Learning objectives for Stat 370

A. Introduction to types of studies, data collection

A1. Explain the difference between a descriptive and inferential statistics
A2. Given a study, identify the population, sample, parameters, and statistics
A3. Given a variable, determine whether it is qualitative or quantitative
A4. Given a quantitative variable, determine whether it is discrete or continuous
A5. Explain the difference between an observation study and a designed experiment
A6. Determine the type of situation where an observational study would be appropriate
A7. Determine the appropriate conclusions from an observational study
A8. Given a study, determine whether it is observational or experimental
A9. Explain the difference between a simple random sample and a convenience sample
A10. Given a sample design, determine whether it is a simple random sample
A11. Given a sample design, analyze whether it is appropriate, or what could go wrong

B. Graphical and numerical summaries (relevant software http://statcrunch.stat.ncsu.edu/)

B1. Given a set of discrete data, make frequency table and histogram by hand
B2. Given a set of continuous data, make classes, frequency table for classes, and then draw histogram by hand
B3. Given a set of raw data, make histogram using software
B4. Given a histogram, determine the number of individuals in a particular range
B5. Given a boxplot, describe the distribution’s shape (left skewed, right skewed, symmetric, unimodal or multimodal)
B6. Given a histogram, identify outliers
B7. Given side-by-side boxplots, compare and contrast key features of the groups represented by the boxplots
B8. Given a boxplot, determine the shape of the distribution (skewed right, skewed left, symmetric)
B9. Given a boxplot, determine the 5-number summary
B10. Given a five-number summary, create the corresponding boxplot
B11. Explain the impact of adding an outlier on summary statistics such as mean, median and standard deviation
B12. Given a raw data set with not many data points, determine the five-number summary and draw the corresponding boxplot
B13. Given a data set, determine the min, max, quartiles, mean, standard deviation (using software, and also by hand if data set is small)
B14. Explain how the mean and median are related for different shapes of a distribution (skewed left, skewed right or symmetric)
B15. List the following characteristics of the standard deviation
   a. The standard deviation must be greater than or equal to zero
   b. When the standard deviation is zero, there is no spread – every number of the data set is the same
B16. Describe how linear transformations affect median, mean, and standard deviation
B17 Apply the empirical rule to a data set

C. Design of experiments

C1. Given a designed experiment, identify the factors, treatments, response variables, and experimental units
C2. List reasons for variability of responses (treatment effect, experimental error)
C3. List sources of experimental error
C4. Explain what it would mean to control for a variable
C5. Explain why we would want to control for a variable
C6. List two ways to deal with experimental error that may remain after controlling variables (randomization, blocking)
C7. Define replication
C8. Give reasons for replication in designed experiments
C9. Define randomization
C10. Explain why one wants to randomize in a designed experiment
C11. Define a block (homogenous subset of experimental units)
C12. Explain why one would want to “block” a subset of experimental units
C13. Given a design, determine which variables are randomized, controlled, blocked

D. Factorial data analysis

D1. Given a factorial design, determine the factors and levels of the factors
D2. For a factorial design, tabulate the treatment means
D3. For a factorial design with two factors, draw both interaction plots
D4. Compute simple and main effects of levels of treatments
D5. Compute interactions for a factorial design with two factors
D6. For a factorial design with two factors, determine whether there is interaction graphically, through computed interactions, and by seeing if effect of factors can be described without mentioning the other factor
D7. Determine which factors have a greater effect through fitted main effects
E. Analysis of Variance

E1. Based on an ANOVA table, determine if interaction or treatment effects are statistically significant
E2. Explain when an additive model is appropriate, as opposed to a full factorial model
E3. Create an additive model table from an ANOVA table for the full model
E4. Given a partial ANOVA table, fill in the missing pieces
E5. Explain that in the presence of significant interaction, simple effects should be reported and not main effects
E6. Explain when an additive model table is appropriate

F. Simple linear regression

F1. For a linear regression model, identify the independent variable, and the dependent variable
F2. Interpret the sample correlation $r$ in terms of strength and direction of the linear relationship
F3. Define residuals
F4. Compute residuals given a data set and least squares line
F5. Explain the method of least squares (how the line is determined)
F6. Interpret the coefficient of determination $R^2$
F7. Given two of $SS(\text{total})$, $SS(\text{model})$, $SS(\text{error})$, compute $R^2$
F8. Interpret a given ANOVA table associated with linear regression
F9. Use the least squares line to predict $Y$ for $X$ values in range of data
F10. Interpret the slope of the least squares line in the context of a specific problem
F11. Interpret the intercept of the least squares line (if applicable)
F12. Explain why extrapolation is dangerous
F13. Explain when the intercept is meaningful

G. Discrete random variables

G1. Given a probability mass function (pmf), compute associated probabilities
G2. Given a pmf, compute cumulative distribution function (cdf)
G3. Given a pmf, compute mean, variance, standard deviation
G4. Given a cdf, compute pmf
G5. Determine whether a given function defines a pmf
G6. Determine whether a given function defines a cdf
G7. Compute probabilities associated with a binomial random variable
G8. Compute mean and variance of a binomial random variable
G9. Apply binomial model to word problems
G10. Explain when a binomial model is appropriate
H. Continuous random variables

H1. Given a probability density function (pdf), compute associated probabilities
H2. Given a pdf, compute cdf
H3. Given a pdf, compute mean, variance, standard deviation
H4. Given a cdf, compute pdf
H5. Determine whether a given function defines a pdf
H6. Determine whether a given function defines a cdf
H7. Given a cdf, determine if associated random variable is discrete or continuous

I. Normal Distribution

I1. Compute probabilities associated with a normal random variable using standard normal table
I2. Compute percentiles for normal random variables
I3. Apply the normal distribution to word problems
I4. Describe and use properties of normal distribution (symmetry, ability to standardize general normal random variable to get standard normal)
I5. Determine values centered about the mean that contain percentages of normal distribution

J. Sampling Distribution of $\bar{X}$

J1. Determine the distribution of $\bar{X}$ when underlying data is normally distributed
J2. For large samples, apply Central Limit Theorem to approximate probabilities associated with $\bar{X}$ when underlying data is not normally distributed
J3. Apply sampling distribution of $\bar{X}$ to word problems
J4. Determine the effect of increasing the sample size on variance of $\bar{X}$
J5. Explain properties of t-distribution relative to normal distribution

K. Confidence Intervals for $\mu$

K1. Construct a confidence interval for $\mu$ when underlying data is normally distributed with $\sigma$ is known
K2. Construct a confidence interval for $\mu$ when underlying data is normally distributed with $\sigma$ is unknown (use t-distribution)
K3. Construct a large-sample confidence interval for $\mu$ when $\sigma$ is known (use normal distribution)
K4. Construct a large-sample confidence interval for $\bar{X}$ when $\sigma$ is unknown (use t-distribution)
K5. Interpret a confidence interval about a population mean

L. Hypothesis Testing for $\mu$
L1. Given a problem involving testing, determine the null and alternative hypotheses
L2. Determine if a test is one- or two-sided
L3. Given level of significance $\alpha$, determine the critical (rejection) region of a test
L4. Given a problem involving testing, identify the test statistic
L5. Explain the role of a test statistic in hypothesis testing
L6. Explain the meaning of the level of significance $\alpha$
L7. Test hypotheses for the population mean using the classical (rejection region) approach when $\sigma$ is known and data is normally distributed
L8. Test hypotheses for the population mean using the classical (rejection region) approach when $\sigma$ is unknown and data is normally distributed
L9. Test hypotheses for the population mean using the p-value approach when $\sigma$ is known and data is normally distributed
L10. Test hypotheses for the population mean using the p-value approach when $\sigma$ is unknown and data is normally distributed
L11. Test hypotheses for the population mean using the classical (rejection region) approach when $\sigma$ is known, data is not normally distributed, and sample size is large
L12. Test hypotheses for the population mean using the classical (rejection region) approach when $\sigma$ is unknown, data is not normally distributed, and sample size is large
L13. Test hypotheses for the population mean using the p-value approach when $\sigma$ is known, data is not normally distributed, and sample size is large
L14. Test hypotheses for the population mean using the p-value approach when $\sigma$ is unknown, data is not normally distributed, and sample size is large
L15. Test two-sided hypotheses concerning a population mean using confidence intervals