AP® Statistics
1998 Scoring Guidelines

Free-Response Scoring Guidelines: Question 1

4 Complete Response

Standard symbols are acceptable without explanation, but non-standard symbols must be defined. The question asks for exact results, but a student can receive full credit for CAREFULLY explaining what might happen in a simulation.

(a) The mean of the sampling distribution is equal to the mean of the population.

(b) The standard deviation of the sampling distribution is equal to the standard deviation of the population divided by the square root of the sample size.
OR
Clearly states that the standard deviation of the sampling distribution decreases as n increases.

(c) The equivalent of the following two statements must be included:

1. The sampling distribution is skewed for small sample sizes. (A statement that does not use the term skewed but says the distribution will be non-normal is OK.)

2. The shape of the sampling distribution gets more and more normal-like (bell shaped) as the sample size increases.

3 Substantial Response

States <u>both</u> (a) and (b) correctly; has a weak, but correct, response on part (c). A weak response for (c) would include correctly one of the two statements in (c) above, but not both.
OR
States <u>either</u> (a) or (b) correctly and gives a complete and correct response to (c).

2 Developing Response

States <u>both</u> of (a) and (b) correctly but gives an incorrect response to (c).
OR
Gives a correct response to <u>either</u> (a) or (b), but gives a weak response to (c).
OR
Gives an incorrect response to both (a) and (b), but gives a complete and correct response to (c).

1 Minimal Response

States <u>one</u> of (a) or (b) correctly.
OR
Has substantive errors in (a) and (b), but gives a weak response to (c).

Notes:

Some students appear to have confused the *sampling distribution* with the histogram for a particular sample. There were a number of papers that had responses containing "the sample mean is close to" or "gets close to the population mean as n increases," or other rewordings of the law of large numbers. These statements, while true, do not answer the questions posed. These incorrectly worded responses may surround the correct formula in parts (a) and (b). If the written response is irrelevant but does not contradict the formula, credit can be awarded; if the written response directly contradicts the formula, credit should not be given.

Free-Response Scoring Guidelines: Question 2

4 Complete Response

(a) Shows a correct histogram (or bar graph) of the number of defective items produced per day, as shown below. The vertical scale could be either frequency or relative frequency $(8/20 = .40)$.
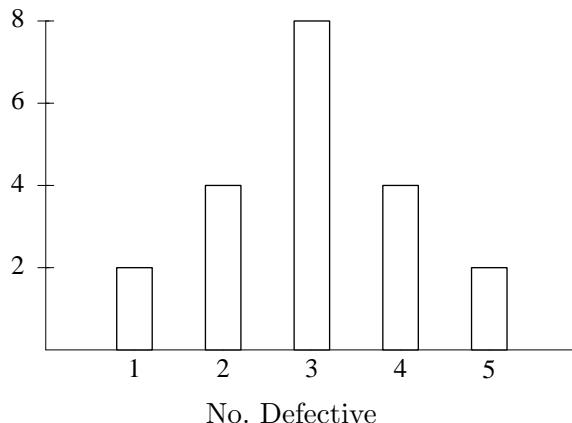
No. Defective

(b) The most obvious fact from the histogram is that the number of defective items has a symmetric, mound-shaped, bell-shaped or <u>approximately normal</u> distribution. Any one of these terms (or phrases that convey the same notion) will be considered as correct. The term <u>normal distribution</u> is not acceptable for a complete response.

Other facts that might be reported include the mean of 3 and/or some measure of spread, such as the standard deviation of approximately 1.1, but numerical summaries are not required, whereas some statement on shape is required.

(c) The most obvious fact from the scatterplot is that the number of defects per day has a decreasing, linear trend across the days. This point should not be missed even if other correct information, such as the larger spread for the later days, is given. Statements such as negative correlation, negative association and negative slope are acceptable as correct answers here.

3 Substantial Response

(a) Draws a histogram that is essentially correct, but may have a minor error such as an error in counting or scaling.

(b) Makes a nearly correct statement about the shape of the histogram. Stating that the number of defects is normally distributed without qualification will reduce the score to a maximum of 3.

(c) Makes a correct, but perhaps weak or incomplete, statement on the decreasing trend over time in the scatterplot.

2 Developing Response

Draws a histogram nearly correctly and has a nearly correct description of the shape of the histogram. (The term normal distribution is acceptable here.) Makes no correct statement about essential features of the scatterplot.
OR
Draws a histogram incorrectly but then makes correct statements <u>based on that drawing</u>. Makes a

correct statement about a feature of the scatterplot that reflects some understanding of the graphical image portrayed.

OR

Draws a histogram nearly correctly but interprets it incorrectly. Makes a correct statement about a feature of the scatterplot that reflects some understanding of the graphical image portrayed.

1 Minimal Response

Has a partially correct response in at least one the three parts.

Free-Response Scoring Guidelines: Question 3

4 Complete Response

(a) A complete response should try to make the six treatment groups look as nearly alike as possible, except for the treatment, so as to balance out any characteristics (seen or unseen) that might affect chances of successful migration.

Some ways to do this are as follows:
- Randomly allocate the 3600 butterflies to groups of 600 for each marking location. A correctly described method does not necessarily have to be practical. Although it would be difficult, the 3600 butterflies could be each assigned a unique number and then a random number table could be used to select 600 for position A, and so on. A more practical randomization could be accomplished by putting 3600 slips of paper n a bag, 600 for each letter. As each butterfly is captured, a slip is randomly selected from the bag (without replacement) to indicate the marking location for the butterfly.
- Divide the 3600 butterflies into groups smaller than 600 and then randomly assign an equal number of groups to each marking location.
- Systematically assign marking locations A through F to successive groups of 6 butterflies until all 3600 have been assigned.

Note: Responses for (a) should be read carefully to assess understanding of randomness. For example, only indicating an allocation of 600 butterflies per location is an incorrect response. Only stating that 600 butterflies are randomly assigned to each marking location is an incomplete response.

(b) Generate six possible outcomes independently, each with probability 1/6, perhaps by tossing a balanced die. Assign one number to each location, such as 1 = A, 2 = B, and so on. Toss the die and assign the marking location sequentially to the 3600 butterflies as they are caught. The order of assignment does not matter.

(c) The analysis requires a test of the null hypothesis that the proportions of butterflies that successfully migrate are the same for all locations. Either a frequency table equivalent to the one shown below can be indicated as the proper way to arrange the data, or a written description of how the counts are tabulated and proportions compared is acceptable.

	A	B	C	D	E	F
Successful						
Not Successful						
Number Launched						

The student must indicate an appropriate test statistic. A chi-square test is anticipated. Multiple comparisons of proportions would also work, but this requires 15 simple comparisons. (The student need not indicate degrees of freedom or formulas. However, if these are included, they must be correct.)

Note: No credit is given for writing down "chi-square test" without an essentially correct method of assigning treatments to the butterflies.

(d) In (a) the number launched is fixed at 600 for each marking location. The chi-square test for the equality of proportions is still an acceptable test, so the method of analysis need not be changed.

1998 AP STATISTICS

3 Substantial Response

Gives essentially correct methods for (a) and (b) but has an error in one of the methods of analysis in (c) or (d). A common error is to think that the chi-square test will not work when the sample sizes are random, part (c), and to try to concoct a new test here.
OR
Gives essentially correct methods for (a) and (b) and, in (c) and (d), indicates how to tally the counts and that the analysis requires a comparison of proportions (but does not name a specific test to analyze this comparison).
OR
Has an incomplete response for either (a) or (b), but indicates appropriate analysis in (c) and (d).

2 Developing Response

Gives an essentially correct method for one of (a) or (b) and a correct analysis that goes with it in (c) or (d). Has a substantive error in one of (a) or (b) and the related analysis.
OR
Gives an essentially correct method for (a) and (b), but (c) and (d) are incorrect or blank.
OR
Gives an essentially correct method for (a), but (b) is incorrect or missing, and (c) and (d) have incomplete analyses.

1 Minimal Response

Gives an essentially correct method for (b), but (a), (c), and (d) are incorrect or missing.
OR
Gives an essentially correct method for (a), but (b), (c), and (d) are incorrect or missing.
OR
Has an incomplete response for (a) or (b), and in (c) or (d) indicates how to tally the counts and that the analysis requires a comparison of proportions.

Free-Response Scoring Guidelines: Question 4

4 Complete Response

(a) Correctly gives equation of the regression line as $\hat{y} = -20.5893 + 24.3929x$ (Could use y.)
and defines both variables: $x = \#$ of teaspoons of weed killer; $\hat{y} = \%$ killed
OR
percent killed $= -20.5893 + 24.3929(\#$ of teaspoons of weed killer$)$

(b) Substitutes $x = 2.6$ into the regression equation to get a predicted value of 42.83224, and notes
that the residuals around the predicted value of 42.8 (or the middle of the predicted values)
are negative. Concludes that since the residual for this prediction is negative, the prediction
is expected to be too large.
OR
Notes that $x = 2.6$ is about in the middle of the explanatory values and, hence, the predicted
percent killed will be close to the middle of the predicted values. Concludes that since the
residual for this prediction is negative, the prediction is expected to be too large.
OR
Notes that $x = 2.6$ is in the middle of the explanatory values and that the residuals as
a function of the explanatory values must exhibit the same pattern of positive and negative
residuals. Since the residuals in the middle of the explanatory values are negative, the predicted
value is expected to be too large.

- Arithmetic errors in (b) that give reasonable predictions (i.e. predictions between 20 and
60) should not be penalized.

3 Substantial Response

Gives a correct answer for either parts (a) or (b) and a partially correct answer to the other part.

Partially correct answers include but are not limited to:

- (a) Gives the correct equation but fails to define both variables.
- (a) Switches the values for slope and y-intercept in the equation but defines both variables.
- (a) Defines both variables but gives only one correct coefficient in the correct place of the linear
equation.
- (b) Correctly explains why the residual at $x = 2.6$ is negative, but incorrectly interprets this
negative residual to mean that the predicted value will be too small.
- (b) States that the residual at $x = 2.6$ is negative and thus the predicted value will be too large
but fails to specify where the predicted residual for $x = 2.6$ is relative to the other residuals on
the residual plot.
- (b) Uses the correct model and gets an incorrect prediction that is not reasonable, but reasons
correctly using this prediction. (Unreasonable predictions are below 20 or greater than 60.)
- (b) Gives incorrect residual but interprets it correctly.

2 Developing Response

Gives a correct answer to one of (a) or (b) but not a partially correct answer to the other
OR
gives a partially correct response to both (a) and (b).

1998 AP STATISTICS

The following responses receive scores of zero for (b):

- Saying that the line cannot be used for prediction or saying that a prediction cannot be based on the computer output because the appropriate model should be quadratic or because the residuals have a non-random pattern. (An r^2 of 97.2% indicates that the fit is good; the residual plot merely reveals that the fit could be improved.)

- Computing a prediction (usually 42.8%) from the equation in (a) and saying that the prediction is "too large" without giving any reasoning from the prediction value to the conclusion of "too large."

1 Minimal Response

Gives a partially correct answer to either (a) or (b), but not both.

Free-Response Scoring Guidelines: Question 5

4 Complete Response

The response is substantially complete for all four sections.

3 Substantial Response

The response is substantially complete for three of the four sections.

2 Developing Response

The response is substantially complete for two of the four sections.

1 Minimal Response

The response is substantially complete for one of the four sections, or the response must show some evidence that they understand the role of variation in making inference from data.

Solutions and Scoring

Section 1: Checking the assumptions in Part (a)

Students must show (1), or they may show both (2) AND (3) in the following list:
1. Shows $np > 10$ and $n(1 - p) > 10$ with numbers substituted for n and p; must use $p = 0.1$.
2. States only that $np > 10$ and $n(1 - p) > 10$ [$np > 5$ and $n(1 - p) > 5$ are OK].
3. States that the observations must be independent or points out SRS conducted.

Section 2: Doing the mechanics in Part (a)

Students should correctly state null and alternative hypotheses.

Because the 481 respondents appear to constitute a random sample, it is appropriate to test the null hypothesis: $p = .10$ against the alternative hypothesis: $p > .10$ where p denotes the true proportion of graduate students who prefer campus housing.

Students should show that the test statistic is:
$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0.129 - 0.100}{\sqrt{\frac{0.1(0.9)}{481}}} = 2.12 \quad \text{where } \hat{p} = \frac{62}{481} = 0.129,$$
the observed sample proportion of those who desire on-campus housing.

Students should show the associated p-value, .017 (one-sided), or they may use an appropriate rejection region approach.

To receive credit for Section 2, the student MUST state the null hypothesis and a 1-sided alternative and give a p-value (or appropriate rejection region). Overall, a student cannot receive a score of 4 on this question if they do not identify, by name or by formula, the statistical procedure they use. For example, they may say "1-sample test of a proportion" or give the formula for the test statistic (using symbols or numbers). If the score is reduced to 3 by some other error, failure to identify the statistical procedure used will not further reduce the score by another point.

Do not give credit for Section 2 if more than one or two of the following errors occur:

- Uses non-standard notation that is not defined
- Uses incorrect notation in either hypothesis
- Gives incorrect Z value because $\hat{p}$ was used in calculation of standard error
- Gives incorrect Z value because of minor arithmetic errors
- Does not show the calculation of the test statistic
- Makes inference based on 1 minus the correct p-value
- Gives a 2-sided p-value for a 1-sided test

Section 3: Writing the conclusion in Part (a)

Students must connect the work done in Section 2 with the conclusion given in Section 3 by noting that the observed test result is a rare event. They might say:

> Because the p-value is less than an alpha-level the student selects, OR because the p-value is so small, OR because there are only 17 chances in 1,000 of getting a z-value this large or larger when null hypothesis is true therefore we reject the null,

The student must ALSO state the conclusion in the context of the problem. For example,

> There is evidence to believe that more than 10% of the graduate students desire on-campus housing.

If a significance level is stated at .01, then the conclusion should be that there is insufficient evidence to say that more than 10% of the graduate students desire on-campus housing. The conclusion must be consistent with the alpha level the student gives.

A student should not lose credit for this section if they make one error such as stating that "the hypothesis has been proven" if this section is otherwise correct.

Section 4: Part (b)

Students must recognize that the situation most detrimental to the conclusion reached in part (a) would be the case in which all 19 said that they were not looking for housing on campus, and they must evaluate their effect on the p-value of the test statistic and on their conclusion in part (a).

In this situation, $\hat{p} = \dfrac{62}{500} = 0.124$

and the test statistic becomes $z = 1.79$ and the p-value is .0368. Although this p-value is larger than the one in part (a), there is still evidence to conclude that more than 10% of the graduate students desire on-campus housing (an any fixed significance level greater than .0368).

If the decision in Part (a) is not to recommend increasing the amount of housing, the student should evaluate
$\hat{p} = \dfrac{81}{500} = 0.162$
and the test statistic becomes $z = 4.62$ and the p-value is less than .000002. In this case the decision in Part (a) would change.

Other Possible Solutions Include:

(1) **Exact Binomial**

   Part (a) $X \sim \text{BINOMIAL}(n = 481, p = .1)$ $P(X \geq 62) = .0238$

Part (b) $X \sim \text{BINOMIAL}(n = 500, p = .1)$ $\text{P}(X \geq 62) = .0465$

(2) **Normal Approximation to Binomial**

Part (a): $X \sim \text{BINOMIAL}(n = 481, p = .1) \Rightarrow X$ approximately $\text{N}(48.10, 6.58)$

$Z = \dfrac{x - \mu}{\sigma} = \dfrac{62 - 48.1}{6.58} = 2.113$ and $\text{P}(Z \geq 2.113) = 0.0173$

OR

with the correction for continuity: $Z = 2.037$ and $\text{P}(Z \geq 2.037) = 0.0208$

Part (b): $X \sim \text{BINOMIAL}(n = 500, p = .1) \Rightarrow X$ approximately $\text{N}(50.00, 6.71)$

$Z = \dfrac{x - \mu}{\sigma} = \dfrac{62 - 50}{6.71} = 1.789$ and $\text{P}(Z \geq 1.789) = 0.0368$

OR

with the correction for continuity: $Z = 1.714$ and $\text{P}(Z \geq 1.714) = 0.0432$

(3) **Chi-Square Test**

(4) **Appropriate One-Sided Confidence Interval**

1998 AP STATISTICS

Free-Response Scoring Guidelines: Question 6

4 Complete Response

The response is substantially correct for all three sections, I, II, and III.

3 Substantial Response

The response is substantially correct for two of the three sections, I, II, and III, or the response is substantially correct for one of the three sections and nearly correct for each of the other two.

2 Developing Response

The response is substantially correct for one of the three sections, I, II, and III, and shows an acceptable approach in at least one of the other sections.

1 Minimal Response

The response is substantially correct for one of the three sections, I, II, and III, or shows an acceptable approach in at least one of the three sections.

Solutions and Scoring

Section 1: Finding a probability

(a) Using the normal density function, determines that $P(7 < X < 9) = 0.38$, which rounds to 0.4.

A "substantially correct" response must show some indication of how this value was determined. For example, either

$P(7 < X < 9) = P(-0.5 < Z < 0.5) = 0.6915 - 0.3085 = 0.3830$

or

Normalcdf$(7, 9, 8, 2)$

is fine. If there is no explanation for the 0.38, this is considered only an "acceptable approach."

Other acceptable approaches, but not complete responses, include answers where normal curve calculations or geometry is incorrect. Such acceptable approaches include answers of $1 - 0.3830 = 0.6170$ and 0.1915. Note that the answer $0.5(0.68) = 0.34$ should receive no credit.

Section 2: Setting up and running a simulation

(b) Shows a listing of four of the ten digits defined to be equivalent to finding a pearl of the correct size (a "success") and states that the other digits represent finding no pearl. The value used in (b) must match the value calculated in (a). If no solution is given in (a), a value can be "made up" and the problem completed. It is OK to use a two digit simulation with $p = 0.38$.

Describes a correct method of sampling digits from a random digit table until two successes are found.

To get any credit for Section II, the student must describe a waiting time simulation. A correct assignment of digits in (b) is not sufficient to get any credit if the waiting time component is missing.

(c) Correctly identifies a success, based on the definition in part (b) and counts the number of digits it took to get two successes.

Correctly runs three trials (more is OK) and records the number of digits it took to get two

successes. (Starting points for the trials will vary.)

If the assignment of digits in (b) is incorrect, the response should be considered an "acceptable approach" if a waiting time simulation is carried out correctly in part (c) according to the assignments of digits in part (b).

Section 3: Interpreting results of a simulation

(d) Identifies A as the correct distribution for $P = 0.4$ by noting that the probability of finding a pearl between 4 and 6.5 mm is smaller than the probability of finding a pearl between 7 and 9 mm, and it therefore takes longer to find two pearls between 4 and 6.5 than two pearls between 7 and 9. If the student computes the probability of finding a smaller pearl, this should be considered a bonus. ($P(4 < X < 6.5) = 0.2039$.) If the argue that a pearl between 7 and 9 is more likely because this interval contains the mean or is closer to the mean than $[4, 6.5]$, this is not a major error but is considered a minus. If the student selects Distribution A with no reason or an incorrect reason, they get no credit on part (d). If the student selects a distribution based on how well it matches their simulation, this is an "acceptable approach."

Another correct response is to observe that the mean should be $2(1/0.38) = 5.26$, which is closer to 5 than to 10.

(e) The best estimate of the expected number of pearls is the sample mean of 5.16, for distribution A. This can be arrived at by direct computation using the formula

$$\overline{x} = \frac{\sum x \cdot f}{100} = 5.16$$

by estimating the balance point of the distribution, or by using the formula $2(1/0.38) = 5.26$. [The standard deviation is 2.56 in case anyone uses it.] A computational mistake using the right formula is okay, if the answer is reasonable.

If no indication is given of how the answer was determined, this is an "acceptable approach," but not a "complete response." However, if no work was shown in part (a) as well, don't count off again.

If a student uses the median of 4 or the mode of 3 as the best measure of center with justification, this is OK. Without justification, it is an "acceptable approach."

If the student selects distribution B in part (d), the mean should be 9.36, the median 8, and the mode 7 (and the standard deviation is 5.27).