2017 AP STATISTICS: Question 1

The intent of this question was to assess:

1) **Students' statistical content knowledge via application of terminology**
2) **Determine if students could accurately interpret a slope in context**
3) **Evaluate if students could calculate an observed value for y given a residual value of y and a regression equation.**

To access the question, please visit https://secure-media.collegeboard.org/ap/pdf/ap-statistics-frq-2017.pdf. Be sure to read the College Board Solutions as well as this commentary.

a. **This question gave students information about a data set involving wolves' length in meters and weight in kilograms, and described the scatterplot. Specifically, students had to describe positive, linear and strong in context of the scenario.**
   A possible solution:
   - Positive: as length goes up, so does weight
   - Linear: the data points generally follow a straight line; or the rate of change of weight in response to length is a constant amount
   - Strong: residuals are small, or correlation coefficient is close to 1.

RUBRIC requires: context, positive, linear, strong

b. **Students were given the equation: $\hat{y} = -16.46 + 35.02\,x$; where x = length of wolves in meters and y = weight of wolves in kg. They were asked to interpret the slope in context.**
   A possible solution:
   - For each additional meter in length of the wolf, the weight increases, **on average,** 35.02 kg.

RUBRIC requires: correct value, interpretation, nondeterministic language

c. **The final part asked students to determine the weight of a wolf with length 1.4 meters and residual, -9.67 meters (using the equation in b).**
   A possible solution:
   - $\hat{y} = -16.46 + 35.02\,(1.4) = 32.568$
   - -9.67 = y -32.568
   - 22.898 = y

Rubric requires: correct value for predicted value and observed value as well as computation for at least one of the two.

COMMENTARY AND TEACHER TIPS

1. **Vocabulary terms need to be used correctly. Many students used two words incorrectly---correlation and normal. Blatant misuse of these words often caused students to miss what might have otherwise been correct. Instead of saying, "the relationship between x and y can be modeled by a line", they said, "the correlation between x and y" can be modeled by a line". Correlation is a term that represents the strength of linearity, so this was considered incorrect.**

2. **Students need to know definitions and how they apply. Students are learning how to determine strength, form and direction of a scatterplot, but it appears they don't really know what each term means. "Positive" was the easiest term for students to get, though many students still did not indicate that as one variable increased, the other also increased.**

3. **Clarity is important. Saying both length and weight go up does not indicate what a positive relationship in a scatterplot represents.**

4. **Linear could not be described with words like "modeled by a linear function". The points can be modeled best a least squares regression line was not acceptable because any set of data can be modeled by a LSRL (it does not mean it is a good fit or that it is useful for describing that set of data).**

5. **Also, saying "line of best fit is straight" is not acceptable. A line of best fit is always straight.**

6. **For linear, students could indicate that the data points generally made a line. However, stating that a line could be drawn through the points is not acceptable because a line could be drawn through any set of points.**

7. **When discussing shape and data, it is important for students to use correct terms. Saying "the scatterplot is a line" is incorrect. The data points in the scatterplot resemble a line is acceptable. They should clearly know and be able to write about the points that are displayed in the scatterplot, but not try to short cut what they are writing by only saying scatterplot.**

8. **A rate of change argument was acceptable if it was written well. The most important aspect of this argument was that students were required to communicate that a 1 unit change in one variable resulted in a constant change in the response variable. This communication had to be clear that the change was not 1 unit in x resulted in a 1 unit change in y. Ex: As the length of a wolf increases, the weight of the wolf changes by a constant amount.**

9. **Another fairly common mistake found in several parts was that "the slope was increasing". It wasn't. For a linear relationship, the slope is constant. Students need to practice writing and say what they mean and mean what they say.**

10. **Algebraic language was not accepted. Length and weight did not have a directly proportional relationship. If a student talked about the increase in length resulting in a constant change in the increase in weight, that was acceptable.**

11. **Strength was hard for many students to interpret. Indicating that the line had a high "r" value did not provide enough information-→ students had to indicate that r>.7, r was close to 1 and could even indicate this on the negative side. However, a value was needed for r to be considered as an argument.**

12. **Students must know how to accurately interpret slope. A large number of students described how to substituted the number into the equation and get the predicted value. That does not explain how to interpret slope.**

13. **Students need to realize they should not round slope. If they did so in the interpretation of slope section, they did not receive full credit because their non-deterministic language applied to their rounding, and no one used the non-deterministic language twice, as would be needed in that situation.**

14. **Trying to point out parts of the LSRL for the interpretation does not work. This is a sentence that should be in context.**

15. **It is imperative that students learn to use "predicted", "on average" or possibly "expected". I would suggest using only these words. On some occasions, other words may be acceptable, but that is rare. Students need to understand that the LSRL is a prediction equation and treat answers in that context.**

16. **Non- deterministic words that are acceptable are "on average, predicted, and expected". Because students were given an equation this year and asked to explain the slope, there was more leniency (don't expect this in the future) and students could use about and approximately. Why? No rounding was necessary in creating the equation or determining a final value. The reason why approximately and about do not typically work is rounding has taken place when getting the predicted value and the reader does not know if "approximately" is talking about the rounding or is talking about "on average" how the model predicts.**

17. **A major component is reappearing in the AP Exam and was present in many parts. Students must justify their work. We are fully aware that the work can be accomplished on the calculator, but that is not the point. We want to see what they are using to get the predicted value and what they are using to get the observed value. There were many students who did not receive full credit on "c" because they**

had bald answers. It is important to emphasize to students to show their work and to be consistent in your grading when they don't. They lost many points this year when work was not shown.