Common Errors on the 2009 AP Statistics Exam

Question 1

<u>Part (a)</u>
- Counts (frequencies) instead of percents (relative frequencies)—not appropriate when comparing groups of unequal size
- Nonstandard graphs of many varieties
- No label or incorrect label on vertical axis
- Condition based on one variable, but draw the graph as if they conditioned on the other

<u>Part (b)</u>
- Describing shape, center, or spread
- Referring to correlation instead of association
- Only commenting on who has more jobs
- Not addressing all three categories clearly

<u>Part (c)</u>
- Incorrect name for chi-square test
- Stating hypotheses that suggest causation, e.g.
  Ho: Gender has no effect on job experience.
- Using symbols in hypotheses, e.g.
  Ho: $\chi^2 = 0$
- "Freelancing" on wording of hypotheses, e.g.
  Ho: There is no significant association…

Question 2

Part (a)

- A number of students set z=0.7 and attempt to calculate the stopping distance. Many had difficulty determining the percentile.
- Several students interpreted the 70%-tile to be 70% centered about the mean. By doing so several used the 68-95-99% rule to try to solve the problem.
- Used "calculator speak" to calculate stopping distance without defining parameters of the distribution. Calculator commands without defining the arguments for the commands are discouraged.
- Sketches of the approximately normal distribution were unclear and unlabeled.

Part (b)

- Several students failed to use the binomial distribution correctly; calculating $1-P(Y\leq 2)$ instead of $1-P(Y\leq 1)$.
- Several students only gave one term, $P(Y=2)$.
- Several students used .7 as their probability of a success, as well as correctly defining the parameters of the binomial distribution
- Several students constructed another normal probability not seeing the binomial was needed.

Part (c)

- Several students failed to get the sampling distribution-didn't define the distribution or its parameters correctly.
- Several students gave a value z=1.72, and a value p=0.0427, without correctly indicating that this probability relates to $P(Z\geq 1.72)=1-P(Z<1.72)=0.0427$.
- Several students thought confused the question with a test of hypothesis and gave the probability $P(Z\geq 1.72)=1-P(Z<1.72)=0.0427$ as a p-value.

Question 3

Part (a)

- Some used a stopping rule with a coin toss (or equivalent method) with out prior randomization.
- Some did not state a device/mechanism for randomization.
- Some assigned numbers to students without randomization; then used evens/odds or 1-12 and 13-24 to select the members of the 2 groups.
- Many did not specify groups (context); i.e., they formed “Group 1” and Group 2,” but did not indicate which was the dissection group and which was the computer software group.
- Some did not give enough information so that two “knowledgeable statistics users” would use the same method to assign the students to the two instructional groups.
- Some made references to simple random samples.
- Many gave the design diagram alone.
- Many who picked names or numbers from a hat, bag, or laundry basket forgot to mix the contents.
- Some who used a paired design did not block on similar pre-test scores.
- Some were not explicit on how they put students in random order (“randomly assign”).
- Some over-randomized and shoot themselves in the foot (e.g. proper randomization into 2 groups then using a poorly-described coin flip (or some such) to assign treatments to groups).
- Some had inappropriate or poorly described blocking schemes.
- Some formed blocks on a characteristic other than pre-test such as gender.

Part (b)

- Most stated a reasonable characteristic (not a mistake); some, however, only say that students “like it.”
- Most did not describe how behaviors associated with the self-selection criterion impact the changes in the differences (post – pre).
- Many referred only to the post-test (instead of the change in score {post – pre}) or mentioned a vague aspect of performance (do better, learn more/less).
- Many required the reader to infer that the student knows an effect on post-test score will affect the response variable (change).
- Many used terms like bias, observation, voluntary response, and so on unclearly.
- Some students only mentioned a characteristic without any connection to performance.

Question 4

In part (a), step 1:

- Many students identified a z confidence interval as the appropriate procedure rather than a t.
- Many students failed to check the sample size condition at all.
- Some students did an inadequate job of checking the sample size conditions by saying that the samples are large enough, with no reference to a number such as 25 or 30, the central limit theorem or sampling distributions.
- Some students stated that 50 is large enough to assume that the *populations* or *samples* or *data* are approximately normal, rather than that the sampling distribution(s) is (are) approximately normal.

In part (a), step 2:

- A few students used 1.645 as the multiplier in their computation.
- A few students neglected to square the standard deviations when computing the standard error, and consequently presented an incorrect final answer.
- A few students thought that the interval could not go below 0, and truncated it at 0.

In part (a), step 3:

- Some students omitted the word "mean" and interpreted the confidence interval as applying to the difference in response times.
- Some students omitted the word "difference" or any similar wording to indicate that the interval is for a difference in means, stating that the interval is for the "mean response time."
- A few students omitted the context.
- A few students interpreted the confidence level instead of the confidence interval.
- A few students interpreted the confidence interval correctly but interpreted the confidence level as well and did it incorrectly.
- A few students wrote that the confidence interval was for a "mean proportion" or "proportion of difference" or similar wording using "proportion."

In part (b):

- Many students made a statistically incorrect statement, such as "because the interval contains 0, the council member's belief is wrong."
- Some students thought that the interval supported the council member's belief because it included more values on one side of 0 than the other.
- Some students thought that the interval supported the council member's belief because it included values as large as 2 minutes.
- A few students based a conclusion solely on testing hypotheses and made no reference to the confidence interval.

Overall mistakes:

- Some students computed two separate confidence intervals rather than a confidence interval for the difference.
- Some students confused notation for sample and population means.
- Some students presented a formula for the confidence interval with incorrect numbers plugged in, but then used a calculator to compute the correct interval, and used the calculator version to answer the remainder of the question.
- A few students referred to the sample standard deviations of 3.7 and 3.2 as σ_1 and σ_2 or called them population standard deviations.

Question 5

Part (a)

- Confusing the p-value with the significance level (saying that the p-value is the probability of rejecting H0)
- Interpreting the p-value as the probability that H0 (or Ha) is true (or false)
- Omitting a reference to the difference between proportions obtained in this study: "There is a 7.61% chance that the treatment that uses CC alone produces a higher survival rate than CC+MMR, if the true difference between the survival rates is 0"
- Omitting "as large as" in the probability phrase
- Saying "by chance alone" or "as a result of sampling variation" in place of the conditional phrase
- Omitting the conditional phrase
- Omitting context

Part (b)

- Stating that we "accept H0"
- Omitting linkage
- Omitting context

Part (c)

- Confusing Type I and Type II errors
- Providing a "consequence" that is a decision, rather than an action
- Lacking specificity with respect to treatments – specifically, failing to distinguish whether "both treatments" means "CC+MMR and CC" or "CC+MMR"

Question 6

Part (a)

- Students did not understand how to define the parameter of interest. Instead we saw:
    a. "The mpg of the cars" (the variable of interest)
    b. "All the cars of this model" (the population of interest)
    c. "To determine if the manufacturer is misleading customers" (the question of interest)
- Students often attempted to define the parameter more than once (e.g. saying "the parameter is …" and then later saying "μ = …"). These were treated as parallel solutions and the worst attempt was scored.
- Students used non-standard notation in the hypotheses, often without explicitly defining their notation.
- Students used a two-tailed alternative hypothesis.

Part (b)

- Students reversed the relationship between the mean and median in a right skewed distribution (saying the mean will be less than the median in a right skewed distribution).
- Students made good statements about the relationship between the mean and median but did not state that large values of the statistic indicate right skewness.
- Students stated that large values of the statistic indicate right skewness but only argued that in a normal (or symmetric) distribution the ratio should be close to 1 and did not explain how the mean and median are related in a right skewed distribution.
- Students said "large" without any explanation.

Part (c)

- Students did not understand that the dotplot approximated the *sampling distribution* of the statistic (sample mean)/(sample median). In other words, students did not understand that it showed what values of the statistic would occur when sampling from a *normal* population.
    a. Some students thought the dotplot showed sample data (as opposed to simulated values of a sample statistic), described the shape of the dotplot as approximately normal, and used this to justify that the original sample came from a normal population.
    b. Other students thought that the values in the dotplot came from new samples of size 10 from the *original* population, instead of from a normal population.
- Students did not understand how to use the dotplot to make an appropriate conclusion. In other words, students did not know to look for where 1.03, the observed value of the sample statistic, fell in the distribution and explicitly indicate whether or not a value of 1.03 would be likely to occur by chance when sampling from a normal population.
    a. Some students stated the relative position of 1.03 without specific numerical evidence from the dotplot ("1.03 is towards the middle of the distribution") and then correctly decided that it was plausible the original sample came from a normal population.

b. Some students thought that 1.03 was unusual enough to conclude that the original population was skewed to the right ("1.03 is in the tail of the distribution so I conclude that the sample came from a right skewed population").

- Other incorrect responses included:
    a. Saying that the dots are centered around 1 so the sample came from a normal population. Because the sampling distribution was generated using samples from a normal population, this is not surprising. However, it doesn't address whether or not 1.03 is unusual.
    b. Arguing that 1.03 is close to 1 without describing its relative position in the dotplot. It was clear that many students were thinking simply about the absolute difference between 1 and 1.03 without considering the variability in the sampling distribution.
    c. Stating that the sample size (or number of samples) is large, so the distribution is normal or that the sample size is too small to make a conclusion
    d. Stating that the sample came from a normal population with no explanation.

<u>Part (d)</u>

- Students did not provide a statistic that measured skewness. This typically occurred if the student only focused on the right half of the distribution (e.g. max/Q3) or used a measure of spread (e.g. (max – min)/median).
- Students provided a method for identifying skewness but not a statistic. For example, "if (max – med) > (med – min), then the distribution is skewed right."
- Students did not correctly identify the values of their statistic which indicate skewness to the right (e.g. looking for values < 1 when using (med – min)/(max – med)).
- Students did not justify the values that indicate right skewness by discussing how right skewness affects the relationship between the components of the statistic.
- Students often tried to use outlier rules to measure skewness. Using these rules correctly and concluding that there is right skewness if there are outliers on the right but not on the left got credit for a reasonable method, but not a reasonable statistic.