Stat 571: Statistical Methods  
List of Topics

This course follows the textbook “Statistics and Data Analysis: From Elementary to Intermediate” by Ajit Tamhane and Dorothy Dunlop. The chapters covered are 1 to 13. Chapter 2 on probability is only briefly covered. The JMP software package is integrated with the course material. The course transparencies and other course information are available in the course home page at http://web.utk.edu/~leon/stat571/

<table>
<thead>
<tr>
<th>Topic</th>
<th>Target Number of Days</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Introduction - Statistical Terminology:</strong> Descriptive statistics or exploratory data analysis, inferential statistics, population, sample, variable, parameter, statistic, random sample, sampling variability, probability versus statistics.</td>
<td>1</td>
</tr>
<tr>
<td><strong>2. Review of Probability:</strong> Approaches to probability (classical, frequentist, subjective, axiomatic), axiomatic approach, frequency tables, conditional probability, Bayes theorem, random variables, distribution function, density function, expected value, variance and standard deviation, quantiles and percentiles, covariance and correlation, Chebyshev’s inequality, weak law of large numbers, discrete distributions (Bernoulli, Binomial, Poisson), continuous distributions (Uniform, Exponential, Gamma), normal distribution.</td>
<td>2</td>
</tr>
<tr>
<td><strong>3. Collecting Data:</strong> Historical data, types of studies (comparative, descriptive or noncomparative, observational, experimental), confounding, lurking variables, control group, treatment or intervention, control group, pretest-only design, pretest-posttest design, placebo effect, single blind study, double blind study, concurrent control group, historical control group, sample surveys, prospective studies, retrospective studies, acceptance sampling, censuses, target population, sampled population, sampling and nonsampling errors, bias, representative sample, cohort studies, case-control studies, judgment sampling, quota sampling, simple random samples, sampling rate, sampling frame, stratified random sampling, multistage cluster sampling, probability-proportional-to-size sampling, systematic sampling, 1-in-k systematic sample, treatment factors, nuisance or noise factors, treatment group, replicate versus repeat measurements, systematic error, random error, measurement error, blocking, regression analysis, covariates, randomization, completely randomized design, randomized block design, iterative nature of experiments.</td>
<td>2</td>
</tr>
</tbody>
</table>
## 4. Summarizing and Exploring Data:
Variable types (categorical, qualitative, nominal, ordinal, numerical, continuous, discrete, interval, ratio), summarizing categorical data (frequency table, bar chart, Pareto chart, pie chart), summarizing numerical data (mean, median), skewness, outliers, measures of dispersion (quantiles, range, variance, standard deviation, interquartile range, coefficient of variation) standardized z-scores, histogram, stem and leaf diagram, box and whiskers plot, fences, normal plots, departures from normality, normalizing transformations, runs chart, summarizing bivariate categorical data (two-way table, mosaic plot), Simpson’s paradox, adjusted (standardized) rates, summarizing bivariate numerical data (scatter plot, simple correlation coefficient, sample covariance), correlation versus causation, straight line regression, regression towards the mean, regression fallacy, summarizing time-series data, data smoothing, forecasting techniques.

## 5. Sampling Distributions of Statistics:
Estimates, sampling error, frequentist approach to statistics, sampling distribution of the sample mean, Central Limit Theorem, Law of Large Numbers, normal approximation to the Binomial, sampling distribution of the sample variance, Chi-Square distribution, Student’s t-distribution, F-distribution.

## 6. Basic Concepts of Inference:
Estimation, hypothesis testing, point estimation, confidence interval estimation, estimator, estimate, bias and variance of estimator, mean square error, precision and standard error, confidence level and limits, frequentist interpretation of confidence intervals, null and alternative hypothesis, type I and II error, probabilities of type I and II error, acceptance sampling, simple and composite hypothesis, P-value, one-sided and two-sided tests, use and misuse of hypothesis tests in practice, multiple comparisons.

## 7. Inference for Single Samples:
Inference for the mean (large samples), confidence intervals for the mean, test for the mean, sample size determination for the z-interval, power calculation for one-sided and two-sided z-test, power function curves, sample size determination for the one-sided and two-sided z-test, inference for the mean (small samples), t distribution, confidence intervals based on the t distribution, inference on variance, confidence intervals for the variance and standard deviation, hypothesis test on variance and standard deviation, prediction intervals, tolerance intervals.
8. **Inference for Two Samples:** Independent sample design, matched pair design, pros and cons of each design, side by side box plots, comparing means of two populations, large sample confidence interval for the difference of two means, large sample test of hypothesis for the difference of two means, inference for small samples (confidence intervals and tests of hypothesis), unequal variance case (confidence intervals and hypothesis tests), sample size determination assuming equal variances, confidence intervals and test of hypothesis for matched pair design, statistical justification of matched pair design, sample size determination for matched pair design, comparing variance of two populations.

9. **Inference for Proportions and Count Data:** Large sample confidence interval for proportion, sample size determination for a confidence interval for proportion, large sample hypothesis test on proportion, comparing two proportions in the independent sample design (confidence interval and test of hypothesis), inference for two-way count data (total sample size fixed, row total fixed), chi-square statistic.

10. **Simple Linear Regression and Correlation:** Dependent and independent variables, probability model for simple linear regression, least squares fit, goodness of fit of the LS line, sums of squares, geometry of sums of squares, coefficient of determination, estimation of error variance, statistical inference for slope and intercept (tests of hypothesis and confidence intervals), analysis of variance, prediction of future observation, confidence and prediction intervals, calibration (inverse regression), regression diagnostics, residual plots, mathematics of residuals, checking for linearity, quadratic model, checking for constant variance, checking for normality of errors, checking for independence of errors, checking for outliers, standardized (studentized) residuals, checking for influential observations, hat matrix, leverage plots, data transformations, variance stabilizing transformations, correlation analysis, bivariate normal density function, statistical inference on the correlation coefficient, correlation between test instruments.
11. **Multiple Linear Regression**: Probability model for multiple linear regression, least squares fit, sums of squares, coefficient of multiple determination, centered and uncentered polynomial regression, statistical inference on the slopes (individually and simultaneously), tests on subsets of the slopes, Type I and III sums of squares, influential observations, transformations, multicollinearity, correlation matrix, Variance Inflation Factor (VIF), regression coefficients in the presence of multicollinearity, dummy predictor variables, JMP’s choice of dummy variables, one dummy and one continuous predictor variables, interaction, using dummy variable to adjust for seasonality, confidence and prediction intervals, logistic regression and logit models, variable selection methods (stepwise regression, best subset regression, optimality criteria).

12. **Analysis of Single Factor Experiments**: Completely randomized design (CRD), randomized block design (RBD), side-by-side box plots, model assumptions for CRD, treatment effect, alternative formulation of CRD model, CRD parameter estimates, confidence intervals for treatment means, mean diamonds, CRD analysis of variance, relationship to dummy variable regression, model diagnostics (residuals versus fitted value, residual versus row order, normal plot of residuals), multiple comparison of means, pairwise equality hypothesis, familywise error rate, Bonferroni Method, Fisher’s protected Least Significant Difference (LSD) method, Tukey’s method, Dunnett’s method for comparison with a control, Hsu’s method for comparison with the best, RBD model assumptions, RBD analysis of variance, degrees of freedom explained, no interactions between treatment and blocks, residual plots.