Indian Hill High School
AP Statistics

The purpose of the AP course in statistics is to introduce students to the major concepts and tools for collecting, analyzing, and drawing conclusions from data. Students are exposed to four broad conceptual themes:

1. Exploring Data: Observing patterns and departures from patterns
2. Planing a Study: Deciding what and how to measure
3. Anticipating Patterns: Producing models using probability theory and simulation
4. Statistical Inference: Confirming models

Students who successfully complete the course and examination may receive credit, advanced placement, or both for a one-semester introductory college statistics course. this does not necessarily imply that the high school course should be one semester long. Each high school needs to determine the length of its AP Statistics course to best serve the needs of its students. Statistics, like some other AP corses, could be effectively studied in a one-semester, a two-trimester, or a on-year course. Most schools, however, offer it as a two-semester course.

Course Content Overview

The topics for AP Statistics are divided into four major themes: exploratory analysis, planning a study, probability, and statistical inference.

I. *Exploratory analysis of data makes use of graphical and numerical techniques to study patterns and departures from patterns.* In examining distributions of data, students should be able to detect important characteristics, such as shape, location, variability, and unusual values. From careful observations of patterns in data, students can generate conjectures about relationships among variables. The notion of how one variable may be associated with another permeates almost all of statistics, from simple comparisons of proportions through linear regression. The difference between association and causation must accompany this conceptual development throughout.

II. *Data must be collected according to a well-developed plan if valid information on a conjecture is to be obtained*. The plan must identify important variables related to the conjecture and specify how they are to be measured. From the data collection plan, a model can be formulated from which inferences can be drawn.

III. *Probability is the tool used for anticipating what the distribution of data should look like under a given model*. Random phenomena are not haphazard: they display an order that emerges only in the long run and is described by a distribution. The mathematical description of variation is central to statistics. The probability required for statistical inference is to primarily axiomatic or combinatorial, but is orient toward describing data distributions.

IV. *Statistical inference guides the selection of appropriate models*. Models and data interact in statistical work: models are used to draw conclusions from data, while the data are allowed to criticize and even falsify the model through inferential and diagnostic methods. Inference from data can be thought of as the process of selecting a reasonable model, including a statement in probability language, of how confident one can be about the selection.

Topic Outline

Following is an outline of the major topics covered by the AP Statistics Examination. The ordering here is intended to define the scope of the course but not necessarily the sequence.

I. Exploring Data: Observing patterns and departures form patterns

Exploratory analysis of data makes use of graphical and numerical techniques to study patterns and departures from patterns. Emphasis should be placed on interpreting information from graphical and numerical displays and summaries.

A. Interpreting graphical displays of distribution of univariate data (dotplot, stemplot, histogram, cumulative frequency plot)
   1. Center and spread
   2. Clusters and gaps
   3. Outliers and other unusual features
   4. Shape

B. Summarizing distributions of univariate data
    1. Measuring center: median, mean
    2. Measuring spread: range, interquartile range, standard deviation
    3. Measuring position: quartiles, percentiles, standardized scores (z-scores)
    4. Using boxplots
    5. The effect of changing units on summary measures
C. Comparing distributions of univariate data (dotplots, back-to-back stemplots, parallel boxplots)
    1. Comparing center and spread: within group, between group variation
    2. Comparing clusters and gaps
    3. Comparing outliers and other unusual features
    4. Comparing shapes
D. Exploring bivariate data
    1. Analyzing patterns in scatterplots
    2. Correlation and linearity
    3. Least-squares regression line
    4. Residual plots, outliers, and influential points
    5. Transformations to achieve linearity: logarithmic and power transformations
E. Exploring categorical data: frequency tables
    1. Marginal and joint frequencies for two-way tables
    2. Conditional and relative frequencies and association

II. Planing a Study: Deciding what and how to measure

Data must be collected according to a well-developed plan if valid information on a conjecture is to be obtained. *This plan includes clarifying the question and deciding upon a method of data collection and analysis.*

A. Overview of methods of data collection
    1. Census
    2. Sample survey
    3. Experiment
    4. Observational study
B. Planning and conducting surveys
    1. Characteristics of a well-designed and well-conducted survey
    2. Populations, samples, and random selection
    3. Sources of bias in surveys
    4. Simple random sampling
    5. Stratified random sampling
C. Planning and conducting experiments
    1. Characteristics of a well-designed and well-conducted experiment
    2. Treatments, control groups, experimental units, random assignments, and replication
    3. Sources of bias and confounding, including placebo effect and blinding
    4. Completely randomized design
    5. Randomized block design, including matched pairs design
D. Generalizability of results from observational studies, experimental studies, and surveys

III. Anticipating Patterns: Producing models using probability theory and simulation

Probability is the tool used for anticipating what the distribution of data should look like under a given model.

A. Probability as relative frequency
    1. "Law of large numbers" concept
    2. Addition rule, multiplication rule, conditional probability,, and independence
    3. Discrete random variables and their probability distributions, including binomial
    4. Simulation of probability distributions, including binomial and geometric
    5. Mean (expected value) and standard deviation of a random variable, and linear transformation of a random variable
B. Combining independent random variables
    1. Notion of independence versus dependence
    2. Mean and standard deviation for sums and differences of independent random variables
C. The normal distribution
    1. Properties of the normal distribution

2. Using tables of the normal distribution
3. The normal distribution as a model for measurements

D. Sampling distributions
1. Sampling distribution of a sample proportion
2. Sampling distribution of a sample mean
3. Central Limit Theorem
4. Sampling distribution of a difference between two independent sample proportions
5. Sampling distribution of a difference between two independent sample means
6. Simulation of sampling distributions

IV. Statistical Inference: Confirming models

Statistical inference guides the selection of appropriate models.

A. Confidence intervals
1. The meaning of confidence interval
2. Large sample confidence interval for a proportion
3. Large sample confidence interval for a mean
4. Large sample confidence interval for a difference between two means (unpaired and paired)

B. Tests of significance
1. Logic of significance testing, null and alternative hypotheses; *p*-values; one- and two-sided tests; concepts of Type I and Type II errors; concept of power
2. Large sample test for a proportion
3. Large sample test for a mean
4. Large sample test for a difference between two proportions
5. Large sample test for a difference between two means (unpaired and paired)
6. Chi-square test for goodness of fit, homogeneity of proportions, and independence (one- and two-way tables)

C. Special case of normally distribute data
1. t-distribution
2. Single sample t procedures
3. Two sample (independent and matched pairs) t procedures
4. Inference for the slope of least-squares regression line