AP Stats Pre-Work

Part 1: Algebra 1/Algebra 2 Review

Show work for ALL problems on a separate sheet. Give answers on this page. Simplify all answers. **Do not approximate or round.**

Solve for the variable.

1) $4(x-3) = 4^2 - x$

$4x - 12 = 16 - x$
$+x \quad +x$
$5x - 12 = 16$
$+12 \quad +12$
$5x = 28$
$x = 28/5$ or $x = 5\ 3/5$

2) $\frac{1}{3}n + 5 = n - 2$

$-n \quad -n$
$-\frac{2}{3}n + 5 = -2$
$-5 \quad -5$
$-\frac{2}{3}n = -7$
$\cdot -\frac{3}{2} \quad \cdot -\frac{3}{2}$
$n = 21/2$ or $n = 10\ 1/2$

3) $\frac{3}{7}x = \frac{9}{35}$

$\cdot\frac{7}{3} \quad \cdot\frac{7}{3}$
$x = \frac{3}{5}$

4) $4(3x-1) > 3x - 1$

$12x - 4 > 3x - 1$
$-3x \quad -3x$
$9x - 4 > -1$
$+4 \quad +4$
$\frac{9x}{9} > \frac{3}{9}$
$x > 1/3$

5) $(x-8)3 \leq 5x + 2$

$3x - 24 \leq 5x + 2$
$-5x \quad -5x$
$-2x - 24 \leq 2$
$+24 \quad +24$
$\frac{-2x}{-2} \leq \frac{26}{-2}$
$x \geq -13$

6) $\frac{m}{12} + \frac{5}{6} = \frac{5}{24}$

$-\frac{5}{6} \quad -\frac{5}{6}\left(\frac{20}{24}\right)$
$\frac{m}{12} = -\frac{15}{24}$
$\cdot 12 \quad \cdot 12$
$m = -\frac{15}{2}$ or $m = -7\ 1/2$

7) $\frac{1}{2}x^2 - 8 = 0$

$\frac{1}{2}x^2 = 8$
$x^2 = 16$
$x = \pm 4$

8) $x^2 - 8x - 20 = 0$

$(x-10)(x+2) = 0$
$x - 10 = 0 \quad x + 2 = 0$
$+10 \quad +10 \quad -2 \quad -2$
$x = 10$ or $x = -2$

9) $\sqrt{2x+10}^2 = x + 1^2$

$(x+1)^2$
$2x + 10 = x^2 + 2x + 1$
$-2x \quad -2x$
$10 = x^2 + 1$
$-1 \quad -1$
$\sqrt{9} = \sqrt{x^2}$
$x = \pm 3$ $\quad x = 3$

$\sqrt{2(3)+10} = 3+1$
$\sqrt{16} = 4$
$4 = 4$

$\sqrt{2(-3)+10} = -3+1$
$\sqrt{-6+10} = -2$
$\sqrt{4} = -2$
$2 \neq -2$

Section 4.1 (p. 228), #E5-9, E11

E5. From respondents 40 years old or older you would expect to get an estimate that is too high. Forty year olds are older than average. As people get older, they tend to visit more and more states. (They can't visit fewer!) From the residents of Rhode Island you might get an estimate that is too high as well, because Rhode Island is a small state close to many other small states. Compare this result to what you might expect from asking people who live in Texas or Montana.

E6. Contacting all of the tens of millions of adult males in the United States would be impossible. Questionnaire bias is unlikely for a factual question like this. Assuming that a nonbiased sampling method is used, there still might be an incorrect response bias because some people might be reluctant to admit to smoking.

E7. a. This is a voluntary response sample.
b. No. With a volunteer response sample, people with stronger feelings one way or the other are more like to respond.
c. Quite a bit less than 92%: The percentage is almost surely inflated by voluntary response bias.

E8. a. This is a volunteer sample.
b. No. With a volunteer sample, people with stronger feelings one way or the other are more likely to respond. Consumer Union is well aware of this, as evidenced by their disclaimer: "This poll is not scientific. It reflects the opinions of only those Internet users who have chosen to participate. The results cannot be assumed to represent the public as a whole. Consumer Union is not responsible for content, functionality, or the opinions expressed therein."
c. It is difficult to say whether people who drive less would be overrepresented or underrepresented in the voluntary sample.

E9. The estimate will be too high for two reasons. First, families with no children have no chance of being sampled. Second, families with many children will be overrepresented. For instance, families with 5 children will be 5 times more likely to be represented than families with one child.

E11. The ABC News poll. From the Gallup poll, 36% were in favor of US air strikes. From the ABC poll, 65% were in favor. Note that the ABC News poll mentions the involvement of allies of the United States and also makes it sound as though the air strikes will hit only objects and not soldiers. It can be hard to predict in advance which direction the psychology of questionnaire bias will go.

Section 4.2 (p. 242), #P7-10, E16, E18-20

P7. a. No. The students at the end of the list have no chance of being chosen.

b. Each student has the same chance of being chosen, and if you think of the phone numbers as being assigned randomly before class, all possible groups of students have the same chance of being in the sample. However, an implicit assumption in simple random sampling is that the sample size is fixed in advance, and in this situation the sample size would be random. So, although this produces a random selection of students, it does not produce a simple random sample of a fixed sample size.

c. No. Although each student has the same chance of being chosen, not all possible groups of students have the same chance of being chosen. Two students sitting in different rows cannot both be in the sample.

d. Yes.

e. No. Although each student has the same chance of being chosen, not all possible groups of students have the same chance of being chosen. A group of six girls cannot all be in the sample.

f. No. Although each student has the same chance of being chosen, not all possible groups of students have the same chance of being chosen. Two students with last names starting with different letters cannot both be in the sample.

P8. a. $0.43 \cdot 1200 = 516$ people who see their dentist twice a year. $0.32 \cdot 1200 = 384$ people who see their dentist once a year. $0.25 \cdot 1200 = 300$ people who see their dentist less than once a year. This could be difficult to do in practice, however. How would you get a list of people who visit their dentist less than once a year?

b. Since you want strata to be as different from each other as possible, you would want to stratify on age. We know there is a difference between the 40 or older and under 40 groups. We know there is little difference between male and females, and we don't really know the effect of apples on gum recession.

P9. $0.65\left(\frac{84}{100}\right) + 0.35\left(\frac{69}{100}\right) = 0.7875$

P10. Take 5% of 200, which is 10 people, and then calculate k: $\frac{200}{10}$ or 20. Then choose a random start between persons 1 and 20, and take every 20^{th} person thereafter. Take 20% of 200, which is 40 people, and find k: $\frac{200}{40}$ or 5. Then choose a random start between 1 and 5, and take every 5^{th} person thereafter.

E16.a. All the cookies must be numbered, which means buying all the bags in the store, numbering all the cookies, and choosing the SRS using random digits.

b. Use bags as clusters. Number the bags and choose an SRS of bags. Use all the cookies in those bags as the sample. A cluster sample is much more practical.

c. As in part b, number the bags and choose an SRS of bags. Then, take an SRS of cookies from each bag in the cluster sample. Which method you prefer depends on the patterns of variation and on how much time and money you have. If you have lots of time and not much money, use a cluster sample: Buy a small number of bags and count the chips in every cookie. If money is no object, but you do not have much time, you could use the two-stage sample: Buy more bags and count fewer cookies from each bag. The two-stage sample would give better estimates if chip contents vary a lot from bag to bag, but tend to be comparatively uniform from cookie to cookie within each bag.

E18. To get 30 pages from 600, $k = \frac{600}{30} = 20$. So, take every 20th page, starting at a page chosen at random from those numbered 1 through 20.

E19. Because there are only five farms and a large number of acres per farm, choosing a sample of farms (cluster sampling) does not seem reasonable. Instead, consider the farm as the five strata and take a random sample of, say, 10 acres from each farm.

E20. Stratification by age and by gender may be appropriate as attitudes toward movies probably differ by age and gender of the patron. Using adults, teens, and children as proxies for age and male and female for gender, there would be six strata. Then take a SRS of 16 each of the adult males and females, 7 or 8 each of the teenage males and females, and 1 or 2 each of the boy children and girl children.

Section 4.3 (p. 257), #P16, P18-20, P24, E28-31

P16.a. The lurking variable is the person's age. Older people are more likely to do both.

b. I causes II. Experiments have shown that drinking more milk, which is high in calcium, results in stronger bones. Of course, there are other causes of strong bones, such as heredity and exercise.

c. II causes I. People who go to school longer tend to earn more money. However, this statement is based on observational studies, not on randomized experiments, so occasionally the cause-and-effect relationship is questioned. Some people believe that family background is a lurking variable: People from well-off families tend to go to school longer and to earn more money, but their family's support in getting them established caused the higher income, rather than the schooling. However, most people believe that although family background certainly contributes, the number of years of schooling is another cause of higher income.

P18.a. The factors are type of lighting (brightness of the room) and type of music. For the type of lighting, the levels are low, medium, and high. For the type of music, the levels are pop, classical, and jazz.

b. Possibilities are heart rate, blood pressure, or self-description of anxiety level.

P19.a. You should not believe this. The pipe and cigar smokers are a bit older on average.

b. observational study

c. The factor is smoking behavior. The levels are nonsmoking, cigarette smoking, and pipe or cigar smoking. The response variable is the number of deaths per 1000 men per year.

P20. Older men have a higher death rate; the pipe and cigar smokers are older than the others. Here are the age adjusted death rates per 1000 men: nonsmokers 20.3, cigarette smokers 28.3, and pipe and cigar smokers 21.2. The new factor is the age.

P24.a. This study has quite a few problems. Students might be pointing home not from some magnetic sense, but because they could feel how much "turning" they had done or because they knew the roads. The might also be able to tell direction because they could sense the direction of the sun even though they were blindfolded.

b. the group that has the magnets

c. Whether the students were assigned randomly to the treatments. Also, the study was not double-blind.

d. The second design is certainly better than the first.

e. The subject could tell whether they had the magnet because of the magnet's attraction to the walls of the van, so this study was not blind. It was not double-blind either because the experimenter who evaluated how well the students pointed home was the same person who chose which kind of bar they wore.

E28.a. the dishes of insects.

b. 8

c. experiment

E29. Death *rate* is used instead of total number of deaths because the size of the populations are different. Death rate may not be a good choice because climate is confounded with proportion of older people in the state. Florida has a much higher proportion of older people than does Alaska because a great number of people move to Florida after retiring. This is an observational study.

E30.a. Graduation rate could be the proportion of enrolled seniors who graduate from high school, or the proportion of the freshman class 4 years prior that graduate within 4 years. (Which of these is likely to produce higher graduation rate?) It is important to look at rates rather than number of graduates because the high schools are likely to be different sizes.

b. The main issue is the question you are trying to answer. If you want to simply see what graduation rates are at different schools, you would do a census or a sample survey. Cost or time would be the main issue to determine which of these to do. If you want to determine something about the cause of graduation rates, you would want to do an experiment or observational study. You could do an experiment if there is some way to randomly assign treatments to the students. If there is not, you would need to do an observational study.

c. No, because you are simply looking at data from the schools and not randomly assigning treatments, you cannot determine cause.

E31. No, not unless the subjects are in random order to begin with. Suppose the subjects in poorest health are all at the end of the list. Then, all of those in poorest health probably are going to be assigned to the same treatment group, so *state of health* and *treatment* will be confounded.

Section 4.4 (p. 275), #P29-P31, E35, E38, E40

P29.a. randomized paired comparison with repeated measures

b. randomized paired comparison with matched pairs

c. randomized paired comparison with matched pairs because they may be a residual effect of one or both of the drugs – that is, the drug does not dear out of the bloodstream in the time allowed between treatments.

P30. Randomized paired comparison with repeated measures

P31.a. Students' memorization skills vary greatly, so some sort of blocking seems desirable. One way is to use a randomized paired comparison design with repeated measures. First, find the seniors in the school who are willing to participate. Then randomly select half to go into a room that has the radio playing and the other half to go into a room that is quiet. The students would be given a familiar task to study, such as a list of new vocabulary words. Afterward, they would be tested on the meaning of the words. The number that they remember is the response variable. Then the students would switch rooms and be given another new list of vocabulary words.

b. For this experiment, the treatments are soup with MSG and soup without MSG. The response variable is the amount of soup eaten by a customer who orders soup. Sources of within treatment variability include the customer's weight (bigger customers tend to eat more) and the temperature of the room during the evening (if it is cooler, customers may eat more soup). Consequently, there is a lot of variation in how much soup people will eat, no matter which treatment they receive. Because it is difficult to estimate a person's size, block could be formed of adults and children, or men, women, and children. For each customer who orders soup that night, a treatment is randomly selected, perhaps by flipping a coin each time.

E35. This is a randomized paired comparison (repeated measures) design. Block: patient (more precisely, two weeks of a patient's time). Treatments: low phenylalanine diet or regular diet. Experimental unit: one week of a patient's time. Design: this is a randomized paired comparison design with repeated measures. Response: dopamine level

E38. The treatments were assigned completely at random in this completely randomized design. Response: time (months) until baby walks unaided. Unit: infant Treatments: four kinds of exercise/follow-up program. Blocks: none

E40.a. This is a completely randomized design. Since there are no obvious variables on which to block, a completely randomized design seems reasonable.

b. The experimental units are the individual rats. The response measurement is the total length of stomach lesions.

c. Yes. While the rats are unlikely to understand the treatment, being forced to take medication (however it is administered) could be upsetting. The added stress could affect the stomach lesions.

Chapter 4 Review (p. 279), #E41, E44, E51, E56, E59

E41.a. Sample. The measurement process is destructive, and the size of the population makes a census too expensive.

b. Census. The population size is small, the information is easy to get, and the measurement process is not destructive.

c. Sample. The population is so large that a census would be too costly and time-consuming.

E44.a. Decrease, because take-home pay is less than gross earnings.

b. The worker might overstate his or her earnings, misclassifying his or her occupation/title, or include overtime, for example.

E51. *New York Times* readers tend to have higher incomes and more years of education than the average New Yorker.

E56.a. The mother has done an observational study. She has not assigned the two treatments of *cola* and *no cola* to her daughter at random at different times.

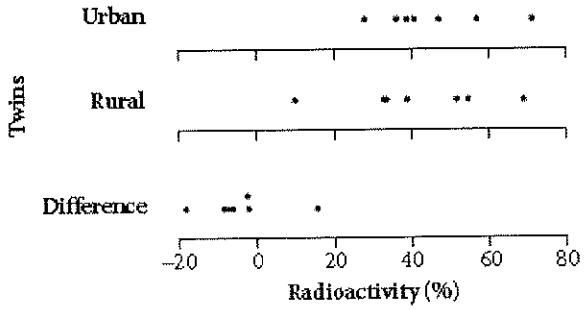
b. The cola is confounded with the fast food and birthday party food. It also is confounded with the fact that the daughter is eating out rather than eating at home, which she may be excited about. Another variable confounded with the cola is that the daughter knows her mother doesn't approve of cola, which may change the daughter's behavior. Thus, instead of the cola causing the hyperactivity, it may be caused by something in the food, or by the excitement of being away from home, or by the daughter wanting the mother to notice she is doing something "rebellious."

c. The daughter could be given cola or not at random over a sufficient number of afternoons and her behavior evaluated. The person who evaluates the level of activity must be blinded to whether the daughter had been given cola or not. However, the daughter would know whether she had had cola or not, and she already understand how she is expected to behave when she has cola. If the trigger for any possible hyperactivity could be isolated as, say, the sugar or caffeine in the cola, it might be possible to remove those from the cola so that the daughter couldn't tell whether she was receiving regular cola or the placebo cola.

E59.a. The factor is location, with level being urban or rural. A block is one pair of twins.

b. This is an observational study. The treatment is already attached to the units.

c. The dot plots show the variation in response is greater in rural twins and least in the differences. This study does demonstrate that the lungs of the urban twins took longer to clear than did the lungs of the rural twins.



d. Twins were used because they are genetically identical. This will reduce any variability caused by genetic differences.

Part 3: AP Problems

1) In response to nutrition concerns raised last year about food served in school cafeterias, the Smallville School District entered into a one-year contract with the Healthy Alternative Meals (HAM) company. Under this contract, the company plans and prepares meals of 2,500 elementary, middle, and high school students, with a focus on good nutrition. The school administration would like to survey the students in the district to estimate the proportion of students who are satisfied with the food under this contract.

Two sampling plans for selecting the students to be surveyed are under consideration by the administration. One plan is to take a simple random sample of students in the district and then survey those students. The other plan is to take a stratified random sample of students in the district and then survey those students.

a. Describe a simple random sampling procedure that the administrators could use to select 200 students from the 2,500 students in the district.

The administrators could number an alphabetical list of students from 1 to 2500. They could then use a random number generator from a calculator or computer to generate 200 unique random integers from 1 to 2500. The students corresponding to those 200 numbers would be asked to participate in the survey.

b. If a stratified random sampling procedure is used, give one example of an effective variable on which to stratify in this survey. Explain your reasoning.

One possible stratification variable might be the school level of the student (elementary, middle, high school). The students' perception of the importance of good nutrition in food served may differ depending on the students' age and therefore on school levels. For example, there may be a difference between what elementary students value in food served as opposed to middle school and high school students.

c. Describe one statistical advantage of using a stratified random sample over a simple random sample in the context of this study.

One statistical advantage of using stratified random sampling as opposed to simple random sampling is, for example, if the elementary, middle and high school strata create groups that differ with respect to what they value – and are therefore more homogeneous with respect to opinion on this issue – then for the same overall sample size a more accurate estimate of the overall proportion of students who are satisfied with the food under this contract may result. Another advantage is that stratified random sampling guarantees that each of the school-level strata will have some representation, because it is possible that a simple random sample would miss one or more of the strata completely.

2) At a certain university, students who live in dormitories eat at a common dining hall. Recently, some students have been complaining about the quality of the food served there. The dining hall manager decided to do a survey to estimate the proportion of students living in the dormitories who think that the quality of the food should be improved. One evening, the manager asked the first 100 students entering the dining hall to answer the following question.

> Many students believe that the food served in the dining hall needs improvement. Do you think that the quality of food served here needs improvement, even though that would increase the cost of the meal plan?
>
> _____Yes _____No _____No Opinion

a. In this setting, explain how bias may have been introduced based on the way this convenience sample was selected and suggest how the sample could have been selected differently to avoid that bias.

Since the manager used a convenience sample (the first 100 students entering the cafeteria), bias may have been introduced. Students who arrive at the cafeteria early may have opinions of food quality that differ in some important way from other students who live in the dormitories.

The bias could be avoided by selecting a random sample of 100 dormitory residents instead of just asking the first 100 students entering the cafeteria.

b. In this setting, explain how bias may have been introduced based on the way the question was worded and suggest how it could have been worded differently to avoid that bias.

The way the question is worded may be leading. The first part of the question included a statement that many students think the food needs improvement. This may lead people to support this view by responding that the food does need improvement.

The inclusion of the phrase "even though that would increase the cost of the meal plan" would lead students to say the food is OK only because they do not want to pay more.

A better wording might be to simply ask "Do you think that the quality of the food served in the cafeteria needs improvement?"

Scoring Guidelines (AP Problems)

#1

Parts (a), (b) and (c) are each scored as essentially correct (E), partially correct (P) or incorrect (I).

Part (a) is scored as follows:

Essentially correct (E) if the response describes a sampling procedure for generating a simple random sample and the description of the sampling procedure includes sufficient detail for implementation.

Partially correct (P) if random selection is used correctly for generating a simple random sample, but the description of the sampling procedure does not provide sufficient detail for implementation.

Incorrect (I) if random selection is not used in a correct way for a simple random sample.

Note: A response in which objects are placed into a hat or a box and then drawn out can only earn an "E" if the response explicitly states that the objects are mixed or that they are drawn out at random.

Part (b) is scored as follows:

Essentially correct (E) if the response identifies a reasonable stratification variable and provides a reasonable justification in context (such as stating, "the groups (strata) might differ *with respect to food preferences or nutritional awareness*").

Partially correct (P) if the response identifies a reasonable stratification variable but provides a weak justification (such as stating only, "the groups (strata) differ").

Incorrect (I) if the response identifies an unreasonable stratification variable, or provides an unreasonable justification or no justification.

Part (c) is scored as follows:

Essentially correct (E) if the response provides a reasonable statistical advantage of stratified random sampling that is not also true of random sampling, and that is clearly communicated and in context.

Partially correct (P) if the response provides a reasonable statistical advantage that is either not well communicated or that is not in context.

Incorrect (I) if the response includes only a vague potential statistical advantage, such as "data more accurate" or "stratified random sampling is better."

Note: Responses to part (c) such as "stratified random sampling allows for inferences to be drawn for the three grade levels separately about the feelings of students in those grade levels" should be considered incorrect unless also accompanied by a statistical advantage specific to stratified random sampling.

4 **Complete Response**

All three parts essentially correct

3 **Substantial Response**

Two parts essentially correct and one part partially correct

2 **Developing Response**

Two parts essentially correct and one part incorrect

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 **Minimal Response**

One part essentially correct and two parts incorrect

OR

Two parts partially correct and one part incorrect

#2

Each part is scored as either essentially correct, partially correct, or incorrect.

Part (a) is essentially correct if the response

1. indicates that selecting the first 100 students to arrive at the cafeteria could introduce bias because opinions of the first 100 students might differ from the opinions of other students who live in the dormitories
2. proposes a reasonable alternative that involves random selection—a simple random sample of dorm residents, or some sort of stratified random sample, or a systematic sample with a random starting point.

Part (a) is partially correct if the response

indicates why selecting the first 100 students to arrive is not reasonable, but proposes an alternative that does not involve random selection

OR

proposes a reasonable alternative that involves random selection, but does not explain how selecting the first 100 students could introduce bias.

Part (b) is essentially correct if the response

1. points out <u>at least one</u> of the two possible problems with the question wording
2. proposes reasonable alternate wording that addresses the concern(s) raised.

Note: The student only needs to identify one problem and take care of it. If only one of the two wording concerns is raised, the alternate wording need only address the one wording problem. However, if the student identifies both problems, appropriate alternate wording must be provided for both problems.

Part (b) is partially correct if the response

identifies one or both of the potential wording problems, but does not propose a new wording that adequately addresses an identified problem.

Part (b) is incorrect if the response

points out both wording problems, but then argues that the question wording is OK as is because the biases are in opposing directions and so will balance each other.

4 Complete Response

Both parts essentially correct

3 Substantial Response

One part essentially correct and the other part partially correct

2 Developing Response

One part essentially correct and the other part incorrect
OR
Both parts partially correct

1 Minimal Response

One part partially correct