Supplemental information is available at **www.collegeboard.org/ap**

Advanced Placement Course Description

Statistics

May 1999, May 2000

THE COLLEGE BOARD
Educational Excellence for All Students

Introduction

The Advanced Placement (AP®) Program consists of 32 college-level courses and exams in 18 disciplines designed for highly motivated students in secondary schools. Its exceptional reputation is made possible by the close cooperation of secondary schools, colleges, and the College Board. Nearly 2,900 universities and colleges worldwide grant credit, advanced placement, or both to students who have performed satisfactorily on the exams, and approximately 1,400 institutions grant sophomore standing to students who have demonstrated their competence in three or more of these exams.

Each course is developed by a committee made up of college faculty and AP teachers. Members of these Development Committees are appointed by the College Board and serve for overlapping three-year terms. Courses and exams are now available in the following subject areas:

Art: History of Art, Studio Art (Drawing and General)
Biology
Calculus: AB, BC
Chemistry
Computer Science: A* and AB
Economics: Macroeconomics*, Microeconomics*
English: English Language and Composition, English Literature and Composition, International English Language
Environmental Science*
French: French Language, French Literature
German Language
Government and Politics: Comparative*, U.S.*
History: European, U.S.
Latin: Literature, Vergil
Music Theory
Physics: B, C: Electricity and Magnetism*, C: Mechanics*
Psychology*
Spanish: Language, Literature
Statistics*

---

*Subjects marked with an asterisk are the equivalent of half-year college courses.

Advanced Placement is a significant force in education, especially at the secondary level. Approximately 14,000 high schools throughout the world participate in the AP Program, and in May 1997, approximately 550,000 of their students took nearly one million AP Exams.

Specific queries about the Program can be addressed to any Regional Office of the College Board or to the National Office in New York.

The AP Exams

AP Exams are offered throughout the world each May. They are administered at participating schools or at multischool centers. Any school may participate; it need only file the AP Participation Form sent to all high schools every fall. Except for Studio Art—which consists of a portfolio assessment—all exams contain a free-response section (either essay writing or problem-solving) and another section consisting of multiple-choice questions. The modern language exams also contain a performance section that includes the recording of students' responses on audiotape, and the Music Theory exam includes sight-singing tasks.

AP Statistics

The Advanced Placement Program offers a course description and examination in statistics to secondary school students who wish to complete studies equivalent to a one-semester, introductory, non-calculus-based, college course in statistics.

Statistics and mathematics educators who serve as members of the Advanced Placement Statistics Development Committee have prepared the course description and examination to reflect the content of a typical introductory college course in statistics. The examination is representative of such a course and therefore is considered appropriate for the measurement of skills and knowledge in the field of introductory statistics.

In colleges and universities, the number of students who take a statistics course is almost as large as the number of students who take a calculus course. An introductory statistics course, similar to the AP Statistics course, is typically required for majors such as social sciences, health sciences, and business. Every semester about 236,000 college and university students enroll in an introductory statistics course offered by a mathematics or statistics department. Perhaps an equal number enroll in an introductory statistics course offered by other departments. Science, engineering, and mathematics majors usually take an upper-level calculus-based course in statistics, for which the AP Statistics course is effective preparation.

The Course

The purpose of the Advanced Placement 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. Planning 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 and/or advanced placement 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 courses, could be effectively studied in a one-semester, a two-trimester, or a one-year course. Most schools, however, offer it as a two-semester course.

Who Should Take AP Statistics

The AP Statistics course is an excellent option for any secondary school student who has successfully completed a second-year course in algebra and possesses sufficient mathematical maturity and quantitative reasoning ability.

Because second-year algebra is the prerequisite course, AP Statistics usually will be taken in either the junior or senior year. The decision whether to take AP Statistics and when to take it depends on a student's plans:

- Students planning to take a science course in their senior year will benefit greatly from taking AP Statistics in their junior year.
- For students who would otherwise take no mathematics in their senior year, AP Statistics allows them to continue to develop their quantitative skills.
- Students who wish to leave open the option of taking calculus in college should include precalculus in their high school program and perhaps take AP Statistics concurrently with precalculus.

Students with the appropriate mathematical background are encouraged to take AP Statistics and AP Calculus in high school.

Admission to an Advanced Placement course ordinarily depends on the student's interest in the subject as well as on a superior academic record. However, many highly motivated students with less outstanding records have successfully completed AP courses and have obtained college credit and/or advanced placement through the Advanced Placement Examinations. Students who take the Advanced Placement Statistics course are strongly encouraged to take the examination.

AP Statistics Course Description

Course Content

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 not primarily axiomatic or combinatorial but is oriented 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.

Outline of Topics

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

I. Exploring Data: Observing patterns and departures from 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 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: frequency tables
   1. Marginal and joint frequencies for two-way tables
   2. Conditional relative frequencies and association

II. Planning 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 a 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 proportions
   5. 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 distributed 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

The Use of Technology in AP Statistics

The AP Statistics course adheres to the philosophy and methods of modern data analysis. Although the distinction between graphing calculators and computers is becoming blurred as technology advances, at present the fundamental tool of data analysis is the computer. The computer does more than eliminate the drudgery of hand computation and graphing—it is an essential tool for structured inquiry.

Data analysis is a journey of discovery. It is an iterative process that involves a dialogue between the data and a mathematical model. As more is learned about the data, the model is refined and new questions are formed. The computer aids in this journey in some essential ways. First, it produces graphs that are specifically designed for data analysis. These graphical displays make it easier to observe patterns in data, to identify important subgroups of the data, and to locate any unusual data points. Second, the computer allows the student to fit complex mathematical models to the data and to assess how well the model fits the data by examining the residuals. The result of this examination is often the need to re-express the data; for example, by taking the logarithm of a variable and repeating the analysis after re-expression. Finally, the computer is helpful in identifying an observation that has an undue influence on the analysis and in isolating its effects.

In addition to its use in data analysis, the computer facilitates the simulation approach to probability that is emphasized in the AP Statistics course. Probabilities of random events, probability distributions of random variables, and sampling distributions of statistics can be studied from a general, conceptual approach through simulation. This frees the student and teacher from the narrow approach that depends on a few simple probabilistic models.

Because the computer is central to what statisticians do, it is considered essential for teaching the AP Statistics course. It is not yet possible for students to have access to a computer during the AP Statistics Exam. Without a computer and under the conditions of a timed exam, students cannot be asked to perform the amount of computation that is needed for many statistical investigations. Consequently, standard computer output will be provided as necessary and students will be expected to interpret it. (See, for example, free-response question 4 on page 34.)

Each student will be expected to bring a graphing calculator with statistical capabilities to the examination and to be familiar with its use. The calculator's computational capabilities should include descriptive statistics such as the standard deviation, the correlation coefficient, and the equation of the least-squares linear regression line. Graphical capabilities should include the ability to make a scatterplot and to graph the least-squares linear regression line. Students find calculators where data are entered into a spread-sheet format particularly easy to use. A graphing calculator is a useful computational aid, particularly in analyzing small data sets, but should not be considered equivalent to a computer in the teaching of statistics. Ideally, students in an AP Statistics course should have access to both for work in and outside the classroom.

Instructional Emphasis in AP Statistics

The AP Statistics course lends itself naturally to a mode of teaching that engages students in constructing their own knowledge. For example, students working individually or in small groups can plan and perform data collection and analyses where the teacher serves in the role of a consultant, rather than a director. This approach gives students ample opportunity to think through problems, make decisions, and share questions and conclusions with other students as well as with the teacher.

Important components of the course should include the use of technology, projects and laboratories, cooperative group problem-solving, and writing as a part of concept-oriented instruction and assessment. This approach to teaching AP Statistics will allow students to build interdisciplinary connections with other subjects and with their world outside school.

The AP Statistics course depends heavily on the availability of technology suitable for the interactive, investigative aspects of data analysis. Therefore, schools should make every effort to provide students and teachers easy access to computers to facilitate the teaching and learning of statistics.

Providing instructional information and educational opportunities for teachers is an important component of the Advanced Placement Program. The College Board offers workshops and summer courses and institutes for teachers in all AP subjects. Further information about these and other training opportunities may be obtained from the College Board Regional Office in your area.

Additionally, the following publications provide some insight into the philosophy of the AP Statistics course.

Curriculum and Evaluation Standards for School Mathematics, The National Council of Teachers of Mathematics, Reston, Virginia, 1989.

Statistics for the Twenty-First Century, Florence and Sheldon Gordon, The Mathematical Association of America, Washington, D.C., 1992 (1-800-331-1622).

Teaching Statistics: More Data, Less Lecturing, a paper by George Cobb in *Heeding the Call for Change: Suggestions for Curricular Action*, Lynn Arthur Steen, Ed., The Mathematical Association of America, Washington, D.C., 1992 (pp. 3–43).

Formulas and Tables

Students enrolled in the AP Statistics course should concentrate their time and effort on developing a thorough understanding of the fundamental concepts of statistics. They do not need to memorize formulas.

The following list of formulas and tables will be furnished to students taking the AP Statistics Examination. Teachers are encouraged to familiarize their students with the form and notation of these formulas by making them accessible at the appropriate time during the course.

I. Descriptive Statistics

$$\bar{x} = \frac{\sum x_i}{n}$$

$$s_x = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}}$$

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

$$b_1 = r\frac{s_y}{s_x}$$

$$s_{b_1} = \frac{\sqrt{\dfrac{\sum (y_i - \hat{y}_i)^2}{n-2}}}{\sqrt{\sum (x_i - \bar{x})^2}}$$

II. Probability

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

$E(X) = \mu_x = \sum x_i p_i$

$Var(X) = \sigma_x^2 = \sum (x_i - \mu_x)^2 p_i$

If X has a binomial distribution with parameters n and p, then:

$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$

$\mu_x = np$

$\sigma_x = \sqrt{np(1 - p)}$

$\mu_{\hat{p}} = p$

$\sigma_{\hat{p}} = \sqrt{\dfrac{p(1 - p)}{n}}$

If $\bar{x}$ is the mean of a random sample of size n from an infinite population with mean μ and standard deviation σ, then:

$\mu_{\bar{x}} = \mu$

$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

III. Inferential Statistics

standardized test statistic: $\dfrac{\text{estimate} - \text{parameter}}{\text{standard deviation of the estimate}}$

confidence interval: estimate ± (critical value) • (standard deviation of the estimate)

Single Sample

statistic	standard deviation
mean	$\dfrac{\sigma}{\sqrt{n}}$
proportion	$\sqrt{\dfrac{p(1-p)}{n}}$

Two Sample

statistic	standard deviation
difference of means (unequal variances)	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$
difference of means (equal variances)	$\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$
difference of proportions (unequal variances)	$\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$
difference of proportions (equal variances)	$\sqrt{p(1-p)}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$

$$\text{Chi-square test statistic} = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Table entry for z is the probability lying below z.

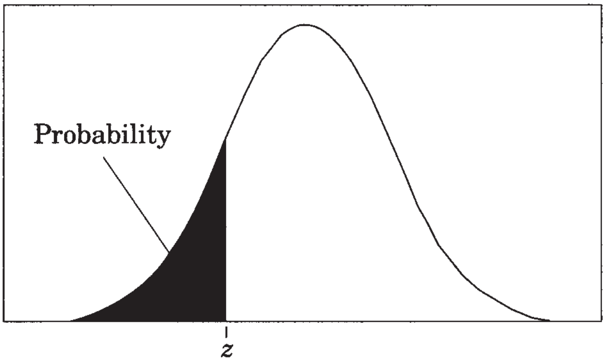


Table A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
–3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
–3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
–3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
–3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
–3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
–2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
–2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
–2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
–2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
–2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
–2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
–2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
–2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
–2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
–2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
–1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
–1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
–1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
–1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
–1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
–1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
–1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
–1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
–1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
–1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
–0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
–0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
–0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
–0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
–0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
–0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
–0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
–0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
–0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
–0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Table entry for z is the probability lying below z.

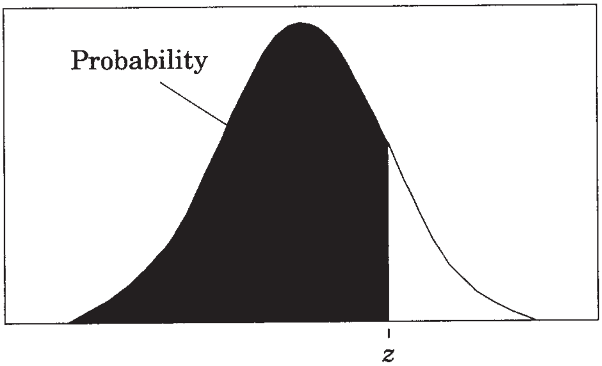


Table A *(Continued)*

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Table entry for *p* and *C* is the point *t** with probability *p* lying above it and probability *C* lying between -*t** and *t**.

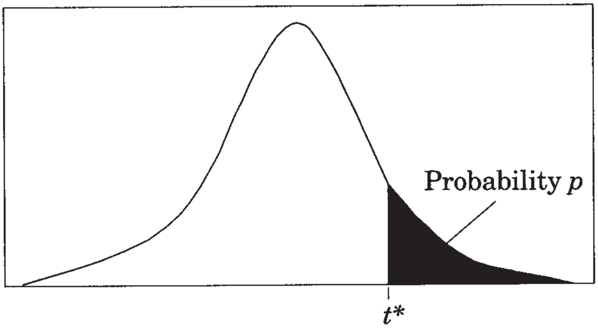


Table B *t* distribution critical values

	Tail probability *p*											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3,435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level *C*											

Table entry for p is the point (χ^2) with probability p lying above it.

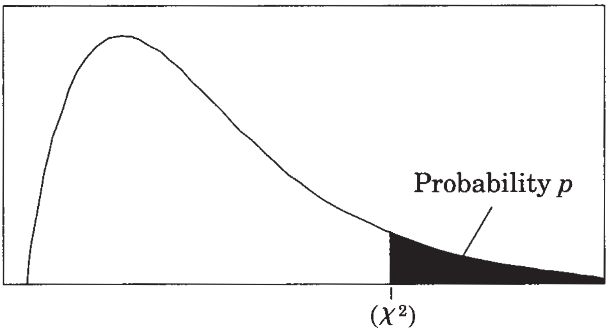


Table C χ^2 critical values

df	Tail probability p										
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4

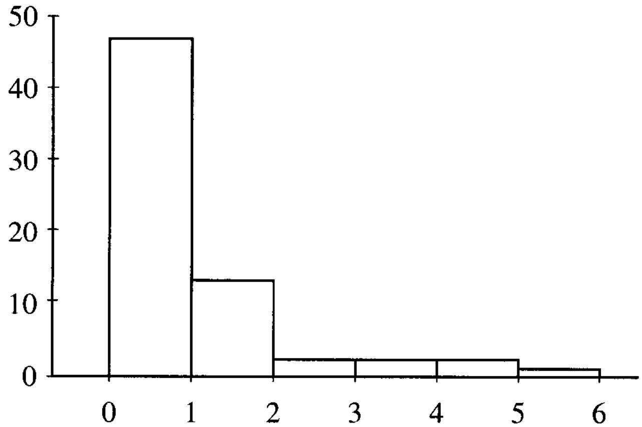
The Examination

The AP Statistics Examination is three hours long and seeks to determine how well a student has mastered the concepts and techniques of the subject matter of the course. This paper-and-pencil examination consists of (1) a multiple-choice section testing proficiency in a wide variety of topics, and (2) a free-response section requiring the student to answer open-ended questions and to complete an investigative task involving more extended reasoning. In the determination of the grade for the examination, the two sections will be given equal weight. Each student will be expected to bring a graphing calculator with statistical capabilities to the examination. Minicomputers, pocket organizers, electronic writing pads (e.g., Newton), and calculators with QWERTY (i.e., typewriter) keyboards will not be allowed. Calculator memories will not be cleared. However, calculator memories may be used only for storing programs, not for storing notes. A student may bring up to two calculators to the examination.

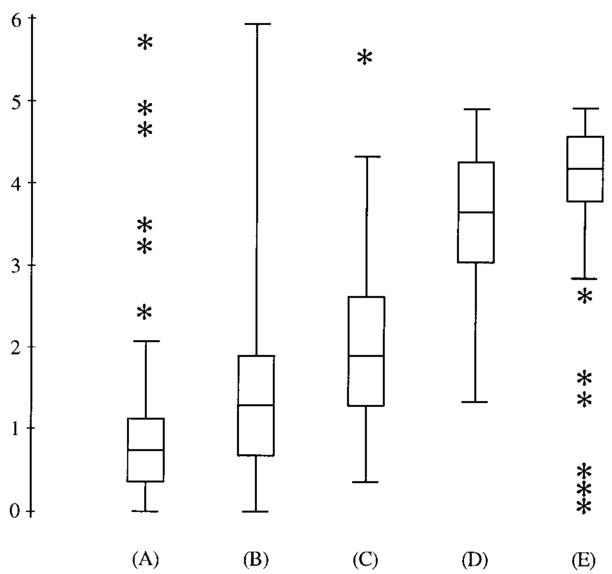
Multiple-Choice Questions

The following are examples of the kinds of multiple-choice questions found on the AP Statistics Examination. The distribution of topics and the levels of difficulty are illustrative of the composition of the examination. Students often ask whether they should guess on the multiple-choice section. Haphazard or random guessing is unlikely to improve scores, because one-fourth of the number of questions answered incorrectly will be subtracted from the number of questions answered correctly. However, students who have some knowledge of a question and can eliminate one or more answer choices will usually find it advantageous to guess from among the remaining choices. An answer key to the multiple-choice questions can be found on page 41.

Directions: Each of the questions or incomplete statements is followed by five suggested answers or completions. Select the one that best answers the question or completes the statement.



1. The histogram above displays a set of measurements. Which of the boxplots below displays the same set of measurements?



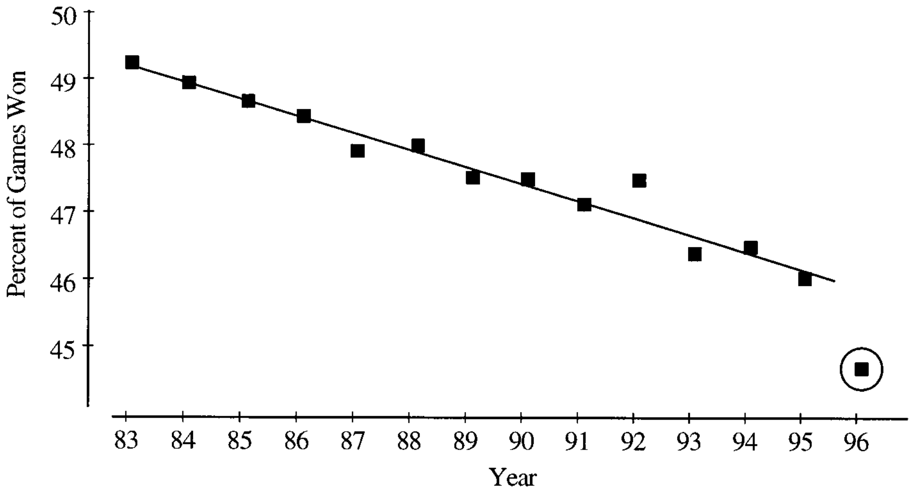
2. A random sample of size 10 was taken from a population. The sample has a variance of zero. Which of the following statements must be true?

I. The population also has a variance of zero.
II. The sample mean is equal to the sample median.
III. The ten data points in the sample are equal in numerical value.

(A) I only
(B) II only
(C) III only
(D) I and II
(E) II and III

3. A professor teaches two statistics classes. The morning class has 25 students and their average on the first test was 82. The evening class has 15 students and their average on the same test was 74. What is the average on this test if the professor combines the scores for both classes?

(A) 76
(B) 78
(C) 79
(D) 80
(E) The average cannot be calculated since individual scores of each student are not available.



4. As shown above, a least-squares regression line has been fitted to the winning percentages for a local sports team in each of the years 1983 through 1995. The percentage for the 1996 season was then plotted (as circled above). Which of the following statements correctly describes how the value for the 1996 season will change the appearance of the least-squares regression line and the correlation coefficient if a new least-squares regression line is fitted to the 1983 through 1996 data?

(A) The 1996 point will make the least-squares regression line steeper and the correlation coefficient stronger.
(B) The 1996 point will make the least-squares regression line steeper and the correlation coefficient weaker.
(C) The 1996 point will make the least-squares regression line closer to horizontal and the correlation coefficient stronger.
(D) The 1996 point will make the least-squares regression line closer to horizontal and the correlation coefficient weaker.
(E) The 1996 point will not have any effect on the least-squares regression line since it follows the same downward trend.

x	$P(X = x)$
1	$\frac{1}{6}$
2	$\frac{2}{3}$
3	?

y	$P(Y = y)$
1	?
2	$\frac{1}{4}$
3	$\frac{1}{4}$
4	?

5. The tables above show part of the probability distribution for random variables X and Y. If X and Y are independent and the joint probability $P(X = 3, Y = 4) = \frac{1}{16}$, then $P(Y = 1) =$

(A) $\frac{1}{8}$

(B) $\frac{1}{6}$

(C) $\frac{1}{4}$

(D) $\frac{3}{8}$

(E) $\frac{1}{2}$

6. For college-bound high school seniors in 1996, the nationwide mean SAT verbal score was 505 with a standard deviation of about 110, and the mean SAT math score was 508 with a standard deviation of about 110. Students who do well on the verbal portion of the SAT tend to do well on the mathematics portion. If the two scores for each student are added, the mean of the combined scores is 1,013. What is the standard deviation of the combined verbal and math scores?

(A) $\frac{110}{\sqrt{2}}$ (approximately 77.78)

(B) 110

(C) $\sqrt{110^2 + 110^2}$ (approximately 155.56)

(D) 220

(E) The standard deviation cannot be computed from the information given.

7. A random sample of two observations is taken from a population that is normally distributed with a mean of 100 and a standard deviation of 5. Which of the following is closest to the probability that the sum of the two observations is greater than 221?

(A) 0.0015
(B) 0.0250
(C) 0.0500
(D) 0.4500
(E) 0.9985

8. A particular psychological test is used to measure academic motivation. The average test score for all female college students nationwide is 115. A large university estimates the mean test score for female students on its campus by testing a random sample of n female students and constructing a confidence interval based on their scores.

Which of the following statements about the confidence interval are true?

I. The resulting interval will contain 115.
II. The 95 percent confidence interval for $n = 100$ will generally be shorter than the 95 percent confidence interval for $n = 50$.
III. For $n = 100$, the 95 percent confidence interval will be longer than the 90 percent confidence interval.

(A) I only
(B) II only
(C) III only
(D) II and III
(E) None of the above gives the complete set of true responses.

9. A survey was conducted at a movie theater to determine movie-goers' preference for different kinds of popcorn. The results of the survey showed that Brand A was preferred by 65 percent of the people with a margin of error of plus or minus 3 percent. What is meant by the statement "plus or minus 3 percent"?

(A) Three percent of the population that was surveyed will change their minds.
(B) Three percent of the time the results of such a survey are not accurate.
(C) Three percent of the population was surveyed.
(D) The true proportion of the population who preferred Brand A popcorn could be determined if 3 percent more of the population was surveyed.
(E) It would be unlikely to get the observed sample proportion of 65 percent unless the actual percentage of people in the population of movie-goers who prefer Brand A is between 62 percent and 68 percent.

10. When performing a test of significance for a null hypothesis, H_0, against an alternative hypothesis, H_a, the p-value is

(A) the probability that H_0 is true
(B) the probability that H_a is true
(C) the probability that H_0 is false
(D) the probability of observing a value of a test statistic at least as extreme as that observed in the sample if H_0 is true
(E) the probability of observing a value of a test statistic at least as extreme as that observed in the sample if H_a is true

11. Twenty men and 20 women with high blood pressure were subjects in an experiment to determine the effectiveness of a new drug in lowering blood pressure. Ten of the 20 men and 10 of the 20 women were chosen at random to receive the new drug. The remaining 10 men and 10 women received a placebo. The change in blood pressure was measured for each subject. The design of this experiment is

(A) completely randomized with one factor, drug
(B) completely randomized with one factor, gender
(C) randomized block, blocked by drug and gender
(D) randomized block, blocked by drug
(E) randomized block, blocked by gender

12. A large elementary school has 15 classrooms, with 24 children in each classroom. A sample of 30 children is chosen by the following procedure.

Each of the 15 teachers selects 2 children from his or her classroom to be in the sample by numbering the children from 1 to 24, then using a random digit table to select two different random numbers between 01 and 24. The 2 children with those numbers are in the sample.

Did this procedure give a simple random sample of 30 children from the elementary school?

(A) No, because the teachers were not selected randomly.
(B) No, because not all possible groups of 30 children had the same chance of being chosen.
(C) No, because not all children had the same chance of being chosen.
(D) Yes, because each child had the same chance of being chosen.
(E) Yes, because the numbers were assigned randomly to the children.

13. The primary reason for using blocking when designing an experiment is to reduce

(A) the sensitivity of the experiment
(B) variation
(C) the need for randomization
(D) bias
(E) confounding

(See page 41 for the answer key to these multiple-choice questions.)

Free-Response Questions

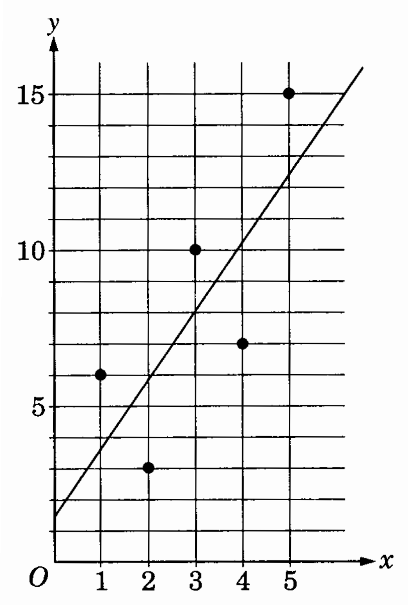
In the free-response section of the AP Statistics Examination, students are asked to answer open-ended questions and to complete an investigative task. Each open-ended question is designed to be answered in approximately 10 minutes, and each examination will contain four to seven of these questions. The longer investigative task is designed to be answered in approximately 30 minutes, and each examination will contain one of these tasks. The questions require students to relate different content areas as they formulate a complete solution to a statistics or probability problem. Students are expected to use their analytical and organizational skills to formulate cogent answers in writing their responses. Students will, as always, be expected to show enough of their work for readers to follow their line of reasoning. (It is not necessary to write out routine arithmetical calculations that can be done on a calculator.) To obtain full credit for a free-response question, students must analyze the situation completely and communicate their analyses and conclusions clearly. Answers should show enough work so that the reasoning process can be tracked throughout the analysis. This is particularly important for assessing partial credit.

Solutions to free-response questions will be graded on how appropriately the student analyzes the situation and on how clearly and completely the student communicates his or her reasoning in each step of that analysis. In the response, the student should identify the important components and assumptions of the problem, synthesize logical and correct relationships among those components using appropriate statistical principles and techniques (including plots when warranted), and state the conclusions in the context of the original problem, with necessary caveats. Answers to statistical inference questions are expected to show knowledge of assumptions and an ability to check assumptions to the extent that the given information will allow. Although correct arithmetic is of secondary importance to clear communication of knowledge of statistical principles, a student should recognize when a numerical answer is unreasonable.

Following are four sample open-ended questions and two sample investigative tasks.

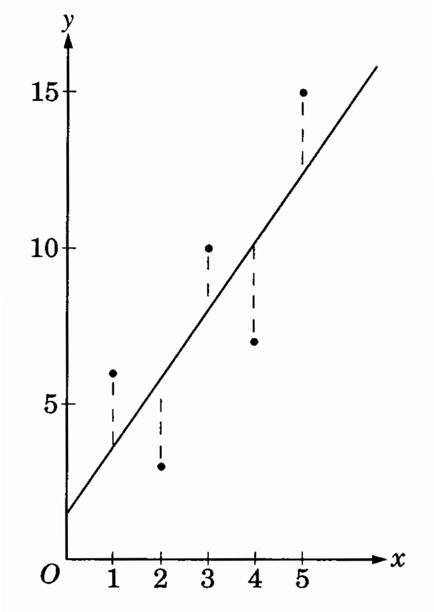
Directions: Show all your work. Indicate clearly the methods you use because you will be graded on the correctness of your methods as well as the accuracy of your final answers.

Four Open-Ended Questions



1. The regression line $\hat{y} = 1.6 + 2.2x$ for the five points on the scatterplot above was computed using the method of least squares. Use the above scatterplot to demonstrate the meaning of the term "least squares."

One Possible Solution to Question 1



The least-squares regression line is used as a model that predicts the value of y when x is given. Since we think of x as being fixed, we measure the error in the prediction of y as the vertical distance from the point to the regression line. These errors (residuals) are the lengths of the dotted lines on the scatterplot above.

The method of least squares finds the line that makes the sum of the squares of these errors as small as possible.

For the given points, the sum of the squared errors from the line $\hat{y} = 1.6 + 2.2x$ is

$[6 - (2.2(1) + 1.6)]^2 + [3 - (2.2(2) + 1.6)]^2 + [10 - (2.2(3) + 1.6)]^2 + [7 - (2.2(4) + 1.6)]^2 + [15 - (2.2(5) + 1.6)]^2 = 34.4$

No other line gives a smaller sum than 34.4.

(Note: An acceptable response for this question need not show any calculations if the explanation is adequate.)

2. Since most people are right-handed, a company that designs machinery has traditionally placed the controls that demand the most hand strength so that they will be used by the right hand. The company decides that it should test its assumption that the right hand of right-handed adults tends to be stronger than the left hand. Nine right-handed adults are randomly selected from employees of this company for the test. Hand strength is measured by using a calibrated hand gripper. It is reasonable to assume that the difference in people's left-hand and right-hand strength is approximately normally distributed. The people test their left hand first and then their right hand. The hand strengths, in kilograms, for each person and the summary statistics are given below.

Person	1	2	3	4	5	6	7	8	9
Right hand	11.7	12.7	11.4	12.4	12.2	11.3	11.1	11.7	11.7
Left hand	11.6	11.2	10.6	11.2	10.9	10.9	10.7	10.3	10.5
Difference	0.1	1.5	0.8	1.2	1.3	0.4	0.4	1.4	1.2

SUMMARY STATISTICS

	Right-hand Strength	Left-hand Strength	Right-hand Strength Minus Left-hand Strength
Mean	11.8	10.9	0.9
Standard Deviation	0.5	0.4	0.5
Variance	0.28	0.16	0.26

Do these data support the conclusion that the right hand of right-handed adults tends to be stronger than the left hand? Give an appropriate statistical justification. Briefly discuss any concerns you have about the design of this study.

One Possible Solution to Question 2

Yes, $t = \frac{(0.9)(\sqrt{9})}{0.5} = 5.4$ with $df = 8$, which is significant at the 0.0005 level.

The matched pairs t-test is appropriate because each right hand is matched with the left hand for that adult and the subjects were randomly selected. We can assume the population of differences is normally distributed. Therefore, it is highly unlikely that this difference in hand strength is due to chance variation alone.

The design of the study may be improved by randomizing the order of which hand grips the hand gripper first or by using a more accurate measuring device.

Another Possible Solution to Question 2

Let the null hypothesis be $p = 0.5$, where p is the proportion of right-handed adults whose right-hand strength is greater than their left-hand strength.

Compute the probability that, in a population with probability $p = 0.5$ of a success, we would get 9 successes in a random sample of size 9. The probability of 9 successes under the null hypothesis is the same as the probability of flipping a coin 9 times and getting 9 heads: $(0.5)^9 = 0.00195$. We reject the null hypothesis. It is highly unlikely that all 9 people in the random sample would have the right-hand strength greater than the left-hand strength if those two events are equally likely in the population.

Randomize the order of which hand grips the hand gripper first. Use other tests of hand strength. Use a more accurate measuring device.

3. Cuckoos lay their eggs in the nests of other birds. Some biologists speculate that the size of the cuckoo's eggs might be different depending on whether the eggs are laid in warblers' nests or wrens' nests. To check this, biologists searched a wildlife refuge for warblers' and wrens' nests. Summary statistics for the lengths (in mm) of cuckoos' eggs found in the nests of warblers and wrens are shown below.

   Eggs from warblers' nests: $n_1 = 29$, $\bar{x}_1 = 22.20$, $s_1 = 0.65$
   Eggs from wrens' nests: $n_2 = 35$, $\bar{x}_2 = 21.12$, $s_2 = 0.75$

   Use a 99 percent confidence interval to determine if these data support the biologists' speculation that the size of the eggs differs depending on whether they are laid in warblers' nests or wrens' nests.

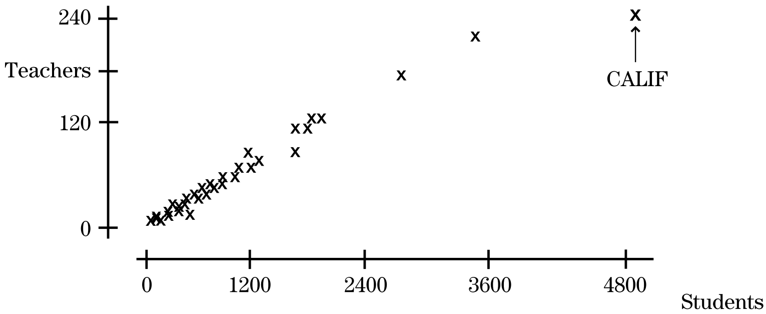
One Possible Solution to Question 3

The data support the biologists' speculation because zero is not in the confidence interval. If the difference between the mean lengths of the eggs in the entire population of eggs was actually zero, then there is less than one chance in a hundred that the 99 percent confidence interval would not contain zero. Therefore we conclude that the difference probably is not zero.

$$(\bar{x}_1 - \bar{x}_2) \pm t\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = (22.20 - 21.12) \pm 2.763\sqrt{\frac{0.65^2}{29} + \frac{0.75^2}{35}}$$

$$= 1.08 \pm 0.484$$
$$= (0.596,\ 1.564)$$

4. The following scatterplot shows the number of teachers (in thousands) for each of the states plus the District of Columbia plotted against the number of students (in thousands) enrolled in kindergarten through grade 12 for those states and the District of Columbia. The computer printout under the plot shows the results of fitting a straight line to the data by the method of least squares.



The regression equation is
Teachers = 4.49 + 0.0534 Students

Predictor	Coef	Stdev	t-ratio	p
Constant	4.486	2.025	2.22	0.031
Students	0.053401	0.001692	31.57	0.000

s = 10.67 R-sq = 95.3%

a. Find the correlation coefficient for the relationship between the number of teachers and the number of students on the basis of these data.
b. Provide an interpretation of the slope of the least-squares regression line in the context of these data.
c. How will the size of the correlation coefficient change if California is removed from the data set and a new correlation coefficient is calculated? (The data point for California is labeled on the scatterplot.) Explain your answer.
d. How will the size of the slope of the least-squares regression line change if California is removed from the data set and a new least-squares regression line is calculated? Explain your answer.

One Possible Solution to Question 4

a. In the case of linear models, the correlation coefficient is plus or minus the square root of R^2 written as a decimal. Because the slope of the line is positive, $r = 0.976$.
b. The slope is the expected change in y per unit change in x. As the number of students increases by 1 unit, which is 1,000 students, the number of teachers tends to change by 0.0534 thousand, or 53.4.
c. If California is removed from the data set, the line will swing upward, and the points will fit the line much better. Thus, the correlation coefficient will increase.
d. As in part c, if California is removed, the line will swing upward and the slope will increase. Notice that California has the effect of lowering the teacher-to-student ratio (or increasing the student-to-teacher ratio) if it is left in the data set.

The idea behind this question is to show that students should have experience in fitting lines to data and observing the effect of points that lie outside the "cloud" represented by most of the points in the scatterplot. The easiest way to do this is to use a computer program for plotting and fitting regression models, which makes it very easy to adjust the data and fit multiple models. By removing California from the data set, it can be shown that the correlation coefficient increases to 0.992 and the slope increases to 0.0625.

Two Investigative Tasks

1. A company has 11 mathematicians on its staff, of whom 3 are women. The president of the company is concerned about the small number of women mathematicians. The president learns that about 40 percent of the mathematicians in the United States are women, and asks you to investigate whether or not the number of women mathematicians in the company is consistent with the national pool. The president knows very little about statistics.

   You decide to answer the president's question using simulation. You will use the Table of Random Digits to simulate the situation of hiring 11 randomly selected mathematicians (and counting the number of women hired).

   Note: A Table of Random Digits will be provided as a part of any question on the examination where it is needed.

   a. Describe how you will use the table to perform one trial of your simulation. Include a description of what each of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9 will represent in your simulation.
   b. Perform your simulation 20 times. Start at the left-most digit in the first row of the table and move across. Make your procedure clear so that someone can follow what you did. You must do this by marking directly on or above the table.
   c. Place the results of your 20 simulations in a frequency table that shows the number of women mathematicians hired.
   d. On the basis of your frequency table, estimate the probability that there are 3 or fewer women in a randomly selected group of 11 mathematicians.
   e. On the basis of your simulation, write a report to the president of the company.

One Possible Solution to Task 1

a. Since about 40 percent of the mathematicians are women, I will let digits 0, 1, 2, and 3 represent women. The remaining digits will represent men. I will select 11 digits from the random digit table and count the number of digits that are 0, 1, 2, and 3. This number represents the number of women selected out of a random sample of 11 mathematicians.

b. Starting at the left of the first line of the table below and working across, out of the first 11 digits, only one represents a woman. That digit is the first underlined 3. I will continue in this way with 19 more groups of 11 digits. The groups are separated by slash marks. The number of digits that represent women is written above each group.

	1		4		6	
48747	76595	3/2588	38392	84/422	80016	378/90
5		5			3	
71950	2249/4	00369	61269/	87073	73694	9/7751
3		8		6		4
17857	52/352	21392	229/30	43776	1050/3	58249
		5		5		4
80993/	52010	88856	2/3882	73613	57/648	47051
	5		4			4
630/16	73572	2268/4	02409	37565/	52457	01257
	6		6		2	
4/0615	63910	09/596	10241	034/13	77576	7487/2
6						
57431	29251/	77848	98037	81230	38561	69580
06181	97842	48327	37976	81333	10264	77769

c.

Number of Women	Frequency
0	0
1	1
2	1
3	2
4	5
5	5
6	5
7	0
8	1

d. From this frequency table, the estimated probability that there are 3 or fewer women in a randomly selected group of 11 mathematicians is 4/20, or 20 percent.

e. Dear President Smith:

As you know, 40 percent of the mathematicians in the United States are women, yet in your company only 3/11, or 27 percent of the mathematicians, are women. I conducted a simulation of what would happen if 20 companies each hired 11 mathematicians chosen at random from the national pool. In my simulation, 4 out of the 20 companies, or 20 percent, hired 3 or fewer women out of the 11 mathematicians hired, while the remaining 16 companies, or 80 percent, hired 4 or more women. Therefore, if mathematicians were hired at random from a national pool containing 40 percent women, we would expect to see 3 or fewer women in a staff of 11 mathematicians about 20 percent of the time. Hence, although your proportion of women mathematicians is low, it is not inconsistent with what would be expected by chance. If this trend continues with future hires, the numbers may fall below what is reasonably expected by chance, so it is worth monitoring your company's hiring practices.

Sincerely,

Janet Jones
Statistician

CALCULUS AB

	Examination Grade	1986 Number of Students	1986 %	1987 Number of Students	1987 %
Extremely well qualified	5	6,676	16.44	37	18.50
Well qualified	4	9,164	22.57	28	14.00
Qualified	3	11,766	28.98	68	34.00
Possibly qualified	2	7,095	17.47	35	17.50
No recommendation	1	5,904	14.54	32	16.00
Total Number of Students		40,605		200	
Mean Grade		3.09		3.02	
Standard Deviation		1.28		1.32	
Number of Schools		4,275		Not Available	
Number of Colleges		1,207		Not Available	

2. The table shown above provides information about the scores on the 1986 and 1987 Calculus AB examinations. The data for 1986 were obtained from all students who took the exam. The data for 1987 were obtained from a random sample of 200 students. The data from this sample will be used to predict the performance of all students who took the 1987 Calculus AB examination.

   a. Do these data provide evidence of a change in the distribution of grades from 1986 to 1987? Give an appropriate statistical justification to support your conclusion.
   b. Is there evidence that there is a change in the mean grade from 1986 to 1987? Give appropriate statistical evidence to support your conclusion.

One Possible Solution to Task 2

a. Let the null hypothesis be that the distribution of grades is the same in 1987 as it was in 1986.

Grade	Observed in 1987 Sample: O	Expected in 1987 Sample: E	$\frac{(O - E)^2}{E}$
5	37	32.88	0.516
4	28	45.14	6.508
3	68	57.96	1.739
2	35	34.94	0.000
1	32	29.08	0.293
Total	200	200	$\chi^2 = 9.06$

From the chart with 4 degrees of freedom and 5 percent level of significance, the critical value of χ^2 is 9.49. The *p*-value is between .05 and .10. This indicates there is borderline evidence that the distribution of 1987 grades differs from the distribution of 1986 grades. The difference appears to be that there are fewer grade 4's in 1987.

b. The 95 percent confidence interval for the mean grade of all Calculus AB examinations in 1987 is

$$3.02 \pm 1.96\left(\frac{1.32}{\sqrt{200}}\right) = 3.02 \pm 0.18,$$

$$= (2.84, 3.20)$$

No, there is no statistical evidence of a change in the mean grade. The mean grade for all 1986 examinations was 3.09. Since this mean is in the confidence interval above, we cannot reject the possibility that the mean grade for all of the 1987 examinations will turn out to be 3.09. In other words, if the mean for all of the 1987 examinations is really 3.09, the same as in 1986, it would not be unlikely to get a mean of 3.02 in a sample of 200 examinations from 1987.

Answer Key to Multiple-Choice Questions

1–A, 2–E, 3–C, 4–B, 5–A, 6–E, 7–A, 8–D, 9–E, 10–D, 11–E, 12–B, 13–B

Distribution of Grades for All 1997 AP Statistics Candidates

	Examination Grade	*Percentage Earning Grade*
Extremely Well Qualified	5	15.7
Well Qualified	4	22.1
Qualified	3	24.4
Possibly Qualified	2	19.7
No Recommendation	1	18.0

Mean Grade	2.98
Standard Deviation	1.33
Number of Candidates	7,667
Number of Schools	752
Number of Colleges Receiving AP Grade Reports	748

1999 Exam Date: Tuesday, May 18, afternoon session

1997-98 Development Committee and Chief Faculty Consultant in Statistics

Ann E. Watkins, California State University: Northridge, *Chair*
Fred C. Djang, Choate Rosemary Hall, Wallingford, Connecticut
Katherine Taylor Halvorsen, Smith College, Northampton, Massachusetts
Christopher R. Olsen, George Washington High School, Cedar Rapids, Iowa
Diann C. Resnick, Bellaire Senior High School, Houston, Texas
Daniel J. Teague, North Carolina School of Science and Math, Durham
Jessica M. Utts, University of California: Davis
Walter O. Walker, Eckerd College, St. Petersburg, Florida
Chief Faculty Consultant: Richard L. Scheaffer, University of Florida, Gainesville
Chief Faculty Consultant-Designate: Roxy Peck, California Polytechnic State University: San Luis Obispo
ETS Consultants: Jeff Haberstroh, Michael A. Ponisciak

I.N. 201704