GOALS FOR THE INTRODUCTORY COURSE IN APPLIED STATISTICS:

1. to understand the meaning of data commonly encountered in newspapers and general magazines.
2. to understand and evaluate common arguments that reference numerical statistics
3. to understand and evaluate the role of confounding factors
4. to understand and evaluate the role of chance

The primary activities involved in achieving these goals includes:
1. Mastering relevant concepts and relationships.
2. Working problems involving these concepts and relationships
3. Using these concepts and relationships to analyzing and interpret real data.

The core concepts in statistics include:
1. Causality versus correlation. Reliability vs. validity
2. Observational Study vs. Experiment: Knowing the different senses of "control"
   a. Controlled study versus Uncontrolled; control group (self-selected vs. assigned)
   b. Having "control of" a variable
   c. "Controlling for" a variable
3. Population vs. Sample; Parameter vs. Statistics
4. Classification of Samples:
   a. Random vs. convenience
   b. Different meanings of Representative
5. Classification of Sample Error
   a. question bias vs. sample bias

==========

6. Classifying Data:
   a. qualitative vs. quantitative
   b. nominal vs. ordinal; mulinomial vs. binomial
   c. interval vs. ratio; continuous vs. discrete
7. Reading Summary Tables involving Percents and rates
   a. Part-whole: rowpercent, column percent and total percent
   b. Change-whole ratio (difference) and whole-whole ratio (ratio)
   c. Percent versus percentage point; Rates in # per 100,000
   d. Ability to form meaningful comparative statements
   e. Odds ratio vs. likelihood ratio
   f. Ability to read tables involving a control variable and to make meaningful statements
   g. Awareness of reasons for Simpson's Paradox
   h. Ability to convert column percent to row percent in a table (Bayes Theorem)

==========

8. Frequency Distributions:
   a. Summaries of Center: Mean, Median, Mode, Mid-range and mid-interquartile range
   b. Summaries of Spread: Std. Deviation, Range and Interquartile range.
   c. Summaries of shape: Symmetric vs. asymmetric; skewed right vs. skewed left
   d. Normalizing (to non-zero base) and Standardizing (to zero base using z)
   e. [Optional: Measure of skewness, coefficient of variation, etc.]
   f. Quantile-Normal plot -- measure of symmetry and Normality.
9. Reading Scatter Plot, Dotplot, Boxplot and Histogram [Stem and Leaf is optional]
10. Bell-shaped Distribution: 68%/95%/99.7% rule

==========

BIVARIATE DATA
11. Association: Direction (positive/negative) and strength/closeness
12. Predictor variables vs. predicted variable (independent vs. dependent??)
13. Longitudinal fallacy: Change in one variable **within a person** will cause a change in other variable
14. Pearson Correlation and slope of OLS line
15. Spearman correlation
16. Pearson Correlation and Coefficient of Determination
17. Regression Effect and Regression fallacy (effect is due to determinate causes -- instead of chance)
18. Understand phrases such as "accounted by", "explained by" in coefficient of determination
19. [Optional: ANOVA table]

MULTIVARIATE REGRESSION -- to illustrate how one "CONTROL'S FOR" something in modeling.
20. Change in Coefficient of Determination and "Explained by"
21. Understand the importance of knowing what has been "controlled for".

TIME SERIES
22. Modeling variables as a function of time.
23. Explaining high correlations as due to time (independent growth in time) as a common cause.
24. Importance of "controlling for" other variables that are highly correlated via time.
25. Able to read survival charts and process control charts

==========

PROBABILITY
26. Random, Exclusive, AND/OR, Independent
27. Binomial Theorem and Binomial Table
28. Normal Distribution [Poisson is optional]

SAMPLING DISTRIBUTION
29. Concept of sampling distribution of statistic from multiple samples
30. Center and spread of sampling distribution as a function of sample size
31. Spread of sampling distribution decreases as sample size increases
32. Spread of sampling distribution is independent of the size of the population (N>100)
33. Mean is more efficient than the median

ESTIMATION
34. Point Estimate
35. Confidence and Precision
36. Confidence Level and Alpha: Reason for allowing error
37. Margin of Error
38. Confidence Interval for one and two populations
39. Required Sample Size
40. Prediction Interval and Confidence Interval for OLS regression

HYPOTHESIS TEST
41. Idea of judging given a presumptive hypothesis; test of significance
42. Types of error: Type 1 and type 2
43. Alpha measures budget for type 1 error (allowed to limit amount of type 2 error)
44. p-value measures probability of type 1 error.
45. Choice of Null. relation to Alternate
46. Technique for testing the strength (weakness of the null hypothesis
47. t-test. Degrees of freedom
48. Testing for two-sided and one-sided null hypothesis
49. testing for one population and two populations
50. testing for OLS regression
51. testing for independence in table using Chi-squared.

KEY PRINCIPLES:
1. CORRELATION/DISTINCTION IS NOT DIRECT CAUSATION
2. CORRELATION/DISTINCTION IS GENERALLY A RESULT OF A COMMON CAUSE
3. "CONTROL OF" A CONFOUNDING VARIABLE IS THE BEST WAY TO ELIMINATE THE EFFECT OF A COMMON CAUE.
4. "CONTROLLING FOR" A CONFOUNDING VARIABLE IS THE SECOND BEST WAY TO MANAGE THE EFFECT OF A COMMON CAUSE.
5. OBSERVATIONAL STUDIES ALWAYS LACK "CONTROL OF"

The four keys to understanding most arguments in statistics:

1. **Was the study an experiment or an observational study?**
2. **What variables are related to the variable of interest?**
3. **In an experiment, which related variables had value which were not assigned (the experimenter not have control of)?**
4. **Which related variables did the study fail to control for?**

"controlled for" are the key to understanding most arguments involving statistics.

Two primary methods of "controlling for"
- TABLING OR SUBSETING
- REGRESSION (constant factor)