

Course Content Overview

The topics for AP Statistics are divided into four major themes: exploratory analysis (20–30 percent of the exam), planning and conducting a study (10–15 percent of the exam), probability (20–30 percent of the exam), and statistical inference (30–40 percent of the exam).

- 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 is to be obtained.* If data are to be collected to provide an answer to a question of interest, a careful plan must be developed. Both the type of analysis that is appropriate and the nature of conclusions that can be drawn from that analysis depend in a critical way on how the data was collected. Collecting data in a reasonable way, through either sampling or experimentation, is an essential step in the data analysis process.
- 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 not primarily axiomatic or combinatorial but is oriented toward using probability distributions to describe data.
- 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 Exam. The ordering here is intended to define the scope of the course but not necessarily the sequence. The percentages in parentheses for each content area indicate the coverage for that content area in the exam.

- I. Exploring Data: Describing patterns and departures from patterns (20%–30%)
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. Constructing and interpreting graphical displays of distributions 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
 1. Frequency tables and bar charts
 2. Marginal and joint frequencies for two-way tables
 3. Conditional relative frequencies and association
 4. Comparing distributions using bar charts

II. Sampling and Experimentation: Planning and conducting a study (10%–15%)

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 sampling and surveys
4. Sampling methods, including simple random sampling, stratified random sampling and cluster 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 and types of conclusions that can be drawn from observational studies, experiments and surveys

III. Anticipating Patterns: Exploring random phenomena using probability and simulation (20%–30%)

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

A. Probability

1. Interpreting probability, including long-run relative frequency interpretation
2. “Law of Large Numbers” concept
3. Addition rule, multiplication rule, conditional probability and independence
4. Discrete random variables and their probability distributions, including binomial and geometric
5. Simulation of random behavior and probability distributions
6. 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 8.3
2. Sampling distribution of a sample mean 8.2
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 8.1
7. t-distribution
8. Chi-square distribution

IV. Statistical Inference: Estimating population parameters and testing hypotheses (30%–40%)

Statistical inference guides the selection of appropriate models.

A. Estimation (point estimators and confidence intervals)

1. Estimating population parameters and margins of error
2. Properties of point estimators, including unbiasedness and variability
3. Logic of confidence intervals, meaning of confidence level and confidence intervals, and properties of confidence intervals
4. Large sample confidence interval for a proportion
5. Large sample confidence interval for a difference between two proportions

6. Confidence interval for a mean

7. Confidence interval for a difference between two means (unpaired and paired)

8. Confidence interval for the slope of a least-squares regression line

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 difference between two proportions
4. Test for a mean
5. 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)
7. Test for the slope of a least-squares regression line