CollegeBoard
Advanced Placement Program

AP® Statistics
2005 Sample Student Responses
Form B

The College Board: Connecting Students to College Success

The College Board is a not-for-profit membership association whose mission is to connect students to college success and opportunity. Founded in 1900, the association is composed of more than 4,700 schools, colleges, universities, and other educational organizations. Each year, the College Board serves over three and a half million students and their parents, 23,000 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 excellence and equity, and that commitment is embodied in all of its programs, services, activities, and concerns.

Visit the College Board on the Web: www.collegeboard.com.
AP Central is the official online home for the AP Program and Pre-AP: apcentral.collegeboard.com.

1A

STATISTICS
SECTION II
Part A
Questions 1-5
Spend about 65 minutes on this part of the exam.
Percent of Section II grade—75

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.

1. The graph below displays the scores of 32 students on a recent exam. Scores on this exam ranged from 64 to 95 points.

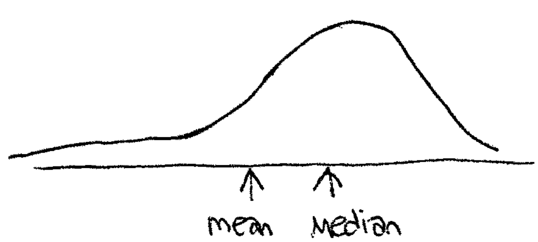
```
6 | * *
6 | * *
7 | * * *
7 | * * * *
8 | * * * *
8 | * * * * * *
9 | * * * * * * *
9 | * * * *
```

(a) Describe the shape of this distribution.

The distribution is spread from 64-95.
It is slightly skewed to the left.
The median is between 85 and 89.
The IQR goes from around 75-79 up to 90-94

(b) In order to motivate her students, the instructor of the class wants to report that, overall, the class's performance on the exam was high. Which summary statistic, the mean or the median, should the instructor use to report that overall exam performance was high? Explain.

The median score; because the distribution is skewed to the left, the mean is pulled toward the long tail, and thus the median will be higher than the mean



GO ON TO THE NEXT PAGE.

1A

(c) The midrange is defined as $\frac{\text{maximum} + \text{minimum}}{2}$. Compute this value using the data on the preceding page.

$$\frac{\overset{\text{min}}{64} + \overset{\text{max}}{95}}{2} = 79.5$$

Is the midrange considered a measure of center or a measure of spread? Explain.

A measure of center, because you cannot tell anything about how spread out the data is by this one number; it only gives the value halfway between the highest and lowest values.

GO ON TO THE NEXT PAGE.

2. For an upcoming concert, each customer may purchase up to 3 child tickets and 3 adult tickets. Let C be the number of child tickets purchased by a single customer. The probability distribution of the number of child tickets purchased by a single customer is given in the table below.

c	0	1	2	3
$p(c)$	0.4	0.3	0.2	0.1

(a) Compute the mean and the standard deviation of C.

mean = $\Sigma(p \cdot c) = (0\times0.4) + (1\times0.3) + (2\times0.2) + (3\times0.1)$

$= 0 + 0.3 + 0.4 + 0.3 = 1.$

variance = $(0-1)^2\times0.4 + (1-1)^2\times0.3 + (2-1)^2\times0.2 + (3-1)^2\times0.1$

$= 0.4 + 0 + 0.2 + 0.4 = 1.$

standard deviation = $\sqrt{\text{variance}} = \sqrt{1} = 1.$

E(x) $\therefore$ mean = 1 ticket

σ standard deviation = 1

(b) Suppose the mean and the standard deviation for the number of adult tickets purchased by a single customer are 2 and 1.2, respectively. Assume that the numbers of child tickets and adult tickets purchased are independent random variables. Compute the mean and the standard deviation of the total number of adult and child tickets purchased by a single customer.

Total mean = child mean + adult mean

$= 1 + 2 = 3$

Total variance = child variance + adult variance

$= 1 + (1.2)^2 = 2.44$

Standard deviation = $\sqrt{2.44} \approx 1.562$

adult variance || (adult standard deviation)^2

$\therefore$ Total ticket by single customer

mean = 3 tickets

standard deviation = 1.562

GO ON TO THE NEXT PAGE.

(c) Suppose each child ticket costs \$15 and each adult ticket costs \$25. Compute the mean and the standard deviation of the total amount spent per purchase.

Number of child tickets

mean = 1 standard deviation = 1

Each ticket costs \$15, so multiplying each value by 15 would cause mean and standard deviation to be multiplied by 15

↳ Price of child tickets [Mean = \$15
Standard deviation = σ = \$15

Number of adult tickets

mean = 2 standard deviation = 1.2

Each ticket costs \$25, so following similar reasoning as above, multiply mean & σ by 25.

↳ price of adult tickets [mean = \$50
σ = \$30

Total Mean price = mean child price + mean adult price

= \$15 + \$50 = \$65

Total standard deviation = $\sqrt{\sigma_{\text{child price}}^2 + \sigma_{\text{adult price}}^2}$

$= \sqrt{15^2 + 30^2} = 33.541$

∴ Total price

Mean = \$65
Standard deviation = \$33.541

GO ON TO THE NEXT PAGE.

2B

2. For an upcoming concert, each customer may purchase up to 3 child tickets and 3 adult tickets. Let C be the number of child tickets purchased by a single customer. The probability distribution of the number of child tickets purchased by a single customer is given in the table below.

c	0	1	2	3
$p(c)$	0.4	0.3	0.2	0.1

(a) Compute the mean and the standard deviation of C.

C = # of child tickets purchased by a single customer.

$\mu_C = \sum x_i p_i$

$= (0 \cdot .4) + (1 \cdot .3) + (2 \cdot .2) + (3 \cdot .1)$

$\mu_C = 1$

$\sigma^2_C = \sum (x_i - \mu_x)^2 p_i$

$= (0-1)^2 .4 + (1-1)^2 .3 + (2-1)^2 .2 + (3-1)^2 .1$

$= .4 + 0 + .2 + .4$

$\sigma^2_C = 1$

$\sigma = \sqrt{1} = 1.$

(b) Suppose the mean and the standard deviation for the number of adult tickets purchased by a single customer are 2 and 1.2, respectively. Assume that the numbers of child tickets and adult tickets purchased are independent random variables. Compute the mean and the standard deviation of the total number of adult and child tickets purchased by a single customer.

$\mu_{C+A} = \mu_C + \mu_A$

$\mu_{C+A} = 1 + 2 = 3$

$\sigma^2_{C+A} = \sigma^2_C + \sigma^2_A$

$\sigma^2_{C+A} = 1 + 1.44 = 2.44.$

$\sigma_{C+A} = \sqrt{2.44} = 1.56$

GO ON TO THE NEXT PAGE.

(c) Suppose each child ticket costs \$15 and each adult ticket costs \$25. Compute the mean and the standard deviation of the total amount spent per purchase.

Mean # child tickets = 1
Mean # of adult tickets = 2

$15 + 2(25) = 65$

mean amount spent per perchase = 65 \$

σ^2 of child tickets = 1
σ^2 of adult tickets = 1.44

$15 + 1.44(25) = 51$

$\sqrt{51} = 7.14$

standard deviation of total amount spent per purchase = 7.14 \$

GO ON TO THE NEXT PAGE.

2C

2. For an upcoming concert, each customer may purchase up to 3 child tickets and 3 adult tickets. Let C be the number of child tickets purchased by a single customer. The probability distribution of the number of child tickets purchased by a single customer is given in the table below.

c	0	1	2	3
$p(c)$	0.4	0.3	0.2	0.1

(a) Compute the mean and the standard deviation of C.

mean: $(0 \cdot .4) + (1 \cdot .3) + (2 \cdot .2) + (3 \cdot .1) = 1$

Standard deviation:

$$\sqrt{\left(\frac{(.3)(.7)}{1} + \frac{(.2)(.8)}{2} + \frac{(.1)(.9)}{3}\right.} = .565685\ldots \approx .566$$

(b) Suppose the mean and the standard deviation for the number of adult tickets purchased by a single customer are 2 and 1.2, respectively. Assume that the numbers of child tickets and adult tickets purchased are independent random variables. Compute the mean and the standard deviation of the total number of adult and child tickets purchased by a single customer.

mean: $\frac{2+1}{2} = 1.5$

Stan. dev: $\sqrt{\frac{1.2^2}{2} + \frac{.566^2}{1}} = 1.019978\ldots \approx 1.02$

GO ON TO THE NEXT PAGE.

2C

(c) Suppose each child ticket costs \$15 and each adult ticket costs \$25. Compute the mean and the standard deviation of the total amount spent per purchase.

$(1 \cdot 15) + (2 \cdot 25) = \$65$

$(.566 \cdot 15) + (1.2 \cdot 25) = \$38.49$

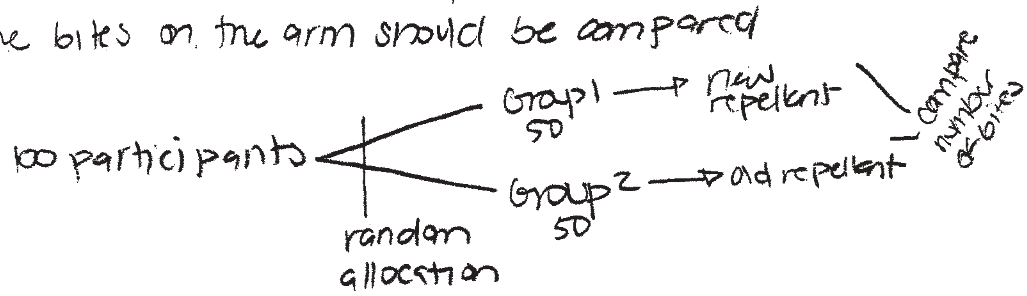
GO ON TO THE NEXT PAGE.

3. In search of a mosquito repellent that is safer than the ones that are currently on the market, scientists have developed a new compound that is rated as less toxic than the current compound, thus making a repellent that contains this new compound safer for human use. Scientists also believe that a repellent containing the new compound will be more effective than the ones that contain the current compound. To test the effectiveness of the new compound versus that of the current compound, scientists have randomly selected 100 people from a state.

Up to 100 bins, with an equal number of mosquitoes in each bin, are available for use in the study. After a compound is applied to a participant's forearm, the participant will insert his or her forearm into a bin for 1 minute, and the number of mosquito bites on the arm at the end of that time will be determined.

(a) Suppose this study is to be conducted using a completely randomized design. Describe a randomization process and identify an inference procedure for the study.

First, you should obtain an alphabetical list of the names of the randomly selected participants. Then number them from 001 to 100 corresponding to the random digit table. Using the table or a calculator, randomly select the first 50 numbers. These participants will be in group 1 and recieve the new repellent. The remaining 50 partiupants will be in group 2 and relieve the old repellent. Each participant will place their arm in a seperate bin and after 1 minute the bites on the arm should be compared for both groups.



• The best inference procedure for this study would be a two sample t test, because you have two independent samples and can compare the mean number of bites for group 1 and group 2.

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 < \mu_2$

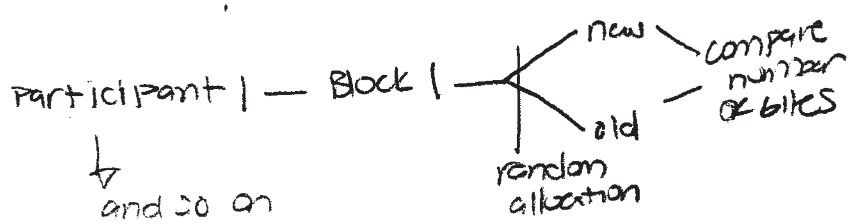
μ_1 = population mean number of bites for new repellent.

μ_2 = population mean number of bites for old repellent.

GO ON TO THE NEXT PAGE.

(b) Suppose this study is to be conducted using a matched-pairs design. Describe a randomization process and identify an inference procedure for the study.

For a matched pairs design, each subject could test both repellents. To decide which repellent they will test first, you can toss a coin for each participant. If it comes up heads they can use the new repellent first and the old one second and vice versa for tails. Then have each subject test both repellents in the same bin of misquitoes and compare the number of bites for each participant.



• The best test for this study would be a matched pairs t test. You can subtract the amount of bites with the old repellent the number with the new repellent and calculate the mean difference.

μ: population mean amount of

$H_0: \mu = 0$

$H_0: \mu < 0$

(c) Which of the designs, the one in part (a) or the one in part (b), is better for testing the effectiveness of the new compound versus that of the current compound? Justify your answer.

The design in part (b) would probably be the best because having each subject try each repellent would prevent the confounding of variables, like whether misquitoes just tend to bite that person more and whether or not the new repellent is more effective, or whether the misquitoes in that particular bin reacted differently.

GO ON TO THE NEXT PAGE.

3. In search of a mosquito repellent that is safer than the ones that are currently on the market, scientists have developed a new compound that is rated as less toxic than the current compound, thus making a repellent that contains this new compound safer for human use. Scientists also believe that a repellent containing the new compound will be more effective than the ones that contain the current compound. To test the effectiveness of the new compound versus that of the current compound, scientists have randomly selected 100 people from a state.

Up to 100 bins, with an equal number of mosquitoes in each bin, are available for use in the study. After a compound is applied to a participant's forearm, the participant will insert his or her forearm into a bin for 1 minute, and the number of mosquito bites on the arm at the end of that time will be determined.

(a) Suppose this study is to be conducted using a completely randomized design. Describe a randomization process and identify an inference procedure for the study.

First, number everyone from 00 to 99. Choose a row number from the random table by randomization. Go through the numbers by 2 digits. The first 50 numbers from that row between 00 and 99 will be chosen. People with the corresponding numbers will be selected to test the new compound. If the same number appears before the 50th number, disregard that number and choose the next one. After getting the 50 people, other 50 people will test the current compound.

After counting the number of mosquito bites, find if the $\text{mean}_{(1)}$ of mosquito bites from the people who were tested with the new compound is less than the $\text{mean}_{(2)}$ of those who were tested with the current compound. Using the 2 sample T Test, if the $\text{mean}_{(1)}$ is significantly lower than the $\text{mean}_{(2)}$, then the new compound is effective. If not, there is no significant evidence that the new compond is better.

(b) Suppose this study is to be conducted using a matched-pairs design. Describe a randomization process and identify an inference procedure for the study.

Repeat the procedure from (a) until we pick the first 50 people. These people will test the new compound with their right arm and the current compound with the left arm. The other 50 people will test the new compound with their left arm and the current compound with the right arm.

Each person will find the difference of mosquito bites on their arms. (Mosquito bites using the new compound − mosquito bites using the current compound). Find the mean and standard deviation of these differences. Then test this mean against $H_0: \mu_{(mean)} = 0$. If the mean of the differences are significantly smaller than 0, the new compound is effective. If not, there is no evidence that the two compounds are different

(c) Which of the designs, the one in part (a) or the one in part (b), is better for testing the effectiveness of the new compound versus that of the current compound? Justify your answer.

A Matched-Pair Design (part B) is better since the two compounds are tested on the same person, who has the same traits. If we just test separately (part A), the two group may have certain traits that will favor one compound over the other, or one group or 50 people may have many people who don't get mosquito bites easily. Therefore, to find if there is a difference, procedure in part (B) is better.

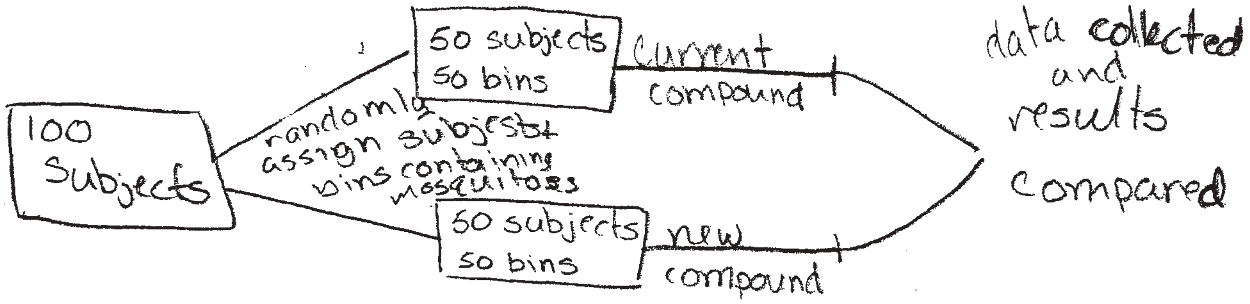
GO ON TO THE NEXT PAGE.

3C

3. In search of a mosquito repellent that is safer than the ones that are currently on the market, scientists have developed a new compound that is rated as less toxic than the current compound, thus making a repellent that contains this new compound safer for human use. Scientists also believe that a repellent containing the new compound will be more effective than the ones that contain the current compound. To test the effectiveness of the new compound versus that of the current compound, scientists have randomly selected 100 people from a state.

Up to 100 bins, with an equal number of mosquitoes in each bin, are available for use in the study. After a compound is applied to a participant's forearm, the participant will insert his or her forearm into a bin for 1 minute, and the number of mosquito bites on the arm at the end of that time will be determined.

(a) Suppose this study is to be conducted using a completely randomized design. Describe a randomization process and identify an inference procedure for the study.



The scientists should randomly assign the 100 subjects to one of two groups. Group A will have 50 subjects and group B will have 50 subjects. The scientists should then randomly assign 50 of the bins containing mosquitos to group A and 50 of the bins to group B. The subjects in group A would apply the current compound to their arm and will place it in the randomly assigned bin for 1 minute. The subjects in group B will do the same only will use the new compound. Once the data has been collected, the results will be compared

GO ON TO THE NEXT PAGE.

3C

(b) Suppose this study is to be conducted using a matched-pairs design. Describe a randomization process and identify an inference procedure for the study.

Each subject will be randomly assigned a bin containing mosquitoes. On their right arm the subjects will put the current compound while on the other they will put the new compound. They will insert each arm into the bin for 1 minute. At the end of 1 minute the number of mosquito bites will be counted and the results will be compared.

(c) Which of the designs, the one in part (a) or the one in part (b), is better for testing the effectiveness of the new compound versus that of the current compound? Justify your answer.

The design in part(b) is better for testing the effectiveness of the new compound versus that of the current compound because 1 bin of mosquitoes may have a tendency to bite more than another. Also, Some people may be more suseptable to mosquito bites than others.

GO ON TO THE NEXT PAGE.

4A

4. A researcher believes that treating seeds with certain additives before planting can enhance the growth of plants. An experiment to investigate this is conducted in a greenhouse. From a large number of Roma tomato seeds, 24 seeds are randomly chosen and 2 are assigned to each of 12 containers. One of the 2 seeds is randomly selected and treated with the additive. The other seed serves as a control. Both seeds are then planted in the same container. The growth, in centimeters, of each of the 24 plants is measured after 30 days. These data were used to generate the partial computer output shown below. Graphical displays indicate that the assumption of normality is not unreasonable.

	N	Mean	StDev	SE Mean
Control	12	15.989	1.098	0.317
Treatment	12	18.004	1.175	0.339
Difference	12	− 2.015	1.163	0.336

(a) Construct a confidence interval for the mean difference in growth, in centimeters, of the plants from the untreated and treated seeds. Be sure to interpret this interval.

$CI = \bar{x} \pm t \cdot \frac{s}{\sqrt{n}}$ t at C = 95% and d.f. = $\cancel{24}$ 12 − 1 = 11

t = 2.201

$= -2.015 \pm 2.201 \frac{(1.163)}{\sqrt{12}}$

$= (-2.754, -1.276)$ in cm.

Meaning:
We are 95% confident that the interval contains the true mean difference in growth

GO ON TO THE NEXT PAGE.

(b) Based only on the confidence interval in part (a), is there sufficient evidence to conclude that there is a significant mean difference in growth of the plants from untreated seeds and the plants from treated seeds? Justify your conclusion.

Yes, based only on the confidence interval in part (a), there is sufficient evidence to conclude that there is a significant mean difference in growth of the plants from untreated seeds and the plants from treated seeds.

If we were to presume there is not a significant mean difference,

$H_0: \mu_d = 0$, when μ_d is the mean difference.

$H_a: \mu_d \neq 0$

However $\mu_d = 0$ is not even contained in the 95% interval obtained in (a). Hence, we can, at 95% ~~&~~ confidence, reject the H_0 and conclude that there is a significant mean growth.

GO ON TO THE NEXT PAGE.

4. A researcher believes that treating seeds with certain additives before planting can enhance the growth of plants. An experiment to investigate this is conducted in a greenhouse. From a large number of Roma tomato seeds, 24 seeds are randomly chosen and 2 are assigned to each of 12 containers. One of the 2 seeds is randomly selected and treated with the additive. The other seed serves as a control. Both seeds are then planted in the same container. The growth, in centimeters, of each of the 24 plants is measured after 30 days. These data were used to generate the partial computer output shown below. Graphical displays indicate that the assumption of normality is not unreasonable.

	N	Mean	StDev	SE Mean
Control	12	15.989	1.098	0.317
Treatment	12	18.004	1.175	0.339
Difference	12	− 2.015	1.163	0.336

(a) Construct a confidence interval for the mean difference in growth, in centimeters, of the plants from the untreated and treated seeds. Be sure to interpret this interval.

95% CI → t.

① Requirements →

1. Assume normal population of growth
2. Sample is randomly chosen
3. s is known.

CI = statistic ± critical value (standard deviation)

$= -2.015 \pm 2.069 \left(\frac{1.163}{\sqrt{24}} \right)$

df = 23.

$= 2.015 \pm 2.069 (0.24)$

$= 2.015 \pm 0.491$

$= \boxed{2.506 \text{ to } 1.523}$

I am 95% confident that the mean population difference in growth, in cm, of the plants from the untreated and treated seed is between 2.506 and 1.523, with the control being the one that grows less.

GO ON TO THE NEXT PAGE.

4B

(b) Based only on the confidence interval in part (a), is there sufficient evidence to conclude that there is a significant mean difference in growth of the plants from untreated seeds and the plants from treated seeds? Justify your conclusion.

$H_0: \mu_d = 0$, where μ_d is the mean diff in growth of plants from untreated and treated seeds.

$H_a: \mu_d \neq 0$

Because 0 is not contained in the 95% CI interval for mean different [2.506, 1.523] it is reasonable to reject the H_0. So, there is sufficient evidence to conclude that there is a significant mean difference, since the difference is not 0.

GO ON TO THE NEXT PAGE.

4C

4. A researcher believes that treating seeds with certain additives before planting can enhance the growth of plants. An experiment to investigate this is conducted in a greenhouse. From a large number of Roma tomato seeds, 24 seeds are randomly chosen and 2 are assigned to each of 12 containers. One of the 2 seeds is randomly selected and treated with the additive. The other seed serves as a control. Both seeds are then planted in the same container. The growth, in centimeters, of each of the 24 plants is measured after 30 days. These data were used to generate the partial computer output shown below. Graphical displays indicate that the assumption of normality is not unreasonable.

	N	Mean	StDev	SE Mean
Control	12	15.989	1.098	0.317
Treatment	12	18.004	1.175	0.339
Difference	12	−2.015	1.163	0.336

(a) Construct a confidence interval for the mean difference in growth, in centimeters, of the plants from the untreated and treated seeds. Be sure to interpret this interval.

conditions: random (stated in problem)
σ unknown (t - test)
~~indep~~ normal (stated, implied)
~~matched~~ paired (implied)

$\bar{x}_d \pm t^* \frac{s_d}{\sqrt{n}}$ 95% → $t^* = 1.96$

~~$\pm$ 1.96~~

$-2.015 \pm 1.96\left(\frac{1.163}{\sqrt{12}}\right)$

$(-2.67, -1.36)$

There's 95% confidence that true mean difference is between the values of −2.67 and −1.36.

GO ON TO THE NEXT PAGE.

4C

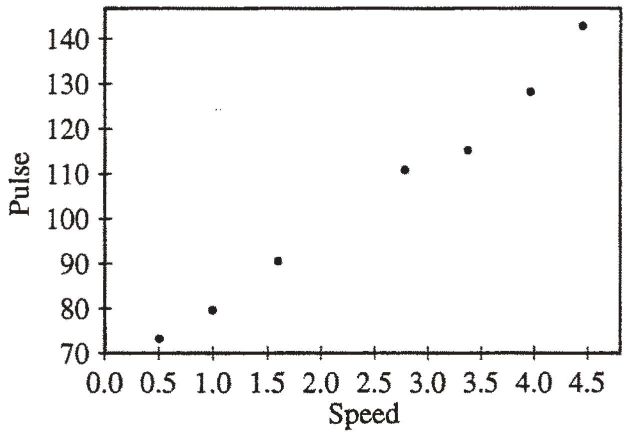
(b) Based only on the confidence interval in part (a), is there sufficient evidence to conclude that there is a significant mean difference in growth of the plants from untreated seeds and the plants from treated seeds? Justify your conclusion.

Yes because 0 is not included in the interval. Hence, there must be a significant mean difference in the growth

GO ON TO THE NEXT PAGE.

5A

5. John believes that as he increases his walking speed, his pulse rate will increase. He wants to model this relationship. John records his pulse rate, in beats per minute (bpm), while walking at each of seven different speeds, in miles per hour (mph). A scatterplot and regression output are shown below.



Regression Analysis: Pulse Versus Speed

Predictor	Coef	SE Coef	T	P
Constant	63.457	2.387	26.58	0.000
Speed	16.2809	0.8192	19.88	0.000

S = 3.087 R-Sq = 98.7% R-Sq (adj) = 98.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	3763.2	3763.2	396.13	0.000
Residual	5	47.6	9.5		
Total	6	3810.9			

(a) Using the regression output, write the equation of the fitted regression line.

$\hat{y} = 63.457 + 16.2809x$

$\hat{y}$ = estimated pulse rate (bpm)

x = walking speed (mph)

GO ON TO THE NEXT PAGE.

(b) Do your estimates of the slope and intercept parameters have meaningful interpretations in the context of this question? If so, provide interpretations in this context. If not, explain why not.

Yes, they do have meaningful interpretations in the context. The intercept parameter is 63.457 bpm, which is when walking speed = 0. This shows that the resting pulse rate would be expected to be close to 63.457 bpm. The slope parameter is 16.2809, which means that for every increase in the walking speed by the increment of 1 mph, the pulse rate is expected to rise by 16.2809 bpm.

(c) John wants to provide a 98 percent confidence interval for the slope parameter in his final report. Compute the margin of error that John should use. Assume that conditions for inference are satisfied.

margin of error = t · SE

SE = .8192 is given

At 98% confidence, with degrees of freedom = 7 - 2 = 5,

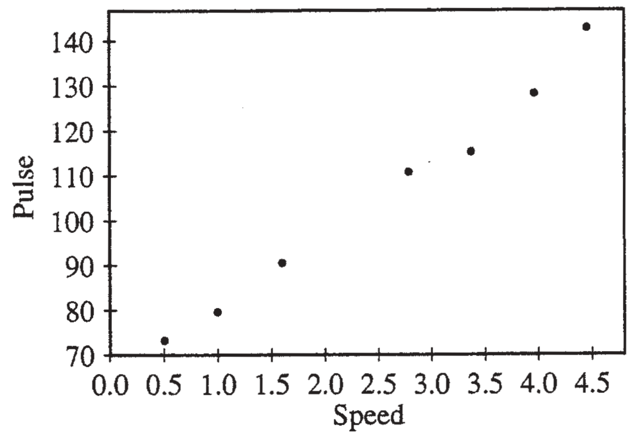
t = 3.365

margin of error = (3.365)(.8192)

= ±2.757

GO ON TO THE NEXT PAGE.

5. John believes that as he increases his walking speed, his pulse rate will increase. He wants to model this relationship. John records his pulse rate, in beats per minute (bpm), while walking at each of seven different speeds, in miles per hour (mph). A scatterplot and regression output are shown below.



Regression Analysis: Pulse Versus Speed

Predictor	Coef	SE Coef	T	P
Constant	63.457	2.387	26.58	0.000
Speed	16.2809	0.8192	19.88	0.000

S = 3.087 R-Sq = 98.7% R-Sq (adj) = 98.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	3763.2	3763.2	396.13	0.000
Residual	5	47.6	9.5		
Total	6	3810.9			

(a) Using the regression output, write the equation of the fitted regression line.

$\hat{y} = 63.457 + 16.2809x$

GO ON TO THE NEXT PAGE.

(b) Do your estimates of the slope and intercept parameters have meaningful interpretations in the context of this question? If so, provide interpretations in this context. If not, explain why not.

The estimates of the slope and intercept parameters are meaningful.

The slope of 16.2809 means that as John's walking speed increases by 1 mile/hour, his heart rate increases on average by 16.2809 beats per minute.

The intercept occurs when John's walking speed is zero. Thus, the intercept of 63.457 means that John's resting heart rate is around 63.457

(c) John wants to provide a 98 percent confidence interval for the slope parameter in his final report. Compute the margin of error that John should use. Assume that conditions for inference are satisfied.

confidence interval :

$b_1 \pm \boxed{t^* SE_{b_1}} \rightarrow$ margin of error.

$t^* = 3.365$ $\quad df = n-2 = 5$

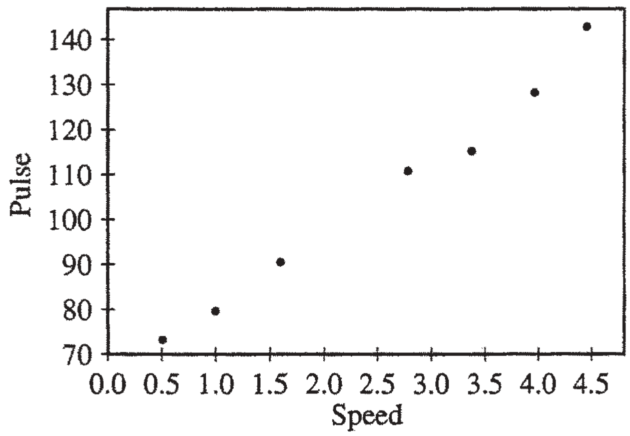
$SE_{b_1} = 0.8192$

$t^* SE_{b_1} = (3.365)(0.8192) = \boxed{2.7566}$

GO ON TO THE NEXT PAGE.

5C

5. John believes that as he increases his walking speed, his pulse rate will increase. He wants to model this relationship. John records his pulse rate, in beats per minute (bpm), while walking at each of seven different speeds, in miles per hour (mph). A scatterplot and regression output are shown below.



Regression Analysis: Pulse Versus Speed

Predictor	Coef	SE Coef	T	P
Constant	63.457	2.387	26.58	0.000
Speed	16.2809	0.8192	19.88	0.000

S = 3.087 R-Sq = 98.7% R-Sq (adj) = 98.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	3763.2	3763.2	396.13	0.000
Residual	5	47.6	9.5		
Total	6	3810.9			

(a) Using the regression output, write the equation of the fitted regression line.

Predicted Pulse = 63.457 + 16.2809 (Speed)

GO ON TO THE NEXT PAGE.

5C

(b) Do your estimates of the slope and intercept parameters have meaningful interpretations in the context of this question? If so, provide interpretations in this context. If not, explain why not.

Yes. The slope parameter tells you how much the pulse rate increases as speed increases. The intercept parameter tells you what the pulse rate is without the speed or while standing still.

(c) John wants to provide a 98 percent confidence interval for the slope parameter in his final report. Compute the margin of error that John should use. Assume that conditions for inference are satisfied.

margin of error $= z^*(SE_b) = 2.326(.8192) = 1.905$

GO ON TO THE NEXT PAGE.

6A

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. Regulations require that product labels on containers of food that are available for sale to the public accurately state the amount of food in those containers. Specifically, if milk containers are labeled to have 128 fluid ounces and the mean number of fluid ounces of milk in the containers is at least 128, the milk processor is considered to be in compliance with the regulations. The filling machines can be set to the labeled amount. Variability in the filling process causes the actual contents of milk containers to be normally distributed. A random sample of 12 containers of milk was drawn from the milk processing line in a plant, and the amount of milk in each container was recorded.

(a) The sample mean and standard deviation of this sample of 12 containers of milk were 127.2 ounces and 2.1 ounces, respectively. Is there sufficient evidence to conclude that the packaging plant is not in compliance with the regulations? Provide statistical justification for your answer.

$n=12$ $\quad$ $H_0: \mu \geq 128$
$\bar{x}=127.2$ $\quad$ $H_a: \mu < 128$
$s=2.1$

Since population standard deviation is unknown, we use the t-test for single sample mean
Assumptions are met, since the question states that contents are normally distributed and that the sample was a simple random sample.

$$t = \frac{\bar{x}-\mu}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{127.2-128}{\left(\frac{2.1}{\sqrt{12}}\right)} = -1.31966$$

Using the graphing calculator, we obtain the p-value of 0.1069.
Since p-value is sufficiently large, we do not have evidence against H_0 and conclude that there is not sufficient evidence that the packaging plant is not in compliance with regulations.

GO ON TO THE NEXT PAGE.

6A

Inspectors decide to study a particular filling machine within this plant further. For this machine, the amount of milk in the containers has a mean of 128.0 fluid ounces and a standard deviation of 2.0 fluid ounces.

(b) What is the probability that a randomly selected container filled by this machine contains at least 125 fluid ounces?

$$z = \frac{125 - 128}{2} = -1.5$$

$$P(z \geq -1.5) = \boxed{.933}$$

(c) An inspector will randomly select 12 containers filled by this machine and record the amount of milk in each. What is the probability that the minimum (smallest amount of milk) recorded in the 12 containers will be at least 125 fluid ounces? (Note: In order for the minimum to be at least 125 fluid ounces, each of the 12 containers must contain at least 125 fluid ounces.)

Let X be a random variable following binomial distribution with $p = .933$, $n = 12$

$$P(X = 12) = \binom{12}{12}(.933)^{12}(1-.933)^{0}$$

$$= (.933)^{12} = \boxed{.435}$$

An analyst wants to use simulation to investigate the sampling distribution of the minimum. This analyst randomly generates 150 samples, each consisting of 12 observations, from a normal distribution with mean 128 and standard deviation 2 and finds the minimum for each sample. The 150 minimums (sorted from smallest to largest) are shown on the next page.

GO ON TO THE NEXT PAGE.

6A

Sample	Minimum	Sample	Minimum	Sample	Minimum
1	121.45	51	124.28	101	125.25
2	122.51	52	124.29	102	125.31
3	122.53	53	124.30	103	125.36
4	122.72	54	124.31	104	125.38
5	122.75	55	124.34	105	125.40
6	122.89	56	124.36	106	125.42
7	122.93	57	124.37	107	125.48
8	122.99	58	124.37	108	125.49
9	123.04	59	124.39	109	125.50
10	123.08	60	124.39	110	125.52
11	123.09	61	124.41	111	125.54
12	123.10	62	124.44	112	125.56
13	123.31	63	124.53	113	125.61
14	123.34	64	124.53	114	125.67
15	123.39	65	124.54	115	125.72
16	123.40	66	124.55	116	125.76
17	123.41	67	124.55	117	125.77
18	123.41	68	124.55	118	125.78
19	123.46	69	124.55	119	125.79
20	123.49	70	124.58	120	125.84
21	123.51	71	124.67	121	125.87
22	123.57	72	124.69	122	125.87
23	123.58	73	124.73	123	125.90
24	123.59	74	124.77	124	125.90
25	123.60	75	124.78	125	125.93
26	123.66	76	124.78	126	125.93
27	123.67	77	124.80	127	125.93
28	123.72	78	124.80	128	125.94
29	123.75	79	124.81	129	125.98
30	123.77	80	124.85	130	126.00
31	123.78	81	124.91	131	126.03
32	123.84	82	124.92	132	126.05
33	123.91	83	124.92	133	126.05
34	123.93	84	124.96	134	126.06
35	123.95	85	125.00	135	126.09
36	123.95	86	125.01	136	126.15
37	123.98	87	125.02	137	126.15
38	123.99	88	125.02	138	126.16
39	124.05	89	125.03	139	126.19
40	124.05	90	125.04	140	126.19
41	124.06	91	125.05	141	126.25
42	124.12	92	125.07	142	126.26
43	124.14	93	125.08	143	126.33
44	124.15	94	125.09	144	126.35
45	124.16	95	125.14	145	126.45
46	124.19	96	125.18	146	126.50
47	124.23	97	125.21	147	126.57
48	124.27	98	125.21	148	126.62
49	124.28	99	125.22	149	126.64
50	124.28	100	125.25	150	126.95

GO ON TO THE NEXT PAGE.

(d) Use the simulation results to estimate the probability that was requested in part (c) and compare this estimate with the theoretical value you calculated.

After sample 84, minimum ≥ 125.
Hence number of samples with minimum value of at least 125 is $n-84 = 150-84 = 66$.

P(minimum is 12 containers at least 125) = $\dfrac{\text{frequency of samples with min} \geq 125}{\text{Total number of samples}}$

$= \dfrac{66}{150} =$ ~~...~~ $.44$

This probability is close to the theoretical value I calculated in part (b) of .435.
This is reasonable, since the sample size (150) is fairly large. By the law of large numbers, as the sample size increases, the observed ~~probability~~ proportion approaches the expected probability (theoretical estimate).

END OF EXAMINATION

THE FOLLOWING INSTRUCTIONS APPLY TO THE BACK COVER OF THE SECTION II BOOKLET.

- **MAKE SURE YOU HAVE COMPLETED THE IDENTIFICATION INFORMATION AS REQUESTED ON THE BACK OF THE SECTION II BOOKLET.**
- **CHECK TO SEE THAT YOUR AP NUMBER APPEARS IN THE BOX(ES) ON THE BACK COVER.**
- **MAKE SURE YOU HAVE USED THE SAME SET OF AP NUMBER LABELS ON ALL AP EXAMINATIONS YOU HAVE TAKEN THIS YEAR.**

GO ON TO THE NEXT PAGE.

6B

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. Regulations require that product labels on containers of food that are available for sale to the public accurately state the amount of food in those containers. Specifically, if milk containers are labeled to have 128 fluid ounces and the mean number of fluid ounces of milk in the containers is at least 128, the milk processor is considered to be in compliance with the regulations. The filling machines can be set to the labeled amount. Variability in the filling process causes the actual contents of milk containers to be normally distributed. A random sample of 12 containers of milk was drawn from the milk processing line in a plant, and the amount of milk in each container was recorded.

   (a) The sample mean and standard deviation of this sample of 12 containers of milk were 127.2 ounces and 2.1 ounces, respectively. Is there sufficient evidence to conclude that the packaging plant is <u>not</u> in compliance with the regulations? Provide statistical justification for your answer.

one sample t-test   Data is SRS ✓   Containers are independant ✓

$10(12) < \text{population}$ ✓

$H_0: \mu = 128$ $H_a: \mu \neq 128$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{127.2 - 128}{\frac{2.1}{\sqrt{12}}} = \frac{-.8}{.6062} = \underline{-1.3197}$$

$t = -1.3197$ $p = .0935$ $df = 11$

We must keep Ho. There is not sufficient evidence to conclude that the plant is not within regulations. This is due to the p-value of .0935 not being significant at the a=.05 level.

GO ON TO THE NEXT PAGE.

6B

Inspectors decide to study a particular filling machine within this plant further. For this machine, the amount of milk in the containers has a mean of 128.0 fluid ounces and a standard deviation of 2.0 fluid ounces.

(b) What is the probability that a randomly selected container filled by this machine contains at least 125 fluid ounces?

$P(x > 125) = P\left(\frac{x-\mu}{\sigma} > \frac{125-128}{2}\right) = P(z > -1.5) = .9332$

$\underline{P(x > 125) = .9332}$

(c) An inspector will randomly select 12 containers filled by this machine and record the amount of milk in each. What is the probability that the minimum (smallest amount of milk) recorded in the 12 containers will be at least 125 fluid ounces? (Note: In order for the minimum to be at least 125 fluid ounces, each of the 12 containers must contain at least 125 fluid ounces.)

$P(\text{All 12 over 125}) = P(x > 125)^{12} = (.9332)^{12} = \boxed{.43621}$

An analyst wants to use simulation to investigate the sampling distribution of the minimum. This analyst randomly generates 150 samples, each consisting of 12 observations, from a normal distribution with mean 128 and standard deviation 2 and finds the minimum for each sample. The 150 minimums (sorted from smallest to largest) are shown on the next page.

GO ON TO THE NEXT PAGE.

6B

Sample	Minimum
1	121.45
2	122.51
3	122.53
4	122.72
5	122.75
6	122.89
7	122.93
8	122.99
9	123.04
10	123.08
11	123.09
12	123.10
13	123.31
14	123.34
15	123.39
16	123.40
17	123.41
18	123.41
19	123.46
20	123.49
21	123.51
22	123.57
23	123.58
24	123.59
25	123.60
26	123.66
27	123.67
28	123.72
29	123.75
30	123.77
31	123.78
32	123.84
33	123.91
34	123.93
35	123.95
36	123.95
37	123.98
38	123.99
39	124.05
40	124.05
41	124.06
42	124.12
43	124.14
44	124.15
45	124.16
46	124.19
47	124.23
48	124.27
49	124.28
50	124.28
51	124.28
52	124.29
53	124.30
54	124.31
55	124.34
56	124.36
57	124.37
58	124.37
59	124.39
60	124.39
61	124.41
62	124.44
63	124.53
64	124.53
65	124.54
66	124.55
67	124.55
68	124.55
69	124.55
70	124.58
71	124.67
72	124.69
73	124.73
74	124.77
75	124.78
76	124.78
77	124.80
78	124.80
79	124.81
80	124.85
81	124.91
82	124.92
83	124.92
84	124.96
85	125.00
86	125.01
87	125.02
88	125.02
89	125.03
90	125.04
91	125.05
92	125.07
93	125.08
94	125.09
95	125.14
96	125.18
97	125.21
98	125.21
99	125.22
100	125.25
101	125.25
102	125.31
103	125.36
104	125.38
105	125.40
106	125.42
107	125.48
108	125.49
109	125.50
110	125.52
111	125.54
112	125.56
113	125.61
114	125.67
115	125.72
116	125.76
117	125.77
118	125.78
119	125.79
120	125.84
121	125.87
122	125.87
123	125.90
124	125.90
125	125.93
126	125.93
127	125.93
128	125.94
129	125.98
130	126.00
131	126.03
132	126.05
133	126.05
134	126.06
135	126.09
136	126.15
137	126.15
138	126.16
139	126.19
140	126.19
141	126.25
142	126.26
143	126.33
144	126.35
145	126.45
146	126.50
147	126.57
148	126.62
149	126.64
150	126.95

GO ON TO THE NEXT PAGE.

6B

*

(d) Use the simulation results to estimate the probability that was requested in part (c) <u>and</u> compare this estimate with the theoretical value you calculated.

$$\frac{\#\text{ over }125}{150} = \frac{66}{150} = .44$$

predicted value is .4362 while simulated value is .44
The simulated proportion is a little higher, but still well with sampling variability.

END OF EXAMINATION

THE FOLLOWING INSTRUCTIONS APPLY TO THE BACK COVER OF THE SECTION II BOOKLET.

- **MAKE SURE YOU HAVE COMPLETED THE IDENTIFICATION INFORMATION AS REQUESTED ON THE BACK OF THE SECTION II BOOKLET.**
- **CHECK TO SEE THAT YOUR AP NUMBER APPEARS IN THE BOX(ES) ON THE BACK COVER.**
- **MAKE SURE YOU HAVE USED THE SAME SET OF AP NUMBER LABELS ON <u>ALL</u> AP EXAMINATIONS YOU HAVE TAKEN THIS YEAR.**

GO ON TO THE NEXT PAGE.

6C

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. Regulations require that product labels on containers of food that are available for sale to the public accurately state the amount of food in those containers. Specifically, if milk containers are labeled to have 128 fluid ounces and the mean number of fluid ounces of milk in the containers is at least 128, the milk processor is considered to be in compliance with the regulations. The filling machines can be set to the labeled amount. Variability in the filling process causes the actual contents of milk containers to be normally distributed. A random sample of 12 containers of milk was drawn from the milk processing line in a plant, and the amount of milk in each container was recorded.

   (a) The sample mean and standard deviation of this sample of 12 containers of milk were 127.2 ounces and 2.1 ounces, respectively. Is there sufficient evidence to conclude that the packaging plant is not in compliance with the regulations? Provide statistical justification for your answer.

1 - Sample t test. ($\alpha = .05$)

- SRS
- Normally distribution
- $s = 2.1$

$n = 12$.

$df = n-1$
$= 11$

$H_0 = \mu_{\bar{x}} = 128$.

$H_a : \mu_{\bar{x}} < 128$.

$$t = \frac{127.2 - 128}{\frac{2.1}{\sqrt{12}}} = -.54$$

P-Value = tcdf(-1000, -.54, 11) = .30

Since the P-Value > .05, we fail to reject Ho.

GO ON TO THE NEXT PAGE.

6C

Inspectors decide to study a particular filling machine within this plant further. For this machine, the amount of milk in the containers has a mean of 128.0 fluid ounces and a standard deviation of 2.0 fluid ounces.

(b) What is the probability that a randomly selected container filled by this machine contains at least 125 fluid ounces?

$P(x \geq 125)$

$z = \frac{125-128}{2} = -1.5$

P-Value = normalcdf(-1000, -1.5) = .0668

(c) An inspector will randomly select 12 containers filled by this machine and record the amount of milk in each. What is the probability that the minimum (smallest amount of milk) recorded in the 12 containers will be at least 125 fluid ounces? (Note: In order for the minimum to be at least 125 fluid ounces, each of the 12 containers must contain at least 125 fluid ounces.)

$P(\bar{x} \geq 125)$

$$z = \frac{125-128}{\sqrt{\frac{.067(1-.067)}{12}}} = \frac{-3}{.07} = -42.86$$

P-Value = normcdf(-1000, -42.86) = 0

Since P-Value is 0, the pobability that the minimum recorded in the 12 containers will not be at least 125 fluid ounces is 0.

An analyst wants to use simulation to investigate the sampling distribution of the minimum. This analyst randomly generates 150 samples, each consisting of 12 observations, from a normal distribution with mean 128 and standard deviation 2 and finds the minimum for each sample. The 150 minimums (sorted from smallest to largest) are shown on the next page.

GO ON TO THE NEXT PAGE.

6C

Sample	Minimum	Sample	Minimum	Sample	Minimum
1	121.45	51	124.28	101	125.25
2	122.51	52	124.29	102	125.31
3	122.53	53	124.30	103	125.36
4	122.72	54	124.31	104	125.38
5	122.75	55	124.34	105	125.40
6	122.89	56	124.36	106	125.42
7	122.93	57	124.37	107	125.48
8	122.99	58	124.37	108	125.49
9	123.04	59	124.39	109	125.50
10	123.08	60	124.39	110	125.52
11	123.09	61	124.41	111	125.54
12	123.10	62	124.44	112	125.56
13	123.31	63	124.53	113	125.61
14	123.34	64	124.53	114	125.67
15	123.39	65	124.54	115	125.72
16	123.40	66	124.55	116	125.76
17	123.41	67	124.55	117	125.77
18	123.41	68	124.55	118	125.78
19	123.46	69	124.55	119	125.79
20	123.49	70	124.58	120	125.84
21	123.51	71	124.67	121	125.87
22	123.57	72	124.69	122	125.87
23	123.58	73	124.73	123	125.90
24	123.59	74	124.77	124	125.90
25	123.60	75	124.78	125	125.93
26	123.66	76	124.78	126	125.93
27	123.67	77	124.80	127	125.93
28	123.72	78	124.80	128	125.94
29	123.75	79	124.81	129	125.98
30	123.77	80	124.85	130	126.00
31	123.78	81	124.91	131	126.03
32	123.84	82	124.92	132	126.05
33	123.91	83	124.92	133	126.05
34	123.93	84	124.96	134	126.06
35	123.95	85	125.00	135	126.09
36	123.95	86	125.01	136	126.15
37	123.98	87	125.02	137	126.15
38	123.99	88	125.02	138	126.16
39	124.05	89	125.03	139	126.19
40	124.05	90	125.04	140	126.19
41	124.06	91	125.05	141	126.25
42	124.12	92	125.07	142	126.26
43	124.14	93	125.08	143	126.33
44	124.15	94	125.09	144	126.35
45	124.16	95	125.14	145	126.45
46	124.19	96	125.18	146	126.50
47	124.23	97	125.21	147	126.57
48	124.27	98	125.21	148	126.62
49	124.28	99	125.22	149	126.64
50	124.28	100	125.25	150	126.95

GO ON TO THE NEXT PAGE.

6C

(d) Use the simulation results to estimate the probability that was requested in part (c) and compare this estimate with the theoretical value you calculated.

$\frac{84}{140} = \hat{p} = .56$.

Prob. from simulation = .56
theoretical value = 0.

.56% difference in comparing the theoretical results and the results from the simulation.

END OF EXAMINATION

THE FOLLOWING INSTRUCTIONS APPLY TO THE BACK COVER OF THE SECTION II BOOKLET.

- **MAKE SURE YOU HAVE COMPLETED THE IDENTIFICATION INFORMATION AS REQUESTED ON THE BACK OF THE SECTION II BOOKLET.**
- **CHECK TO SEE THAT YOUR AP NUMBER APPEARS IN THE BOX(ES) ON THE BACK COVER.**
- **MAKE SURE YOU HAVE USED THE SAME SET OF AP NUMBER LABELS ON ALL AP EXAMINATIONS YOU HAVE TAKEN THIS YEAR.**

GO ON TO THE NEXT PAGE.