Studies in Logic, Vol. x, No. x (xxxx): 1–10
PII: 1674-3202(xxxx)-0x-0001-10

Statistical Thinking and Data Analysis

MIT

hienminhnguyen711@gmail.com

Abstract.

Book Tamhane, Ajit C., and Dorothy D. Dunlop. Statistics and Data Analysis: From Elementary to Intermediate. Prentice Hall, 1999. ISBN: 9780137444267.

1 Review of probability

Computing probabilities: conditional probabilities and Bayes' rule Quantiles/percentiles, CDF's, mean, median, variance, standard deviation, covariance, various distributions and what they are used for (particularly Bernoulli, Binomial, Multinomial, Hypergeometric, Poisson, normal)

2 Collecting data

3 Summarizing and exploring data

Summarizing univariate data: numerically (sample mean, IQR, etc.) and by plotting (pie/bar/pareto chart for categorical data, histogram, box plot, normal plot) Summarizing bivariate data: Simpson's paradox, scatter plot, sample correlation coefficient Time series: MA, EWMA, forecast error and MAPE, auto-correlation coefficient

4 Sampling distribution of statistics

Normal approximation to binomial distribution (which relies on the CLT), computing probabilities with chi-square distribution, t-distribution, F-distribution

5 Basic concepts of inference

Bias, MSE, setting up hypotheses, Type I error, Type II error, power For z-test: z-scores, p-values, confidence intervals

6 Inferences for single samples

Sample size calculation for confidence intervals on z-test, sample calculation for z-test, sample size calculation for power on z-test, t-test, chi-square test for variance

7 Inferences for 2 samples

QQ plots Comparison of two means for independent samples design (large samples z-test, small sample t-test using either a pooled variance or the Welch-Sattethwaite method) Comparison of two means for matched pairs design (t-test, power and sample-size calculation for power) Comparison of variance using the F-test

8 Inferences for proportions and count data

Comparison to a given proportion using large sample z-test, sample size calculation for confidence intervals Comparison of two proportions using large sample z-test Chi-square test (multinomial and goodness of fit)

9 Similar linear regression and correlation

Computing the least square line, computing r^2, $hypothesistestingonbeta_1$, $understandingANC$

10 Multiple linear regression

Understanding ANOVA regression tables, t-tests on individual regression coefficients Multicollinearity Logical regression

11 Nonparametric statistical methods

Comparison to a given median using: sign test, Wilcoxon signed rank test (these tests can also be used on the di's for matched pairs) Comparison of two distributions using Rank Sum test or MWU test Rank correlation methods: Spearman's rank coefficient, Kendall's Tau