AP® Statistics
2003 Sample Student Responses
Form B

These materials were produced by Educational Testing Service® (ETS®), which develops and administers the examinations of the Advanced Placement Program for the College Board. The College Board and Educational Testing Service (ETS) are dedicated to the principle of equal opportunity, and their programs, services, and employment policies are guided by that principle.

The College Board is a national nonprofit membership association whose mission is to prepare, inspire, and connect students to college and opportunity. Founded in 1900, the association is composed of more than 4,300 schools, colleges, universities, and other educational organizations. Each year, the College Board serves over three million students and their parents, 22,000 high schools, and 3,500 colleges through major programs and services in college admissions, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT®, the PSAT/NMSQT®, and the Advanced Placement Program® (AP®). The College Board is committed to the principles of equity and excellence, and that commitment is embodied in all of its programs, services, activities, and concerns.

For further information, visit www.collegeboard.com

For the College Board's online home for AP professionals, visit AP Central at apcentral.collegeboard.com.

A

(a) Suppose that at the end of the study, 10 percent of the females in the 40-44 age-group contracted the illness. Calculate a 95 percent confidence interval to estimate the population proportion of females in this age-group that contracted the illness.

The use of 95 % confidence test is justified because the sample is an SRS and $np = 37 (> 10)$ and $n(1-p) = 333 (> 10)$ The population is at least 10 times the sample.

$p = .1$ $\quad z^* = InvNorm(.975) = 1.95996$

$\hat{p} \pm z^* \sqrt{\frac{p(1-p)}{n}}$ $\quad .1 \pm 1.96 \sqrt{\frac{.1(.9)}{370}} = .1 \pm .0306$

$(.069, .131)$

We are 95 % confident that the true proportion of females between 40 and 44 years of age contracting the illness lies between .069 and .131.

Interpret this confidence interval in the context of this situation.

We are 95 % confident that the true proportion of females between age 40 and 44 contracting the illness lies between .069 and .131.

Interpret the confidence level of 95 percent.

In the long run, 95 % of all the possible samples taken from the population will present intervals that would include the true parameter. (in this case, the proportion of people that become ill.)

GO ON TO THE NEXT PAGE.

(b) Suppose that at the end of the study, 10 percent of the males in the 40-44 age-group contracted the illness. The corresponding 95 percent confidence interval to estimate the population proportion of males in this age-group that contracted the illness is (0.061, 0.139).

Note that this interval and the interval in part (a) are of different lengths even though the two sample proportions were identical. What would be an alternative way to allocate a sample of 2,000 subjects so that the 95 percent confidence interval widths for all male age-groups and for all female age-groups (i.e., for all 8 groups) would be the same when the sample proportions are the same? Justify your answer.

The length of confidence interval depends on the sample size, when all other variables are constant. $p \pm 2*\sqrt{\frac{p(1-p)}{n}}$ n, or sample size is inversely proportional to the interval length. An alternative way to allocate 2000 subjects with same C. interval lengths would be to make each of the 8 groups contain the same number of subjects, 250. Stratification method is needed here. The population should be divided into each age and sex group, and then SRS of 250 subjects should be taken.

(c) Based on previous studies, researchers believe that the percentages of those who contract the illness will be similar for males and females, and therefore plan to ignore gender when selecting a sample for this study. Previous studies also indicate that the percentages of adults who will contract this illness in the 35-39, 40-44, 45-49, and 50-54 age-groups are anticipated to be 5%, 8%, 20%, and 35%, respectively. How should the sample of 2,000 subjects be allocated with respect to age-groups so that the widths of the 95 percent confidence intervals for the four groups will be approximately the same? Justify your answer.

The length of the confidence interval is also dependent on the proportion. $\hat{p} \pm 2*\sqrt{\frac{p(1-p)}{n}}$ As p increases, the margin of error increases. Therefore, if the proportion of sickness increases as the age increases, the sample size should also increase in order to balance out the increase in m. (larger n makes margin of error smaller)

	fraction of total $p(1-p)$		Δ
.05 . .95 = .0475	.0933	.15056 →	186.6
.08 . .92 = .0736	.14471	→	287.42
.2 x .8 = .16	.31458	→	629.16
.35 . .65 = .2275	.4473	→	874.6

The 2000 subjects should be allocated as follows:

	35-39	40-44	45-49	50-54
n (sample size)	187	289	629	895

B

STATISTICS

SECTION II

Part B

Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II grade—25—

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 on the accuracy of your results and explanation.

6. Researchers at a large health maintenance organization (HMO) are planning a study of a certain mild illness. They will select a random sample of patients who are ages 35 to 54 and see if they contract the illness in the next year. The researchers are interested in estimating the proportions of men and of women who are likely to develop the illness in each of 4 age-groups: 35-39, 40-44, 45-49, and 50-54.

   The researchers plan to include 2,000 patients in the study. Suppose the researchers draw a random sample from all of the patients at this HMO who are ages 35 to 54 and find the following numbers within each gender and age-group.

	Age-Group			
	35-39	40-44	45-49	50-54
Male	350	230	150	60
Female	445	370	245	150

.175 .115 .08 .03

GO ON TO THE NEXT PAGE.

(b) Suppose that at the end of the study, 10 percent of the males in the 40-44 age-group contracted the illness. The corresponding 95 percent confidence interval to estimate the population proportion of males in this age-group that contracted the illness is (0.061, 0.139).

Note that this interval and the interval in part (a) are of different lengths even though the two sample proportions were identical. What would be an alternative way to allocate a sample of 2,000 subjects so that the 95 percent confidence interval widths for all male age-groups and for all female age-groups (i.e., for all 8 groups) would be the same when the sample proportions are the same? Justify your answer.

We will have to have the same amount of people in each group.

for males: $.1 \pm 1.96\sqrt{\frac{(.9)(.1)}{250}}$

for females: $.11 \pm 1.96\sqrt{\frac{(.89)(.11)}{250}}$

The only difference in the above equation is only the denominator. So if we had 250 (2000/8) people in each group we will have same length intervals.

(c) Based on previous studies, researchers believe that the percentages of those who contract the illness will be similar for males and females, and therefore plan to ignore gender when selecting a sample for this study. Previous studies also indicate that the percentages of adults who will contract this illness in the 35-39, 40-44, 45-49, and 50-54 age-groups are anticipated to be 5%, 8%, 20%, and 35%, respectively. How should the sample of 2,000 subjects be allocated with respect to age-groups so that the widths of the 95 percent confidence intervals for the four groups will be approximately the same? Justify your answer.

maximum of error: $2^* \sqrt{\frac{pq}{n}}$

$\sqrt{\frac{(.9)(.1)}{250}} = .01897367$

$.01897367 = \sqrt{\frac{(.05)(.95)}{n}}$ $\quad .01897367 = \sqrt{\frac{(.08)(.92)}{n}}$

$n_1 = 131.9$ $\quad$ " $n_2 = 204.4$

$.01897367 = \sqrt{\frac{(.2)(.8)}{n}}$ $\quad .01897367 = \sqrt{\frac{(.35)(.65)}{n}}$

$n_3 = 444.4$ $\quad n_4 = 631.9$

2000 / total of n_i : $\quad 2000/1412.7 = 1.415695$

$n_1 (35-39) = 186.7 \approx 187$ people

$n_2 (40-44) = 289.3 \approx 289$ people $\quad q$ = constant

$n_3 (45-49) = 629.1 \approx 629$ people

$n_4 (50-54) = 294.6 \approx 295$ people

all 95% conf int (illegible) approx

GO ON TO THE NEXT PAGE.

(a) Suppose that at the end of the study, 10 percent of the females in the 40-44 age-group contracted the illness. Calculate a 95 percent confidence interval to estimate the population proportion of females in this age-group that contracted the illness.

$p = .10$ contract illness (40-44)

$$0.1 \pm z^* \left(\sqrt{\frac{pq}{n}}\right)$$

$$0.1 \pm 1.96 \left(\sqrt{\frac{(.1)(.9)}{270}}\right)$$

$$0.1 \pm 1.96 (.01559625 73)$$

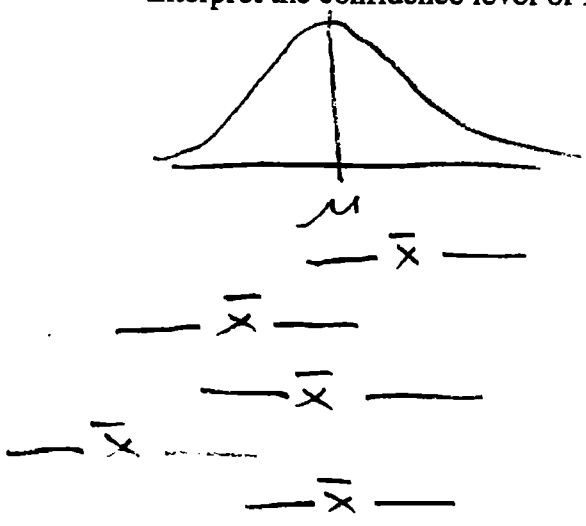
$$= 0.1 \pm .03057$$

$$[.06943, .13057]$$

Interpret this confidence interval in the context of this situation.

About 6.943 % to 13.057 % of the population of females in the age group of 40 - 44 will contract this illness.

Interpret the confidence level of 95 percent.



If we were to take an infinitely many samples and used the same margin of error, we would capture the true μ 95% of the time, As shown on the diagram on the left.

GO ON TO THE NEXT PAGE.