CollegeBoard
Advanced Placement Program

AP® Statistics
2005 Scoring Commentary

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.

Question 1

Overview

This question described a study comparing the caloric intake for a random sample of 20 students from one rural high school to a random sample of 20 students from one urban high school. The variable measured was the number of calories of food per kilogram of body weight consumed in one day by a student. A back-to-back stemplot displayed the number of calories consumed for each student on that day. Part (a) of the question asked students to use the stemplot to compare the data distributions from both schools. The response should have included a clear, correct comparison statement for each of the three characteristics: shape, center (position), and spread. It was not sufficient simply to list and describe those characteristics separately for each distribution. Part (b) asked whether it was reasonable to generalize the findings of the study to all rural and urban ninth-grade students in the United States. The response should have clearly indicated that generalizing was not appropriate, since only one urban school and one rural school were used. The sampling unit was the school, not the students chosen for the study or the area from which those students were selected. Part (c) presented two plans for consideration in conducting a similar study. Plan I used only one day; Plan II used the same 7-day period with a 7-day average computed for the number of calories consumed by each student. Plan II better met the goal of the study because it accounted for the effect of day-to-day variability. A correct justification should have indicated an understanding of what might cause systematic day-to-day variability in the difference between rural and urban students in the number of calories consumed on different days of the week (different amounts of calories consumed on a weekday versus weekend day), not just the advantage of a 7-day average.

Sample: 1A
Score: 4

Each part of this response is complete and clearly communicated. In part (a) the correct distribution shapes are stated. Measures of position (median, Q1, and Q3) for urban are compared ("all less than") to those for the rural. The spread of rural (using range) is compared to that for urban. There is a minor error (the IQR for urban is 7 rather than 5, and Q1 for rural is incorrect); however, specific values are not required, and the student was not penalized. In part (b) it is clear why generalization is not possible since students in the sample were from only two schools. The last sentence is not needed. In part (c) the first sentence states the advantage of "a 7-day period which would average out any days that a student might have eaten an extremely large amount or a very small amount"—conveying an understanding of the day-to-day variability. The last sentence further supports the justification of what might cause day-to-day variability. Students tend to personalize the day-to-day variability by talking about an individual student. This is not ideal but judged to include enough of the idea of day-to-day variability.

Sample: 1B
Score: 2

This is an example of a developing response. In part (a) the rural distribution is described as roughly symmetrical, and the urban distribution is described as skewed toward lower values. Although the skew is actually toward higher values, this is considered a minor error since visually turning the paper to make the urban stemplot horizontal appears to make the graph skewed left. Although values for center and spread are given for each distribution, the values are described separately for each distribution and no comparison is made. Only one out of three characteristics is correctly compared. In part (b) the first paragraph mentions the sample size is too small and only includes two schools. It is not clear if the small sample size refers to students or to schools. However, the following

Question 1 (continued)

two sentences clarify that the two schools have been correctly recognized as the sampling unit with too small of a sample size. In part (c) Plan II is selected, and the justification of day-to-day variability is indicated by discussing "the impact of unusually high or low days (such as a party or a day in which a meal was missed)."

Sample: 1C
Score: 1

This response provides a minimal, but complete, answer in part (a). The first sentence states: "Students in rural high school has [*sic*] higher median and range compared to students in urban high school." This sentence nicely emphasizes immediately the comparison of center and spread. The comparison of shape is given in the last sentence. Part (a) is a simple, straightforward presentation. In part (b) the first sentence states that "the sample size is small." There is no reference to whether the sampling unit is the number of schools or the number of students. The comment on confounding variables is considered extraneous. Thus, the student does not adequately answer this question. In part (c) Plan II is selected with the statement that "a week period of food he or she consume [*sic*], the data would be more reliable" giving a hint of the day-to-day variability. This is a weak justification for saying more days are better than one day.

Question 2

Overview

This problem presented students with a discrete probability distribution for the number of telephone lines in use by a technical support center at noon each day. In part (a) of the question, students were asked to find the expected value of the random variable. In part (b) they had to compare the behavior of the mean of a random sample of size 1,000 from the given distribution to the mean of a random sample of size 20 from that distribution. Part (c) gave a definition for the median of a discrete random variable and asked students to compute the median of this random variable. In part (d) students had to comment on the relationship between the skewness in the given distribution and the mean and median calculated in parts (a) and (c).

Sample: 2A
Score: 4

The response shows excellent understanding of the concepts being tested. For parts (a) and (c) both the mean and the median are computed correctly, with correct application of the definition of median that is given in the statement of the problem. In part (b) the response includes a description of the effect of the increase in sample size on the sample mean and supports that statement with an excellent description of how the variability in the sample mean decreases as the sample size increases. The response in part (d) includes a correct statement that the distribution is right-skewed and a description of the effect of the right-skewed distribution on the relationship between the mean and the median. It is not necessary to restate the values of the mean and the median found earlier in the problem.

Sample: 2B
Score: 3

Conceptually this response is quite good. However, in parts (a) and (c) even though correct values are given for the mean and the median, there is no work to support the value of the median. Students are expected, for any numerical result, to give some indication of how that value was computed. Part (b) is well done both in the statement of what should occur and in the justification for the statement. The response indicates clearly that larger samples should produce means closer to the expected value. Part (d) is answered very well, with supporting evidence in the form of a probability histogram.

Sample: 2C
Score: 2

The response indicates knowledge of the concepts but does not articulate those concepts well. In parts (a) and (c) the mean is computed correctly, but the median is incorrect even though there is an attempt to apply the definition that was given. Part (b) is not complete. The response includes a correct statement of the effect of the increase in sample size on the sample mean, but the justification is weak. To get full credit for this part, the response should have included an explanation of how the Law of Large Numbers applies in this context. Part (d) is answered correctly, stating that the right-skewed distribution is the reason why the median is less than the mean.

Question 3

Overview

In this question, students were presented with a data table, scatterplot, residual plot, and computer output from a linear regression analysis. Part (a) of the question asked students to evaluate the appropriateness of a linear model. They were expected to use graphical evidence to make this determination. Some students argued that a value of r^2 close to 1 or a value of r close to ± 1 indicated that a linear model was appropriate, but this was not correct. There are numerous examples of relationships between two quantitative variables in which r^2 is close to 1 and r is close to ± 1, but that is where the scatterplot shows a nonlinear relationship and the residual plot shows a clear pattern. In part (b) students had to recognize that the estimated slope from the computer output was needed and then use the estimated slope to compute a point estimate of the change in average cost of fuel per mile for each additional railcar. Part (c) required students to identify the value of r^2 from the computer output and to interpret this value in context. Ideally, students would have described the r^2-value as the proportion (percent) of variation in the fuel consumption that was accounted for by the linear model relating fuel consumption and number of railcars. In part (d) students were asked whether it was reasonable to use the linear model to make a prediction for a value of the explanatory variable that is far beyond the range of the data.

Sample: 3A
Score: 4

In part (a) the student's comment that "the original data appears linear" is too vague on its own to earn credit. However, the subsequent statement about the residual plot being "randomly distributed" is sufficient. The student gives a clear explanation for a correct calculation in part (b). Although the response makes no mention of the linear model in part (c), it does convey a generally correct understanding of what r^2 measures. In part (d) the response shows a clear understanding of why extrapolation is not appropriate.

Sample: 3B
Score: 3

The student appropriately refers to the residual plot as evidence supporting the use of a linear model in part (a). However, the student also incorrectly appeals to r^2 as justification for using a linear model. In addition, the response incorrectly identifies r^2 as 96.3 percent and gives an incomplete interpretation of this value. In part (b) the student correctly computes the average cost and clearly describes the computation. The student again uses the incorrect value of r^2 in part (c) and then proceeds to interpret it awkwardly. In part (d) the response indicates a clear understanding of extrapolation.

Question 3 (continued)

Sample: 3C
Score: 2

In part (a) the student refers incorrectly to r^2 but correctly to the relevant characteristic (no pattern) of the residual plot. Note that a high value for r^2 does not provide evidence that a linear model is appropriate. The first line in the student's response to part (b) is unexpected. However, the student seems to ignore this line when computing the point estimate. In part (c) the student appears to have a general sense of what r^2 measures but is unable to interpret r^2 in context. The answer to part (d) is incorrect and does not address the concept of extrapolation.

Question 4

Overview

Question 4 involved a hypothesis test of the proportion of boxes of a breakfast cereal that contained a voucher. The hypotheses should have been stated using standard notation for a proportion (p or π) and with a lower tail alternative, since the students' claim was that the proportion of boxes with vouchers was less than 20 percent. Since the problem stated that the sample of boxes could be considered a random sample of the population, students needed to determine whether the sample size was adequate to allow a normal approximation by showing a computation to check that $np_o > 10$ and $n(1 - p_o) > 10$. They should have identified the one sample z-test for a proportion (or an acceptable alternative such as an exact binomial calculation) as the appropriate procedure to apply, and included a calculation to find the value of the test statistic and either a p-value or a critical value for a rejection region. Finally they should have properly interpreted the results in the context of the problem. The conclusion should have included justification linking the decision (do not reject the null hypothesis) to the p-value (or rejection region) by comparing to a specific significance level (e.g., $\alpha = 0.05$) or including a general comment such as "Since this p-value is so large … ." When interpreting the conclusion, students should not have indicated that they were "accepting" the null hypothesis or stated that the company's claim that $p = 0.2$ was correct. Rather, the conclusion should have indicated that the data did not provide sufficient evidence to refute the company's claim that 20 percent of the cereal boxes contained vouchers.

Sample: 4A
Score: 4

This response identifies an appropriate test, states the necessary conditions, and demonstrates how they were verified. Hypotheses are written using standard notation for population proportion with a lower tail alternative. Proper formula and correct substitution are used to calculate the correct test statistic and p-value. The conclusion is written with justification (linkage) and correct interpretation in the context of the problem.

Sample: 4B
Score: 3

This response shows hypotheses using standard notation for population proportion with a lower tail alternative. The response identifies an appropriate test, states the necessary conditions, and demonstrates how they were verified. Proper formula and correct substitution are used to calculate the correct test statistic and p-value. Although the conclusion is stated in context, it is missing linkage (no α stated or indication that the p-value is large) and erroneously concludes that the null hypothesis is correct. Either of these two errors alone would be enough to score part 4 as incorrect.

Sample: 4C
Score: 2

Hypotheses are written using standard notation for population proportion with a lower tail alternative. This response identifies an appropriate test; however, it fails to state or verify the necessary conditions. Proper formula and correct substitution are used to calculate the test statistic and p-value. The conclusion is written with justification (linkage) but lacks context and incorrectly writes that the null hypothesis is accepted.

Question 5

Overview

In this question, students were given information on an upcoming survey. The goal was to estimate the proportion of heads of households in the United States with (or without) a high school diploma. Random-digit dialing was to be used to select heads of households for inclusion in the sample. In part (a) of the question, students had to identify a potential source of bias for this survey, explain how that source of bias would be related to whether or not the head of household had a high school diploma, and describe the impact of the bias on the estimated proportion. Part (b) assessed whether students could determine the sample size that would be needed to obtain an estimate of the proportion with a desired level of precision. In part (c) students had to recognize that stratified random sampling should be used with states as strata, and random sampling should be done within each state. This process would yield both state and national estimates of the proportion of heads of households without a high school diploma.

Sample: 5A
Score: 4

The student correctly identifies that "Not all households in the US have telephones" in part (a). This source of selection (or undercoverage) bias is linked to the lack of a high school diploma. The potential effect of the bias is correctly stated as "this may result in an estimate that is too low for the proportion of adult heads of households in the US who do not have a high school diploma." In part (b) the student provides the correct formula and computation, rounding appropriately. The critical value used is 1.95996 instead of the more commonly used 1.96. The student's response correctly identifies "a stratified random sample" as the sampling method in part (c). The states are identified as the strata, and a correct random method is indicated. A mail survey is suggested to avoid the bias of random-digit dialing, and concern about potential "nonresponse" is expressed.

Sample: 5B
Score: 3

In part (a) a source of bias is identified as undercoverage and is correctly described as those who do not own a phone and linked to the lack of a high school diploma. The direction of undercoverage is correctly described to be because "the survey will leave out a large portion of households that would otherwise increase the estimate." The formula chosen and the numbers substituted are both correct in part (b). A minor arithmetic mistake yields 733.333, which is then appropriately rounded. The calculations are explained clearly. Part (c) has a nice description of taking a stratified random sample, using states as strata, but the sampling technique is never identified as stratified random sampling.

Sample: 5C
Score: 2

Part (a) has a nice description of the source of bias (undercoverage). The student explains that this would lead to people without a high school diploma being underrepresented in the sample. The effect of this bias would be to decrease the estimated proportion of heads of households without a high school diploma, not to increase it as stated by the student. The student gives the proper formula in part (b) and substitutes the appropriate values. A minor arithmetical error is made. However, the student gives the decimal value and then rounds properly. Using the term "block" instead of stratum in part (c), the student clearly describes taking a stratified random sample but fails to identify the sampling method as stratified random sampling.

Question 6

Overview

This question described an experiment in which a random sample of children from suburban day-care centers was randomly divided into a group that played outside and a group that played inside. The experiment was duplicated for children randomly selected from urban day-care centers. The response variable was the amount of lead on the child's dominant hand after an hour of play. A 95 percent two-sample *t*-confidence interval was given for the difference in the mean amount of lead for children playing inside and the mean amount of lead for children playing outside at the suburban day-care centers. In part (a) of the question, students were asked to construct the confidence interval for the two samples of children at the urban day-care centers. The response should have included identification of the procedure used, a check of the necessary conditions for the validity of the procedure, the computation of the interval, and an interpretation in context. Omitting a check of conditions or giving a list of "assumptions" with no work done to check them were common errors. Another common error was to compute a confidence interval for the mean difference as if the samples were paired. In part (b) students constructed a plot of the four means (inside suburban, outside suburban, inside urban, outside urban), which should have included a scale and labels. Part (c) assessed whether students could describe the effect of setting (mean lead level inside was lower than outside, for both suburban and urban children), environment (mean lead level for suburban children was lower than for urban children, both inside and outside), and the combined effect (relationship) of the two on the response (the mean lead level was not much different between inside and outside for suburban children, but the mean level was very different between inside and outside for urban children). Justification of the conclusions in part (c) could have included reference to the means, the plot in part (b), and, preferably, the confidence intervals given in the stem of the problem and constructed in part (a). Note that analysis of variance is not a topic on the AP Statistics syllabus, so this question was most students' first experience with describing main effects (inside/outside and suburban/urban comparisons) and interaction (the compounded relationship between environment and setting).

Sample: 6A
Score: 4

Each part of this outstanding response is complete and well organized. In part (a) the necessary conditions for a two-sample *t*-procedure (two independent random samples or random assignment of subjects to treatments and no reason to suspect that the populations are not approximately normal) are checked. The response is vague as to the reason that a boxplot should be roughly symmetric with no outliers (because then we may reasonably assume that the population from which the sample was drawn is approximately normal). The confidence interval is computed on the calculator, which uses fractional degrees of freedom. The interpretation is clear that the confidence interval is meant to capture the parameter of the difference in the population means rather than the mean of the differences. The three conclusions asked for in part (c) are given in order (setting, environment, relationship between them), with the justification for each conclusion included. The justification given for the conclusion about setting is excellent. The two confidence intervals do not overlap zero so, in both environments, there is a statistically significant difference between the mean lead levels of children who play inside and children who play outside. The response correctly states that there is not enough information to construct confidence intervals to justify the conclusion about the difference in environments. However, the response might also have justified the conclusion about the relationship using the confidence intervals: because the two confidence intervals do not overlap, and the interval for urban children is farther away from zero, the difference between playing inside and outside is larger for urban children than for suburban.

Question 6 (continued)

Sample: 6B
Score: 3

This substantial response does not consider conditions in part (a). Part (c) clearly describes all conclusions requested (setting, environment, and their relationship) but doing so in the order of setting, relationship, and environment. Each conclusion is justified, using the confidence intervals for setting and relationship and the means from the table of part (b) for environment.

Sample: 6C
Score: 2

In part (a) this developing response does not consider the crucial condition of normality for constructing a two-sample *t*-interval. The interpretation of the confidence interval is not correct. Unless requested, an interpretation of a confidence interval does not have to include an interpretation of confidence level, but if it is included, the statement must be correct. We expect the parameter will be captured in 95 percent of the different intervals generated by repeated random samples. We do not expect the parameter would be captured in this particular interval in 95 percent of repeated random samples. In part (c) the response clearly describes the effect of environment and setting, although it only gives a justification in terms of the means for environment. No conclusion about the relationship between the two is stated.