Serik Sagitov, Chalmers Tekniska Högskola, May 22, 2004

Course overview

1. Descriptive statistics

Cdf estimation: empirical cdf
  survival function and the hazard function
Density estimation: vertically scaled histogram,
  kernel density estimate, and steam-and-leaf plot
Q-Q plots and normal probability plots
Measures of location: sample mean $\bar{X}$,
  sample median $\hat{M}$, and α-trimmed mean $\bar{X}_\alpha$
Measures of dispersion: sample variance s^2, IQR, MAD
Skewness and kurtosis. Boxplots, outliers

2. Parameter estimation

Population parameters and sample parameters
  point and interval estimates
Sampling distribution and sampling error
  systematic and unsystematic errors
  unbiased and consistent estimates
  estimated standard error
Parametric statistical models
  method of moments and ML method
Confident intervals exact and approximate
  nonparametric CI for the population median
Parametric and non-parametric bootstrap

3. Hypotheses testing
Statistical hypotheses: simple and composite
   one-sided and two-sided
Two types of error, error sizes: α and β
   significance level and the P-value, two-sided P-value
   exact and approximate null distributions
   the power of the test, planning the sample size
Parametric and non-parametric one-sample tests
   one-sample t-test, sign test, the signed rank test
   the small-sample test for the proportion
   large-sample tests for proportion and for mean
The duality of CI and hypotheses testing

4. Two or more samples
Two-sample t-test, Wilcoxon rank sum test
Simple linear regression model
   least square estimates, normal equations
   coefficient of determination
   sample covariance, sample correlation
   assesing the fit with residuals
CI and hypothesis testing concerning the slope β_1
   model utility test of H_0: $\beta_1 = 0$
   confidence bands for individual observations
Multiple regression model
   adjusted coefficient of multiple determination

Anova tests
   the one-way layout F-test, Kruskal-Wallis test
   the two-way layout F-test, Friedman's test
Likelihood ratio tests and generalized LRT
   the power of test, optimal tests
   goodness-of-fit chi-square test, degrees of freedom
Categorical data: Fisher's exact test
   chi-square tests of homogeneity and of independence
   McNemar's test, odds ratio
Simultaneous confidence intervals
   Bonferronni's method and Tukey's method

5. Experimental design
Sampling with and without replacement
   finite population correction
Sratified random sampling
   optimal and proportional allocations
Two-sample tests: independent and paired samples
   double-blind, randomized controlled experiments
   confounding, Simpson's paradox
Anova tests: one-way layout and two-way layout
   randomized block design
Contingency tables
   independent samples or a cross-classified sample
   matched-pairs design
   prospective and retrospective studies

6. Decision theory and Bayesian inference

Risk function for a decision rule based on a loss function
    minimax decision rules

Bayesian approach: prior and posterior distributions
    Bayes risk and Bayes decision rule
    Posterior risk and Bayes action

Conjugate priors
    Beta and Dirichlet distributions, pseudocounts
    Gamma distribution

Bayesian estimation: MAP and PME
    credibility interval
    Bayesian hypotheses testing and LRT