Question 6

Intent of Question

The primary intent of this question is to assess a student's ability to: (1) make an inference about the difference in two population proportions; (2) examine a regression model for a linear trend in proportions; and (3) construct a confidence interval for a slope. The investigative part of this question requires a student to use a regression model to estimate survival probabilities for two different situations and make an inference about the expected number of surviving species that would be achieved.

Solution

Part (a):

Part 1: States a correct pair of hypotheses

$$H_o : p_L - p_S = 0 \quad \text{versus} \quad H_a : p_L - p_S < 0$$

OR

$$H_o : p_S - p_L = 0 \quad \text{versus} \quad H_a : p_S - p_L > 0$$

OR

$$H_o : p_L = p_S \quad \text{versus} \quad H_a : p_L < p_S$$

Where

p_L is the proportion going extinct on large islands,
and
p_S is the proportion going extinct on small islands.

Part 2: Identifies a correct test (by name or by formula) and checks appropriate assumptions.

Two-sample test for proportions

$$z = \frac{\hat{p}_L - \hat{p}_S}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n_L} + \dfrac{\hat{p}(1-\hat{p})}{n_S}}}$$

Assumptions: independent observations and large sample sizes.

The problem states that whether one species becomes extinct is independent of whether another species becomes extinct, and that the probability of extinction is the same for all species on large islands and for all species on small islands, so it is reasonable to assume that observations are independent.

Question 6 (continued)

$\hat{p}_L = 0.091 \quad \hat{p}_S = 0.221$
$n_L \hat{p}_L = 19 \quad n_L(1 - \hat{p}_L) = 189$
$n_S \hat{p}_S = 66 \quad n_S(1 - \hat{p}_S) = 233$

All are greater than 5 (or 10), so the sample sizes are large enough to proceed.

Part 3: Correct mechanics, including the value of the test statistic and p-value (or rejection region).

$$\hat{p} = \frac{19 + 66}{208 + 299} = \frac{85}{507} = 0.168$$

$$z = \frac{\hat{p}_L - \hat{p}_S}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_L} + \frac{\hat{p}(1-\hat{p})}{n_S}}} = \frac{0.091 - 0.221}{\sqrt{\frac{(0.168)(0.832)}{208} + \frac{(0.168)(0.832)}{299}}} = \frac{-0.130}{0.034} = -3.82$$

p-value = 0.00006

(from table p-value ≈ 0; graphing calculator: $z = -3.836233478$, p-value = 0.00006)

Part 4: Stating a correct conclusion in the context of the problem, using the result of the statistical test.

Because the p-value is less than the stated α (or because the p-value is so small, or because the test statistic is in the rejection region), reject H_o. There is sufficient evidence that the proportion of species becoming extinct is smaller for large islands than for small islands.

If both an α and a p-value are given, the linkage is implied. If no α is given, the solution must be explicit about the linkage by giving a correct interpretation of the p-value or explaining how the conclusion follows from the p-value.

If the p-value in part 3 is incorrect but the conclusion is consistent with the computed p-value, part 4 can be considered as correct.

Part (b):

Compute a 95 percent confidence interval for the slope of the regression line.

Part 1: Identifies appropriate confidence interval by name or by formula.

The confidence interval for the slope of the regression line is $b \pm ts_b$.

Part 2: Checks appropriate assumptions.

Assumptions: The residual plot shows no unusual patterns that would suggest violation of the assumptions, so it is reasonable to proceed.

Part 3: Correct mechanics.

df = $n - 2$ = 13-2 = 11

$$-0.05323 \pm 2.20(0.00618)$$
$$-0.05323 \pm 0.013596$$
$$(-0.0668, -0.0396)$$

Part 4: Interpretation.

We are 95 percent confident that the mean proportion of species going extinct decreases by somewhere between 0.03 and 0.06 with each increase of 1 unit in ln(area). The proportion of species going extinct decreases with increasing area.

Part (c):

From part (b) it appears that the proportion of species going extinct decreases with increasing area. Therefore the proportion of species going extinct is related to the size of the island. Because the island sizes differed within the large island group and within the small island group, the assumption is probably not reasonable.

Scoring

Each part is scored as either essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is essentially correct (E) if three or four parts of the hypothesis test are correct.

Part (a) is partially correct (P) if one or two parts of the hypothesis test are correct.

NOTE: For part 2 of (a), the independent observations assumption does not have to be addressed in the response to get credit for this part, since this is given in the stem of the problem.

Part (b) is essentially correct (E) if three or four parts of the confidence interval are correct.

Part (b) is partially correct (P) if one or two parts of the confidence interval are correct.

Part (c) is essentially correct (E) if the response:

1. states the assumptions are not reasonable, *AND*
2. gives a justification based on the information in part (b);

OR

says that the assumptions are reasonable based on an incorrect conclusion in part (b) that island size is not related to extinction proportion, with an appropriate explanation.

Question 6 (continued)

Part (c) is partially correct (P) if:

it says that the assumption is not reasonable, but the explanation is weak or does not appeal to the information in part (b);

OR

it says that the assumption is not reasonable because the negative estimate of the slope given in part (b) is misinterpreted to suggest that survival rates decrease as area increases;

OR

it appeals to part (b) but says that the assumption is reasonable because within each group (large/small), the island sizes don't vary too much;

OR

it says the assumption is reasonable because the negative estimate of the slope given in part (b) is misinterpreted;

OR

the justification appeals to the differing proportions in the original data table only.

Part (c) is incorrect if a choice is made but no justification is given.

Part (d) is essentially correct (E) if the large preserve is chosen and the decision is well supported based on the expectation that a larger number of species will be preserved, in comparison to the expected number preserved on the five small islands.

Part (d) is partially correct (P) if:

the large preserve is chosen based on the results from parts (a) and/or (b);

OR

the large preserve is chosen but the justification is weak;

OR

the five small preserves are chosen based on an incorrect computation of the number of species saved for the two scenarios.

Part (d) is incorrect if:

a choice is made (large or five small) but no justification is given;

OR

five small preserves are chosen based only on the fact that there are 80 rather than 70 species at the outset.

Question 6 (continued)

4 Complete Response

All four parts essentially correct

3 Substantial Response

Three parts essentially correct and no parts partially correct

OR

Two parts essentially correct and two parts partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and two parts partially correct

OR

Four parts partially correct

1 Minimal Response

One part essentially correct and no parts partially correct

OR

No parts essentially correct and two parts partially correct

If a response is between two scores (for example, 2½ points), use a holistic approach to determine whether to score up or down depending on the strength of the response and communication.

(a) One scientist involved in the study believes that large islands (those with areas greater than 25 square kilometers) are more effective than small islands (those with areas of no more than 25 square kilometers) for protecting at-risk species. The scientist noted that for this study, a total of 19 of the 208 species on the large islands became extinct, whereas a total of 66 of the 299 species on the small islands became extinct. Assume that the probability of extinction is the same for all at-risk species on large islands and the same for all at-risk species on small islands. Do these data support the scientist's belief? Give appropriate statistical justification for your answer.

Population parameters of interest

π_1 = proportion of species to become extinct on large islands

π_2 = proportion of species to become extinct on small islands

Research hypotheses

$H_0: \pi_1 = \pi_2$

$H_a: \pi_1 < \pi_2$

Test of significance

We will use a ~~1~~ two proportion Z-test to evaluate these hypotheses. This test is justified because:

- The sample is of sufficient size. $n\pi_1 = 19$ $n\pi_2 = 66$
  $n(1-\pi_1) = 189$ $n(1-\pi_2) = 233$
  All of these are greater than 5.

- The data was obtained through a ~~simple~~ census, which is equivalent to a simple random survey for this test.

Test Statistic

$z = -3.836$ $\hat{p}_1 = 0.09$

$p < 0.001$ $\hat{p}_2 = 0.22$

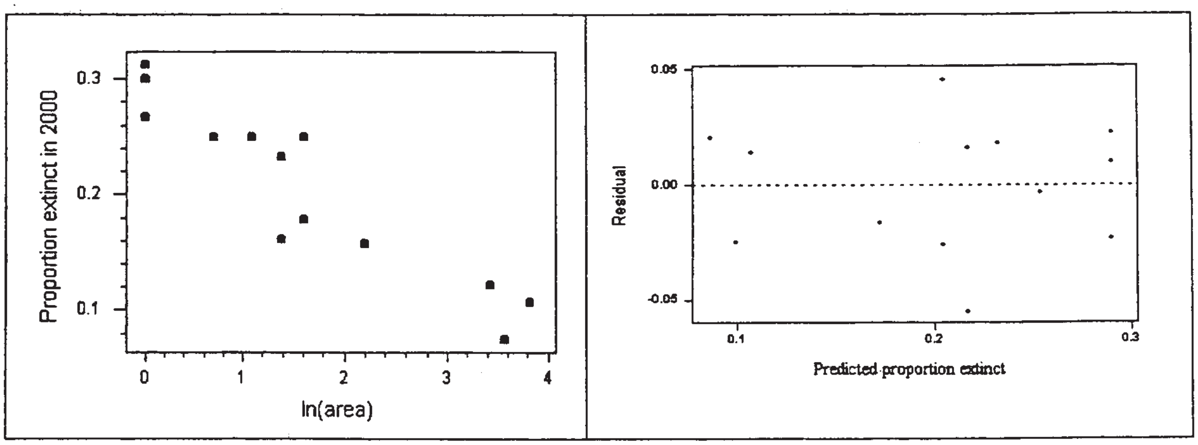
Conclusion

We reject H_0 at any reasonable level of significance. This test supports the scientist's belief; large islands seem to have a smaller proportion of extinctions than small islands.

GO ON TO THE NEXT PAGE.

(b) Another scientist who worked on this study thinks that the proportion of species that become extinct is more directly related to the size of the islands than simply to whether the islands are grouped as large or small. This scientist investigated the relationship between the proportion of extinct birds and the area, in square kilometers, of islands. A least squares analysis was conducted on the proportion extinct and ln(area). The regression analysis output, the scatterplot, and the residual plot are shown below.

Predictor	Coef	StDev	T	P
Constant	0.28996	0.01269	22.85	0.000
ln(area)	-0.05323	0.00618	-8.61	0.000
S = 0.02863	R-Sq = 87.1%			



Estimate the slope of the least squares regression line using a 95 percent confidence interval. Interpret your answer in the context of this situation.

95% CI for slope (B): statistic ± critical value · st dev

~~0.28996 ± 1.96 · 0.00618~~

~~0.28996 ± 0.0121128~~

~~⊖ 0.2778   ⊕ 0.3021~~

$-0.05323 \pm 1.96 \cdot 0.00618$

-0.05323 ± 0.0121128

⊖ -0.0653 ⊕ -0.0411

$(-0.0653, -0.0411)$

The interval ~~(0.2778, 0.3021)~~ captures the true slope with 95% confidence. This implies that the slope is negative; as land area increases, the proportion of extinct species decreases.

The use of this technique is justified as the residuals are small and random.

Approximately 87.1% of variation in extinction proportion can be explained by a linear relationship ~~variation~~ with ln(area).

GO ON TO THE NEXT PAGE.

(c) In part (a), the scientist assumed that the probability of a species becoming extinct is the same for each of the large islands. Similarly, the scientist assumed that the probability is the same for each of the small islands. Based on your answer in part (b), do you think this is a reasonable assumption? Explain.

Part B shows that the proportion of extinct species varies with the natural log of island area. This would imply that arbitrary groupings may be inappropriate. However, the island size in the study could be considered bimodal. Scientist A's classifications reflected this aspect of the observational study and they may be valid in this case.

(d) A conservation group with a long-term goal of preserving species believes that all at-risk species will disappear whenever land inhabited by those species is developed. It has an opportunity to purchase land in an area about to be developed. The group has a choice of creating one large nature preserve with an area of 45 square kilometers and containing 70 at-risk species, or 5 small nature preserves, each with an area of 3 square kilometers and each containing 16 at-risk species unique to that preserve. Which choice would you recommend and why?

Based on A's research, we can expect about 0.09 of the 70 at risk species in the large preserve to become extinct and about 0.22 of the 80 total at risk species in the small preserves to become extinct.

B's research implies that the proportion of species to become extinct can be predicted by the equation:

$$\text{prop. extinct} = -0.05323(\ln \text{area}) + 0.28996$$

This table summarizes potential extinction ~~rates~~ proportions for each option:

	Size	Total at risk	# extinct predicted by A	# extinct predicted by B	# Saved
Large	45 sq Km	70	9% = 6.3 species	8.73% = 6.11 species	70 - (~6) = 64
Small	3 sq Km	80	22% = 17.6 species	23% = 17.6 species	80 - (~18) = 62

The conservation group should create one large preserve. The research of both scientists suggests that about 6 species will become extinct in the large preserve versus more than 17 species in the smaller preserves. Creating one large preserve will save the greatest number of species from extinction. It is important to note that the small preserves will contain 80 at risk species versus 70 in the large preserve; however, the large preserve still saves a larger number of species (64 as opposed to 62).

GO ON TO THE NEXT PAGE.

(a) One scientist involved in the study believes that large islands (those with areas greater than 25 square kilometers) are more effective than small islands (those with areas of no more than 25 square kilometers) for protecting at-risk species. The scientist noted that for this study, a total of 19 of the 208 species on the large islands became extinct, whereas a total of 66 of the 299 species on the small islands became extinct. Assume that the probability of extinction is the same for all at-risk species on large islands and the same for all at-risk species on small islands. Do these data support the scientist's belief? Give appropriate statistical justification for your answer.

Assumptions:

$n_L p_L = 19 \geq 10$

$n_L(1-p_L) = 189 \geq 10$

$n_S p_S = 66 \geq 10$

$n_S(1-p_S) = 233 \geq 10$

since all are greater than 10 distributions are normal by CLT.

Randomly selected islands stated,

π_L = pop. proportion of extinction on large islands

π_S = pop. proportion of extintion on small islands

$H_0: \pi_L - \pi_S = 0$

$H_a: \pi_L - \pi_S < 0$

$\alpha = 0.01$, because the scientists don't want false support for their theories

$$z = \frac{p_L - p_S - (\pi_L - \pi_S)}{\sqrt{\frac{p_L(1-p_L)}{n_L} + \frac{p_S(1-p_S)}{n_S}}}$$

$$z = \frac{\frac{19}{208} - \frac{66}{299} - 0}{\sqrt{\frac{\frac{19}{208}(1-\frac{19}{208})}{208} + \frac{\frac{66}{299}(1-\frac{66}{299})}{299}}}$$

$z = -4.145$

p-value = 0

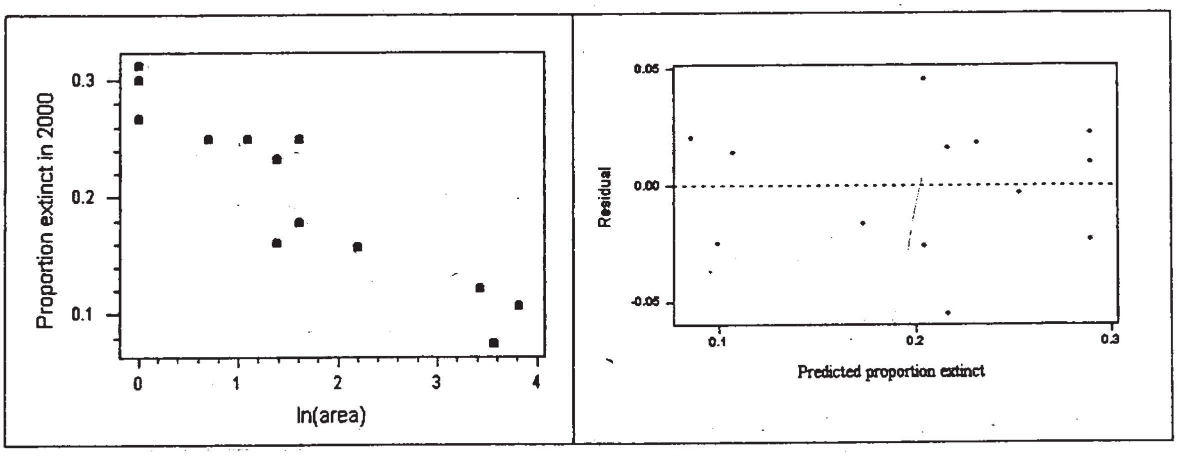
Because p-value $\leq \alpha$, I will reject H_0 is favor of H_a and conclude the scientists belief is correct that large islands are more effective than small islands for protecting at-risk species.

Unauthorized copying or reuse of any part of this page is illegal.

GO ON TO THE NEXT PAGE.

(b) Another scientist who worked on this study thinks that the proportion of species that become extinct is more directly related to the size of the islands than simply to whether the islands are grouped as large or small. This scientist investigated the relationship between the proportion of extinct birds and the area, in square kilometers, of islands. A least squares analysis was conducted on the proportion extinct and ln(area). The regression analysis output, the scatterplot, and the residual plot are shown below.

Predictor	Coef	StDev	T	P
Constant	0.28996	0.01269	22.85	0.000
ln(area)	-0.05323	0.00618	-8.61	0.000
S = 0.02863	R-Sq = 87.1%			



Estimate the slope of the least squares regression line using a 95 percent confidence interval. Interpret your answer in the context of this situation.

$\beta \in b \pm t\text{-crit.}\left(\frac{s_b}{\sqrt{n}}\right)$ $\quad df = n-1$

$\beta \in -0.05323 \pm 2.179\left(\frac{0.00618}{\sqrt{13}}\right)$ $\quad df = 13-1$

$df = 12$

$\beta \in (-0.056965, -0.049495)$

This interval is the interval that if samples of 13 islands were taken from the pop., 95% of these intervals would contain β, the least squares regression line of the population's slope.

GO ON TO THE NEXT PAGE.

(c) In part (a), the scientist assumed that the probability of a species becoming extinct is the same for each of the large islands. Similarly, the scientist assumed that the probability is the same for each of the small islands. Based on your answer in part (b), do you think this is a reasonable assumption? Explain.

I think it is unreasonable because the least squares regression line shows how proportion varies with area, instead of all large islands having same prob. and all small islands having same prob.

(d) A conservation group with a long-term goal of preserving species believes that all at-risk species will disappear whenever land inhabited by those species is developed. It has an opportunity to purchase land in an area about to be developed. The group has a choice of creating one large nature preserve with an area of 45 square kilometers and containing 70 at-risk species, or 5 small nature preserves, each with an area of 3 square kilometers and each containing 16 at-risk species unique to that preserve. Which choice would you recommend and why?

I would recommend the larger preserve because the at-risk species have a lesser chance of extinction. Even though the smaller reserves contain 80 at-risk species to the larger reserves 70, The proportion of extinction would be about 0.0873 at a larger reserve and 0.2315 at a small reserve. So at the larger reserve you could protect about 64 species while at the smaller reserve you could only protect about 51 species, so the larger reserve would allow you to protect more species.

GO ON TO THE NEXT PAGE.

(a) One scientist involved in the study believes that large islands (those with areas greater than 25 square kilometers) are more effective than small islands (those with areas of no more than 25 square kilometers) for protecting at-risk species. The scientist noted that for this study, a total of 19 of the 208 species on the large islands became extinct, whereas a total of 66 of the 299 species on the small islands became extinct. Assume that the probability of extinction is the same for all at-risk species on large islands and the same for all at-risk species on small islands. Do these data support the scientist's belief? Give appropriate statistical justification for your answer.

π_1 = proportion of species on the large islands who become extint

π_2 = proportion of species on the small islands who become extinct.

$H_0: \pi_1 = \pi_2$

$H_a: \pi_1 < \pi_2$

$\alpha = .01$

Assumptions:

Random sample (as stated in problem)

Sample is normal ($\pi \cdot n \geq 10$; $(1-\pi) \cdot n \geq 10$ $\rightarrow$ $\frac{19}{208} \cdot 208 \geq 10$, $\frac{189}{208} \cdot 208 \geq 10$, $\frac{66}{299} \cdot 299 \geq 10$; $\frac{233}{299} \cdot 299 \geq 10$)

I will conduct this test using the 2-Prop Z test on my calculator.

x1 : 19
n1 : 208
x2 : 66
n2 : 299
p1 : <p2

The results of this test:

$z = -3.836$

$p = 6.249 \times 10^{-5}$

$\hat{p}_1 = .09135$

$\hat{p}_2 = .2207$

$P(p_2 - p_1) = .12935$

$6.249 \times 10^{-5} < .01$

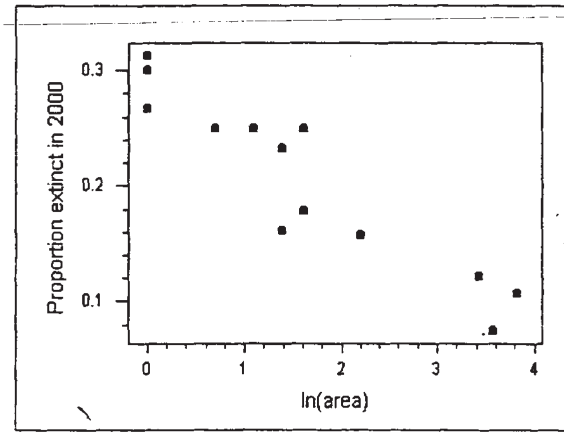
Therefore, reject Ho

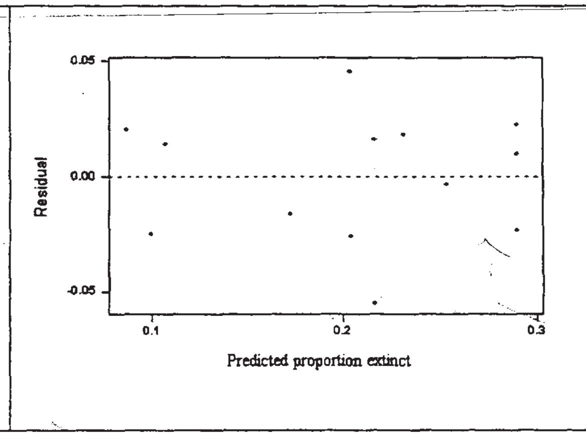
There is sufficient evidence to conclude that the proportion of species on the large islands who become extinct is less than the proportion of species on the small islands who become extinct.

GO ON TO THE NEXT PAGE.

(b) Another scientist who worked on this study thinks that the proportion of species that become extinct is more directly related to the size of the islands than simply to whether the islands are grouped as large or small. This scientist investigated the relationship between the proportion of extinct birds and the area, in square kilometers, of islands. A least squares analysis was conducted on the proportion extinct and ln(area). The regression analysis output, the scatterplot, and the residual plot are shown below.

Predictor	Coef	StDev	T	P
Constant	0.28996	0.01269	22.85	0.000
ln(area)	-0.05323	0.00618	-8.61	0.000
S = 0.02863	R-Sq = 87.1%			





Estimate the slope of the least squares regression line using a 95 percent confidence interval. Interpret your answer in the context of this situation.

least squares line ⇒ $\hat{y} = .28996 - 0.05323x$

95% confidence interval

stat ± z-score (std. dev)

$-.05323 \pm 1.96(.00618)$

$-.05323 \pm .0121128$

$(-.0653428, -.0411172)$

We are 95% confident the the true slope / β (for every 1 unit increase in ln(area), the proportion of extinct birds decreases) is between $-.0653428$ and $-.0411172$.

GO ON TO THE NEXT PAGE.

(c) In part (a), the scientist assumed that the probability of a species becoming extinct is the same for each of the large islands. Similarly, the scientist assumed that the probability is the same for each of the small islands. Based on your answer in part (b), do you think this is a reasonable assumption? Explain.

I think that the assumption is reasonable because according to the 95% confidence interval, the proportion of extinct birds decreases for every 1 unit increase in area.

(d) A conservation group with a long-term goal of preserving species believes that all at-risk species will disappear whenever land inhabited by those species is developed. It has an opportunity to purchase land in an area about to be developed. The group has a choice of creating one large nature preserve with an area of 45 square kilometers and containing 70 at-risk species, or 5 small nature preserves, each with an area of 3 square kilometers and each containing 16 at-risk species unique to that preserve. Which choice would you recommend and why?

I would recommend that the group purchase the one large nature preserve because both the the 95% confidence interval ~~the~~ results of part a show that the larger Islands have a smaller proportion of species who will be extinct.

GO ON TO THE NEXT PAGE.

Question 6

Sample: 6A
Score: 4

This is a complete response that provides an appropriate test of the null hypothesis that large and small islands have the same probability of species extinction. It also constructs and interprets a confidence interval for the slope of a regression line, uses the information provided in part (b) to determine that an assumption used in part (a) is inappropriate, and makes appropriate use of information provided in the problem to determine which of two situations would be expected to preserve more species. An essentially complete response is given in part (a). It provides the appropriate null and alternative hypotheses with good labeling of notation. A two-sample z-test is identified and justified by noting that all observed counts are sufficiently large. The test statistic and p-value are correctly evaluated, and an appropriate conclusion is reached about the scientist's belief. An essentially complete response is also provided for part (b). A formula is provided for a confidence interval for a slope, and it is evaluated with the appropriate standard error. The confidence interval is a bit too narrow because the 97.5-th percentile of the standard normal distribution is used instead of a more appropriate percentile of a t-distribution with $13 - 2 = 11$ degrees of freedom, but this does not seriously detract from the overall strength of the response. Use of the method for constructing the confidence interval is justified by appealing to the random pattern in the residual plot. The appropriate conclusion that probability of extinction decreases as land area increases is expressed with 95 percent confidence. The response to part (c) appeals to the result in part (b) to conclude that the assumption of homogeneous extinction probabilities across small islands is unreasonable because extinction probabilities decrease as island area increases. The response to part (d) computes expected numbers of species that would become extinct and survive for the two situations under consideration. This is done using the estimated extinction probabilities from the models in part (a) and part (b), although only the calculations for the regression model in part (b) are needed for a complete response. The response shows that the expected number of surviving species would be greater if 70 at-risk species are accommodated in a 45 square kilometer area than if 16 at-risk species are accommodated in each of five 3 square kilometer areas.

Sample: 6B
Score: 3

This is a substantial response that also provides an essentially complete test of the null hypothesis that large and small islands have the same probability of species extinction in part (a). It differs from the previous response in that the pooled estimate of the overall probability of extinction is not used to evaluate a two-sample z-test. Given the moderately large expected counts, the two tests are nearly equivalent and lead to the same conclusion. The test statistic and p-value are correctly evaluated, and an appropriate conclusion is reached about the scientist's belief. An incorrect standard error is used in the construction of the confidence interval in part (b). While a general statement is made about construction of confidence intervals, no conclusion is stated about the slope of the regression line fit to the species extinction data. Other errors include incorrect degrees of freedom and failure to use the information in the residual plot to help justify the method used to construct the confidence interval. An appropriate response is made to part (c) that uses the estimated regression line from part (b) to conclude that the assumption of homogeneous extinction probabilities across small islands is unreasonable. The response to part (d) is essentially correct, although the expected number of surviving species in the 5 smaller areas is incorrectly reported as 51 instead of 61.

Question 6 (continued)

Sample: 6C
Score: 2

This is a developing response that provides an essentially complete response to part (a). A formula is provided in part (b) for a confidence interval for a slope, and it is evaluated with the appropriate standard error; but the confidence interval is a bit too narrow because the 97.5-th percentile of the standard normal distribution is used instead of a more appropriate percentile of a t-distribution with $13 - 2 = 11$ degrees of freedom. This response does not refer to the uniformly random pattern in the residual plot to justify the method for constructing the confidence interval. The response to part (c) incorrectly concludes that the assumption of homogeneous extinction probabilities is reasonable based on a contradictory statement that extinction probabilities decrease as land area increases. The response to part (d) simply concludes that a larger land area should be used because it was shown in a previous part of this response that extinction probabilities are smaller in larger areas. The response to part (d) is not complete, because it does not consider that only 70 at-risk species can be accommodated in the 45 square kilometer area, while 80 species can be accommodated in the five smaller 3 square kilometer areas.