Midterm Review

DSCC 462
Computational Introduction to Statistics

Anson Kahng
Fall 2022

Chapter 1: Statistical Data

	Nominal	Ordinal	Interval	Ratio
Categorizes and labels variables	✓	✓	✓	✓
Ranks categories in order		✓	✓	✓
Has known, equal intervals			✓	✓
Has a true or meaningful zero				✓

Categorical **Quantitative**

Chapter 2: Descriptive Statistics and Displays

- Categorical:
  - Frequency tables, bar charts
- Continuous:
  - Histograms, Sturges' formula
  - Boxplots, modified boxplots
  - Center: mean, trimmed mean, median, mode
  - Dispersion: quantiles, IQR, variance, standard deviation, coefficient of variation, skewness
  - Concentration: z-scores, empirical rule, Chebyshev's inequality
  - Normality: quantile plots, linear transformations, Box-Cox transformations

Chapter 3: Relationships Between Variables

- CQ: side-by-side histograms or boxplots
- CC: two-way tables
- QQ: Scatterplot
  - Strength, direction, form, outliers
  - Correlation and covariance

Chapter 4: Probability and Combinatorics

- Probability:
  - Intersection, union, complement of events
  - Probability rules, conditional probability, multiplicative rules
  - Independence (pairwise and mutual), mutual exclusivity, LoTP
  - Bayes' Theorem, PPV, NPV
- Combinatorics:
  - Permutations (ordered), combinations (unordered), product and sum rules, stars and bars

Chapter 5: Distributions

- Discrete and continuous random variables:
  - PMFs/PDFs, CDFs, quantiles
  - Expected value, linearity of expectation, variance, transformations
  - Independence, covariance
- Distributions
  - Discrete: Bernoulli, Binomial, Poisson, Geometric
  - Continuous: Uniform, Exponential, Normal
  - Sampling distributions, CLT

Chapter 6: Confidence Intervals

- Unknown mean, known variance: normal distribution
  - One- and two-sided, margin of error, sample size
- Unknown mean, unknown variance: Student's t distribution
  - One- and two-sided, margin of error, sample size
- Unified approach:
  - Determine distribution, convert to standard normal / t, then backsolve

Chapter 7: Hypothesis Tests

- One- and two-sided z- and t-tests
- CIs are the same here; rejection regions
- Type I and type II errors, power
- Sample size given α and β (one-sided and two-sided z- and t-tests)
- (Unified framework for running hypothesis tests…)

Chapter 8: Hypothesis Testing with Two Samples

- Paired samples
  - Basically one-sample tests
- Independent samples
  - Equal, known variance: known σ^2, z-test
  - Equal, unknown variance: pooled s_p^2, t-test
  - Unequal, unknown variance: Welch's t-test, no pooled variance, ν df

Chapter 9: Inference for Variances

- Sampling distribution: chi-squared distribution
- Hypothesis tests for one population variance: chi-squared
  - CIs
- Hypothesis tests for (ratio of) two population variances: F distribution
  - CIs

Chapter 10: Inference for Proportions

- Sampling distribution: Normal approximation
- One sample:
  - CIs: Wald estimate for p
  - Sample size estimate
  - Hypothesis testing: use $p = p_0$
- Two samples:
  - Hypothesis testing: pooled $\hat{p}$
  - CIs: Wald estimate: $p_1 = \hat{p}_1, p_2 = \hat{p}_2$

Chapter 11: Chi-squared Tests

- Goodness-of-fit
  - True proportion = expected proportion?
  - Generalizes proportion tests
- Chi-squared test of independence
  - Are categorical variables related?
- Everything follows chi-squared distributions; only interested in upper tail probabilities here (“more extreme”)

Chapter 12: Nonparametric Tests

- Wilcoxon Signed-Rank Test
  - One-sample / paired samples (analog to t test)
- Wilcoxon Rank-Sum Test / Mann-Whitney U Test
- Two-sample (analog to two-sample t test)