

Important ideas (likely to be on midterm Mar 16):

Sampling Theory of Means

- And of Proportions

- Square Root Law

- Normal Approximation (Central Limit Theorem)

Insurance: spreading the risk

- Premiums and Claims

- Advantage of Size

- Relationship to Theory of Means

Public Opinion Polls

- 19/20

- Connection with Theory of means

- Sample Size and Population Size

Graphs for Data Display

- Histograms (turkey mail)

- Scatter plots for two variables (tiger prey)

- Logarithms (tiger prey, cell phone fraud)

Regression for Prediction

- Predicting quantities

 - Gerow && (tiger prey); De Veaux (data mining)

- Predicting qualities

 - Kahn & Roseman (turkey mail), De Veaux (PVA data)

Experiments vs Observational Studies (Hill article)

- Human Subjects

 - Clinical trials

 - Double Blind, Placebo

 - Random allocation of treatments (Kahn & Rozeman – turkey mail)

Lurking Variables in Observational Studies (Simpson's Paradox)

Design of Experiments

- (Eddy article on Memory)

 - Repetition, Control, Randomization

- (Hill article on School Choice Programs)

 - Missing Values. Non-Compliance, Investigator Choices

Decision Making (Madigan Article on SPAM)

- Error Reduction

- Multiple Indicators (bag of words)

Population Growth Models (Matis article – bee invasion)

Lotteries

- Cash Flow

- Randomized Selection for Prizes

- Requirements for a fair choice

Coincidences

- Why they are common

- Birthday Paradox

Trial and Error