational Studies

Difficulties with Human Subjects: Missing, noncompliance, ethics

Clinical trials

Double Blind, placebo

Heilig, Hill & Karon 227-242: Leveraging Chance in HIV Research

List Requirement for random sampling

Odds Ratio

Difficulties with Experiments with Human Subjects: Consent

Ethics of Study continuation.

Lotteries

Cash Flow

Randomized Selection for Prizes

Requirements for a fair choice

Coincidences

Why they are common

Birthday Paradox