

Quantitative Reasoning II
Midterm Study Guide
Midterm Date: Tuesday Feb 4

1. Identifying and Explaining Cognitive Biases

Students are expected to be able to identify instances of the following 'instincts' (i.e. cognitive biases) and explain how these instincts impair development of a fact-based worldview (from Rosling).

The Gap Instinct: The tendency to divide the world into two distinct groups—rich vs. poor, developed vs. developing—when in reality, most people live in the middle-income category.

The Negativity Instinct: The tendency to focus more on negative news and assume the world is deteriorating, despite evidence of progress.

The Fear Instinct: The human tendency to overestimate danger and react disproportionately to perceived threats.

The Size Instinct: The tendency to misjudge the importance of a single number without comparison, leading to distorted perceptions of scale and proportion.

The Straight Line Instinct: The tendency to assume that trends will continue in a straight line, leading to misinterpretations of future developments.

2. Statistical Literacy

The midterm will test understanding of the following concepts from Spiegelhalter:

Categorical Data:

- Binary vs. Multicategory
- Ordered vs. Unordered Categories

Proportions and Percentages:

- Framing effects (e.g., "98% survival rate" vs. "2% mortality rate")
- Communicating risk effectively

Risk Comparisons:

- Absolute vs. Relative Risk
- Odds Ratios and their Misinterpretation
- "Number Needed to Treat" (NNT)

Survey Data and Inductive Inference:

- The process of moving from data → sample → study population → target population
- Challenges at each step (biases, sample representation)

Measurement Validity & Reliability:

- Social desirability bias, memory recall issues

Sampling Issues:

- Importance of **random sampling**
- Nonresponse bias in surveys

Distinctions Between Data and Population:

- Population distribution vs. sample distribution
- Normal Distribution (Bell Curve)

Types of Algorithms:

- Classification vs Prediction
- Supervised vs Unsupervised Learning

Evaluating Algorithms:

- Accuracy, Sensitivity, and Specificity
- ROC Curves and AUC (Area Under the Curve)
- Overfitting

Regression and Prediction:

- Galton's study on height and the concept of "Regression to the Mean"
- Correlation vs. causation
- Least-squares regression and best-fit line

Understanding Regression Models:

- Independent vs. Dependent variables
- Regression coefficients and their interpretation

Multiple Regression:

- Using multiple predictors in a model (e.g., predicting height using both parents' heights)
- Adjusting for confounders to better isolate relationships

Different Types of Regression:

- Logistic Regression (used for proportions, such as survival rates)
- Limitations of linear regression in cases where proportions cannot exceed 100%

Common Pitfalls in Regression:

- Confounding variables and their impact
- The importance of adjustment in multiple regression

3. From Theory to Practice: Statistical Concepts in R

Students are expected to understand how R is used to support statistical data analysis. Students may be expected to interpret or explain code on the exam, but will not be required to write code.

Data Frames: Structured data storage with rows (specimens) and columns (variables).

Types of Variables:

- **Quantitative** (e.g., height, age)
- **Categorical** (e.g., gender, species)

Basic Data Operations in R

- Inspecting datasets.
- Wrangling (filtering, summarizing, arranging).

Statistical Graphics:

- **Point Plots:** Graphically representing relationships between variables.
- **Annotated Models:** Used to show statistical trends over data points.
- **Violin Plots:** Represent probability distributions.

Building Prediction Models: Train models on past data to make predictions.

Prediction Intervals: Quantify uncertainty in predictions.

Residuals: Measure prediction errors.

Classifier Performance Metrics: Sensitivity, specificity, Brier score.

Regression Analysis: Quantify relationships between variables.

Adjustment for Covariates: Control confounding variables.

Regression to the Mean: Understanding statistical trends.