Tentative List of Review Topics

This is a ***tentative*** list of test relevant topics covered in class. Any topic labeled by (*) is ***not*** test relevant.
Revisit the topics that you practiced in the labs. These are highly test relevant.

Key Focus:

- Review your labs; concepts practiced in the labs will become part of the final exam.

Miscellaneous

- Equations: Internal structure of equations. Inline versus display equations
- Key functions, data structures etc. in 

Chapter 1

- Data -> Information -> Knowledge -> Understanding -> Wisdom
- Theory -> hypothesis -> prediction
- Empirical versus Experimental Research (*)
- Measurement Theory -> importance of variability -> measurement scales
- Measurement Scales: Hierarchy among four scales. Varying Information content and transformations among measurement scales. Cartographic objects and measurement scales. Discrete and continuous variables
- Data problems: Measurement validity, bias, accuracy versus error, precision of measurement tool, outliers not representative of the underlying population.
- Meaning of statistics (*)

Chapter 2

- Advantages of data visualization and statistical graphs.

- Elements of the grammar of graphics (*)
- Color coding of categorical, gradient and diverging map themes.
- Color blindness
- Statistical graphs in relation to measurement scales
- Construction of frequency distributions
- Construction rules for histograms: number of classes, bins and breaks, visual resolution, comparison of distributions side-by-side
- Stem-and-Leaf Graphs (*)
- Bar charts: ordering of columns
- Pie charts: problem of comparison, scaling pies (*)
- Shape of distributions: symmetry, skewness, modality, outliers
- Construction of cumulative frequencies, concept of quartiles and median
- Visual comparison of histograms versus empirical cumulative distribution functions
- Quantiles and Percentiles: ordered data
- Box-Plots with construction rules.
- Time Series/Line Graphs (*)
- Design guidelines for statistical graphs

Chapter03

Math review

- Summation notation with basic rules
- Double summation
- Computational and definitional expressions of variance
- Concept of derivatives (*)
- Exponential and logarithmic functions (*)
- Population parameters and sample statistics

Central Tendency

- Central tendency: Impact of measurement scale, pro and cons of each measure

- Mid-range, mode, median (odd/even number of observations, optimal criterion), arithmetic mean
- Constraints of mean, optimal criterion, impact of outliers
- Relationship of skewness and central tendency measures
- Trimmed and Winsorized mean, outliers
- Impact of outliers
- Geometric mean
- Bimodality index (*)

Variability

- Accuracy and variability
- Range and Interquartile range
- Reference point to measure variability, zero-sum restriction
- Mean absolute deviation
- Variance and standard deviation, rational for loss of degrees of freedom
- 2/3 rule
- Concepts of unbiasedness and efficiency.
- Winsorized trimmed variance versus trimmed variance.
- Higher order moment measures: skewness and kurtosis and their interpretation.
- Coefficient of variation.

Chapter 4

Correlation

- Concept of relationships versus cause -> effect
- Impact and causes of biasedness
- Bivariate relationships for metric and categorical variables
- Concept of statistical dependency/independence
- Cross-tabulations
- Linear correlation for different scatterplot relationships, importance of data visualization, scatterplot matrices, correlation matrices

- Concept of confounding variables and their impact on correlation/and cross tabulations
- Structure and properties of the correlation coefficient, *z*-transformation
- Spearman rank correlation

Bivariate regression

- Exogenous and endogenous variables
- Controlled measurements in statistical experiments
- Conditional expectation and linear conditional expectation; systematic and random components.
- Notation
- Interpretation of intercept and slope.
- Degrees of freedom and parsimony principle
- Prediction: Interpolation and extrapolation
- Underlying assumptions
- Ordinary least squares: Minimization of sum of squared residuals
- Variance decomposition (TSS=ESS+RSS) and R^2 and $R^2_{adjusted}$
- Properties of line: Mean of $\bar{X}$ and mean of $\bar{Y}$ are pivot point. Zero slope, relationship between intercept and mean of Y. Negative correlation between slope and intercept.
- Properties of residuals: zero sum property, uncorrelated with exogenous variable and predicted value.
- Root mean square error.
- Interpretation of computer output.

Multiple regression

- Structure of equation, ***interpretation*** of partial regression coefficients
- Regression coefficient and scale of variables
- Residuals uncorrelated with exogenous variables
- Confounding variables and change of regression coefficient
- Controlling for confounding effects
- Modeling non-linear relationships with an added quadratic term
- Basic model diagnostics: Outliers and non-linearity

Chapter 5

Sets and Probability

- Definitions: Elementary Outcome, sample space, event, empty set, complementary set, subset, intersection, mutually exclusive, union.
- Basic set operations
- Three Kolmogorov's probability axioms.
- Perspectives on probability: analytical, relative frequency, subjective with discussion
- Counting rules: Combination, permutation, product rule, hyper-geometric rule
- Sampling with/without replacement, with/without considering the order.
- Addition theorem, complementation theorem, conditional probability, statistical independence, multiplication theorem
- Bayesian theorem, *posteriori* and *a priory* probabilities
- Concept of a random variable

Univariate Distributions

- Definition and properties: Probability distribution function and density function for discrete and continuous random variables
- Concept of expectation and variance (* integration skipped for continuous random variables)
- Specific distributions with properties: discrete uniform, continuous uniform, normal distribution, standard normal distribution.
- Detailed derivation of binomial distribution with its properties.
- Other distributions (*): Poisson, geometric, negative binomial, multinomial, hypergeometric, exponential, gamma, beta, log-normal, chi-square, t, F
- Working with the standard normal table.

Bivariate distributions

- Contingency tables for nominal and ordinal scaled variables.
- Probabilities in contingency tables
- Visualization of conditional probabilities in contingency tables with stacked bar charts.
- Concept of covariance
- Shapes of bivariate normal densities in relationship to inter-variable correlation

Chapter 6:

- Concepts of representative, bias, and sampling error
- Basic steps of empirical research (*)
- Delineation of relevant population under investigation
- Definitions: Target population and sampling frame (*)
- Component of sampling frame: non-response, sampled and ineligible (*)
- Sampling design guidelines
- Properties of simple random sampling
- Independence approximation in sampling without replacement
- Definition of sample statistic as function of random variables
- Distribution of a sample statistics form population properties. Example of development for a small population.
- Definition: standard error
- Central limit theorem with implications.
- Random sample designs (*): Technical details of stratified and clustered sampling
- Random sample designs: Properties of stratified and clustered sampling
- Spatial sampling designs: Generation of random points. Quadrate sampling, traverse sampling

Chapter 7:

- Impact of sampling variability
- Difference between estimation and hypothesis testing
- Point estimator with bias, efficiency, asymptotic properties, root-mean-square error
- Tradeoff between bias and root-mean-square error
- Interval Estimation: interval width, uncertainty, error probability, precision, interpretation of interval estimators
- Properties of specific point estimators
- Relationships: [a] error probability and interval width, [b] interval width and sample size
- Intervals for non-normal populations, central limit theorem, *t*-distribution, binomial distribution
- Calculation of sample sizes
- Variance of binomial distribution and sample size

- Confidence intervals around regression coefficients

Chapter 8:

- What are statistical hypothesis: formulation of null and alternative hypothesis
- Why can only the alpha error be controlled
- One- and two-sided hypothesis
- Sampling error versus systematic difference from the value under the null hypothesis
- Relationship between the distributions of the test statistic under the null hypothesis and a hypothetical value under the alternative hypothesis
- Relationship between [a] alpha error and critical value, and [b] value of sample test statistic and prob-value (significance level)
- What errors can be committed in statistical hypothesis testing: difference between alpha error and beta error
- Steps of classical hypothesis testing
- Test for population expectation under varying assumptions
- Exact and approximate test for population proportion based on the binomial distribution
- Relationship between tests and confidence intervals
- Problems of hypothesis testing
- Significance tests for correlation coefficients
- Distribution of the correlation coefficient under alternative hypothesis (*)
- Fisher *z*-transformation of the correlation coefficient (*)

Chapter 9:

- Matched and independent sample designs
- Different null- and alternative hypotheses associated with test for differences in means
- t-test for matched and independent samples
- construction of pooled variances
- Test for equality of variances (homo- and heteroscedasticity)
- Test for equality of proportions
- Simulation of distribution of test statistics (*)

Chapter 10: Non-parametric Statistics

- Difference between parametric and non-parametric tests
- Assumptions of χ^2-test: Counts from a Poisson distribution, minimum size of expected frequencies
- Goodness of fit test: Calculation of expected frequencies, degrees of freedom.
- Test for independence in contingency tables, calculation of degrees of freedom, impact of sample size
- Kolmogorov-Smirnov Test
- Univariate and bi-variate kernel densities. Role of the bandwidth

Chapter 12: Inferential Aspects of Linear Regression

- Assumptions of regression analysis
- Homoscedasticity
- Standard error of a regression coefficient and its associated t-test
- Evaluating residuals and identifying outliers and/or unreasonable observations
- Diagnostic plots: residuals versus predicted values
- Effect of the spread of an independent variable on the on standard error of a regression coefficient
- Difference of regression with rates and absolute values
- Use of the log-transformation to cope with positively skewed variables

Chapter 13: Extending Regression Analysis

- Standardized regression coefficient
- Multicollinearity
- Spatial relationships: multidirectional, un-equal spaced, multilateral (*)
- Spatial relationship matrix: adjacency (topological), inverse distance based (*0
- Adjacency and islands (*)
- Row-sum standardization (*)
- Causes of spatial autocorrelation: model misspecification, spatial interaction process, aggregation problem (*)
- Misspecification perspective: Increase of the number of independent variables and reduction of autocorrelation (*)

- Extreme positively and negatively autocorrelated patterns (*)
- Moran's plot (*)
- Moran's I, expected value and variance under the assumption of independence and statistical hypothesis testing (*)