AP Statistics: Ch 5 and 6 Assignment on Experiments and Simulations:

1. 2003 AP FRQ:

Because of concerns about employee stress, a large company is conducting a study to compare two programs (tai chi or yoga) that may help employees reduce their stress levels. Tai chi is a 1,200-year-old practice, originating in China, that consists of slow, fluid movements. Yoga is a practice, originating in India, that consists of breathing exercises and movements designed to stretch and relax muscles. The company has assembled a group of volunteer employees to participate in the study during the first half of their lunch hour each day for a 10-week period. Each volunteer will be assigned at random to one of the two programs. Volunteers will have their stress levels measured just before beginning the program and 10 weeks later at the completion of it.

(a) A group of volunteers who work together ask to be assigned to the same program so that they can participate in that program together. Give an example of a problem that might arise if this is permitted. Explain to this volunteer group why random assignment to the two programs will address this problem.

(b) Someone proposes that a control group be included in the design as well. The stress level would be measured for each volunteer assigned to the control group at the start of the study and again 10 weeks later. What additional information, if any, would this provide about the effectiveness of the two programs?

(c) Is it reasonable to generalize the findings of this study to all employees of this company? Explain.

'2. 2006 AP Frq

A biologist is interested in studying the effect of growth-enhancing nutrients and different salinity (salt) levels in water on the growth of shrimps. The biologist has ordered a large shipment of young tiger shrimps from a supply house for use in the study. The experiment is to be conducted in a laboratory where 10 tiger shrimps are placed randomly into each of 12 similar tanks in a controlled environment. The biologist is planning to use 3 different growth-enhancing nutrients (A, B, and C) and two different salinity levels (low and high).

(a) List the treatments that the biologist plans to use in this experiment.

(b) Using the treatments listed in part (a), describe a completely randomized design that will allow the biologist to compare the shrimps' growth after 3 weeks.

(c) Give one statistical advantage to having only tiger shrimps in the experiment. Explain why this is an advantage.

(d) Give one statistical disadvantage to having only tiger shrimps in the experiment. Explain why this is a disadvantage.

3. 2009 AP Frq

Before beginning a unit on frog anatomy, a seventh-grade biology teacher gives each of the 24 students in the class a pretest to assess their knowledge of frog anatomy. The teacher wants to compare the effectiveness of an instructional program in which students physically dissect frogs with the effectiveness of a different program in which students use computer software that only simulates the dissection of a frog. After completing one of the two programs, students will be given a posttest to assess their knowledge of frog anatomy. The teacher will then analyze the changes in the test scores (score on posttest minus score on pretest).

(a) Describe a method for assigning the 24 students to two groups of equal size that allows for a statistically valid comparison of the two instructional programs.

(b) Suppose the teacher decided to allow the students in the class to select which instructional program on frog anatomy (physical dissection or computer simulation) they prefer to take, and 11 students choose actual dissection and 13 students choose computer simulation. How might that self-selection process jeopardize a statistically valid comparison of the changes in the test scores (score on posttest minus score on pretest) for the two instructional programs? Provide a specific example to support your answer.

4. 2011 AP FrQ:

An apartment building has nine floors and each floor has four apartments. The building owner wants to install new carpeting in eight apartments to see how well it wears before she decides whether to replace the carpet in the entire building.

The figure below shows the floors of apartments in the building with their apartment numbers. Only the nine apartments indicated with an asterisk (*) have children in the apartment.

11* 12 1st Floor 14 13	21 22* 2nd Floor 24 23*	31 32 3rd Floor 34 33	
41 42 4th Floor 44 43	51* 52 5th Floor 54 53	61 62 6th Floor 64 63	* = Children in the apartment
71 72 7th Floor 74* 73*	81 82 8th Floor 84* 83	91 92* 9th Floor 94 93*	

(a) For convenience, the apartment building owner wants to use a cluster sampling method, in which the floors are clusters, to select the eight apartments. Describe a process for randomly selecting eight different apartments using this method.

(b) An alternative sampling method would be to select a stratified random sample of eight apartments, where the strata are apartments with children and apartments with no children. A stratified random sample of size eight might include two randomly selected apartments with children and six randomly selected apartments with no children. In the context of this situation, give one statistical advantage of selecting such a stratified sample as opposed to a cluster sample of eight apartments using the floors as clusters.

5. 2013 AP Frq

2. An administrator at a large university wants to conduct a survey to estimate the proportion of students who are satisfied with the appearance of the university buildings and grounds. The administrator is considering three methods of obtaining a sample of 500 students from the 70,000 students at the university.

(a) Because of financial constraints, the first method the administrator is considering consists of taking a convenience sample to keep the expenses low. A very large number of students will attend the first football game of the season, and the first 500 students who enter the football stadium could be used as a sample. Why might such a sampling method be biased in producing an estimate of the proportion of students who are satisfied with the appearance of the buildings and grounds?

(b) Because of the large number of students at the university, the second method the administrator is considering consists of using a computer with a random number generator to select a simple random sample of 500 students from a list of 70,000 student names. Describe how to implement such a method.

(c) Because stratification can often provide a more precise estimate than a simple random sample, the third method the administrator is considering consists of selecting a stratified random sample of 500 students. The university has two campuses with male and female students at each campus. Under what circumstance(s) would stratification by campus provide a more precise estimate of the proportion of students who are satisfied with the appearance of the university buildings and grounds than stratification by gender?

Q6: Zack claims that if 4 unrelated, randomly selected people were assembled, the chance that at least 2 of them would have the same birth month is greater than 50-50. This claim sounds far-fetched. You decide to simulate the process and repeat it many times.

1. What are your assumptions?

2. Carry out a simulation to see if Zack is right. Use your calculator, but first enter `18→rand` to "seed" the calculator. State how you assign digits to represent outcomes. What calculator command could you use to simulate 4 birth months?

3. Perform 30 repetitions and report the results. Use tally marks and fill in the table below.

Duplicates	
No duplicates	

4. What is your estimate of the probability of duplicate birth months?

Q7: Joey says that if every student in his AP Statistics class of 18 students picked a number at random from 1 to 226, and did this very many times, that about 50% of the time at least two students would pick the same number. You think this is an amazing claim, because there are only 18 students, and there are more than a dozen times that many numbers to choose from. You and your classmates are *very* skeptical, so you want to carry out a study to prove that Joey is wrong. But because this is a quiz, you have to carry out a simulation all by yourself.

1. To reduce any chance of bias (some people have favorite small numbers, like 7), have your calculator randomly pick a lower bound number between 1 and 750: **`randInt(1,750).`** This number will be the lower bound for your range of numbers. Then add 225 to your starting number to get the upper bound. Record your range of 226 numbers and how you obtained this range below.

2. Now use the random digits table, starting at line **113,** to generate 18 numbers, representing the numbers chosen by your classmates. Record the numbers from the table, and underscore the numbers that are in *your* range (including the lower and upper bound). When you find 18 numbers in your range, write them in a list and check to see if there is a match. If there is a match, write MATCH. If there is no match, write NO MATCH. Make sure you provide enough details so that your teacher can follow your reasoning.

```
Line 113   62568   70206   40325   03699   71080   22553   11486   11776
Line 114   45149   32992   75730   66280   03819   56202   02938   70915
Line 115   61041   77684   94322   24709   73698   14526   31893   32592
Line 116   14459   26056   31424   80371   65103   62253   50490   61181
```

3. Repeat the process once more, and record the results as before. Start at the beginning of the line following the point where you left off previously.

4. Finally, do you think it's possible that Joey's claim might be correct? Explain briefly

Q8:

A manufacturer of boots plans to conduct an experiment to compare a new method of waterproofing to the current method. The apperance of the boots is not changed by either method. The company recruits 100 volunteers in Seattle, where it rains frequently, to wear the boots as they normally would for 6 months. At the end of the 6 months, the boots well be returned to the company to be evaluated for water damage.

(a) **(10 points)** Describe a design for this experiment that uses the 100 volunteers. Include a few sentences on how it would be implemented.

(b) **(5 points)** Could your design be double blind? Explain.

Q9:

Every Monday a local radio station gives coupons away to 50 people who correctly answer a question about a news fact from the previous day's newspaper. The coupons given away are numbered from 1 to 50, with the first person receiving coupon 1, the second person receiving coupon 2, and so on, until all 50 coupons are given away. On the following Saturday, the radio station randomly draws numbers from 1 to 50 and awards cash prizes to the holders of the coupons with these numbers. Numbers continue to be drawn without replacement until the total amount awarded first equals or exceeds $300. If selected, coupons 1 through 5 each have a cash value of $200, coupons 6 through 20 each have a cash value of $100, and coupons 21 through 50 each have a cash value of $50.

(a) **(5 points)** Explain how you would conduct a simulation using the random number table provided below to estimate the distrubution of the number of prize winners each week.

72749	13347	65030	26128	49067	02904	49953	74674	94617	13317
81638	36566	42709	33717	59943	12027	46547	61303	46699	76423
38449	46438	91579	01907	72146	05764	22400	94490	49833	09258

(b) **(10 points)** Perform your simulation 3 times. (That is, run 3 trials of your simulation.) Start at the leftmost digit in the first row of the table and move across. Make your procedure clear so that someone can follow what you did. You must do this by marking directly on or above the table. Report the number of winners in each of your 3 trials.

Q1: 2003 FRQ

Solution

Part (a):

For example, a deadline in the department where the group of volunteers works has been moved back, lowering the stress levels of those working in the department. If the volunteers from this department were all in the same treatment group, this change in stress level could mistakenly be attributed to the treatment.

Without random assignment of volunteers to the two programs, it is possible that the two treatment groups could differ in some way that affects the outcome of the experiment. Randomization "evens out" the possible effects of potentially confounding variables.

Part (b):

Without the control group, the company could compare the two treatments, but would not be able to say whether the observed reduction in stress was attributable to participation in the programs. For example, a change in the work environment during this period might have reduced the stress level of all employees. The addition of a control group would enable the company to assess the magnitude of the mean reduction attributable to each treatment, as opposed to just determining if the two programs differ.

Part (c):

It is not reasonable to generalize the findings of this study to all employees, because

the participants in this experiment were volunteers and volunteers may not be representative of the population
OR
the participants were not randomly selected from the company employees.

Scoring

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

Part (a) has two components: the example, and the randomization.

- The example is scored as essentially correct (E) if it contains each of the elements in the table below:

Elements	Sample statements
1. Identify a plausible example of a problem	"Because a deadline has been moved back…"
2. Relate the identified problem to the change in stress level (the response)	"…lowering the stress levels of those working in the department. This change in stress level…"
3. …and state that the identified problem effects can not be distinguished from the difference in treatment effects	"…could mistakenly be attributed to the treatment." (Note: A construction such as "can't tell the difference" is OK here.)

The example is scored as partially correct (P) if the response contains 2 of the 3 components.

- The randomization is scored as essentially correct (E) if the student gives a reason for the necessity of random assignment. Possibilities include:

clearly stating in context that randomization is relied upon to create comparable groups

clearly stating in context that randomization controls for the effects of potentially confounding variables or reduces bias. (Both "Avoiding" bias and "Eliminating" bias are incorrect (I).)

The randomization is scored as partially correct (P) if the statement about randomization is not in context or is poorly communicated.

Note: Constructions such as "split up" and "divided into" can be interpreted to indicate randomization.

Q2:

Intent of Question

The primary goals of this question are to evaluate a student's ability to: (1) identify the treatments in a biological experiment; (2) present a completely randomized *design* to address the research question of interest; (3) describe the benefit of limiting sources of variability; and (4) describe the limitations to the scope of inference for the biologist.

Solution

Part (a):

The three different growth-enhancing nutrients (A, B, and C) and two different salinity levels (low and high) yield a total of $3 \times 2 = 6$ different treatment combinations for this experiment.

Treatment Combination	Nutrient	Salinity Level
1	A	Low
2	A	High
3	B	Low
4	B	High
5	C	Low
6	C	High

Part (b):

Since 10 tiger shrimps have already been randomly placed into each of 12 similar tanks in a controlled environment, we must randomly assign the treatment combinations to the tanks. Each treatment combination will be randomly assigned to 2 of the 12 tanks. One way to do this is to generate a random number for each tank. The treatment combinations are then assigned by sorting the random numbers from smallest to largest.

Treatment Combination	Nutrient	Salinity Level	Tanks with
1	A	Low	Smallest and second smallest random numbers
2	A	High	Third and fourth smallest random numbers
3	B	Low	Fifth and sixth smallest random numbers
4	B	High	Seventh and eighth smallest random numbers
5	C	Low	Ninth and tenth smallest random numbers
6	C	High	Next to largest and largest random numbers

After three weeks the weight gain (after – before) is computed for each tank, and the treatments are compared using appropriate averages.

Part (c):

Using only tiger shrimp will reduce a source of variation in the experimental units, the tanks of shrimp in this experiment. By eliminating this possible source of variation, type of shrimp, we are better able to isolate the variability due to the factors of interest to us (nutrient and salinity level). This will make it easier to identify any treatment effects that may be present.

Part (d):

Using only tiger shrimp will limit the scope of inference for the biologist. Ideally, the biologist would like to identify the treatment combination that leads to the most growth for all shrimp. However, the biologist will only be able to identify the best treatment combination for tiger shrimp because other types of shrimp may respond differently to the treatments.

Scoring

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

Part (a) is essentially correct (E) if all six treatments are correctly listed. This may be done in a 2 x 3 table or tree diagram but must clearly indicate the six treatments. A correct but incomplete listing of treatments in part (a) can be recovered in part (b) if the six treatments are clearly stated.

Listing the factors (nutrients A, B, C and salinity high, low) is incorrect and cannot be recovered in part (b).

Part (b) is essentially correct (E) if:

- each treatment combination is randomly assigned to 2 of the 12 tanks

  AND
- a correct procedure for randomization is described (so that two knowledgeable statistics users would use the same method to assign treatments to the tanks).

Part (b) is partially correct (P) if only one of these components is present. For example,

- Each treatment is randomly assigned to 2 of the 12 tanks, but the method of randomization is not fully described (i.e., just say randomly assign each treatment to 2 of the 12 tanks).

  OR
- A correct procedure for randomization of the treatments to the tanks is described, but each treatment does not necessarily appear twice.

Part (b) is incorrect (I) if there is no randomization or randomization of treatments is applied to the shrimps only (not the tanks).

Notes:

- If the randomization has been correctly applied to the tanks, additionally randomizing the shrimps or treatments will be regarded as extraneous.
- Because the stem indicates shrimp growth is to be compared, students are not required to identify a response variable in part (b) as was done in the model solution.

Part (c) is essentially correct (E) if

- the statistical advantage of reduced variability is identified

  AND
- an appropriate explanation that relates reduced variability to increasing the likelihood of determining differences among treatments is clearly provided.

Part (c) is partially correct (P) if only one of the two components is correct.

Part (c) is incorrect (I) if neither of the two components is present.

Notes:

- In this completely randomized design, confounding is not possible. Therefore a reference to confounding or lurking variables always incurs a penalty.

Part (d) is essentially correct (E) if

- the statistical disadvantage of limited scope of inference is identified

  AND
- an explanation that different species of shrimp may respond differently to treatments is provided.

(If the different responses to the treatments by other species of shrimp have been established in part (c), then it need not be repeated in part (d).)

Part (d) is partially correct (P) if only one of the two parts of the essentially correct response is provided.

Q3: 2009

Intent of Question

The primary goals of this question were to assess a student's ability to (1) describe a randomization process required for comparing two groups in a randomized experiment; and (2) describe a potential consequence of using self-selection instead of randomization.

Solution

Part (a) (completely randomized design):

Each student will be assigned a unique random number using a random number generator on a calculator, statistical software, or a random number table. The assigned numbers will be listed in ascending order. The students with the lowest 12 numbers in the ordered list will receive the instructional program that requires physically dissecting frogs. The students with the highest 12 numbers will receive the instructional program that uses computer software to simulate the dissection of a frog.

Part (a) *alternative* (randomized block design):

Students will be paired or placed into blocks of size two, based on having similar pretest scores. So, the first block will contain the two students with the two lowest pretest scores, the second block will contain the two students with the third- and fourth-lowest pretest scores, and so on, with the last block containing the two students with the two highest pretest scores. In each block, the students will be assigned a unique random number using a random number generator on a calculator, statistical software, or a random number table. The student in each block with the lower random number will receive the instructional program that requires physically dissecting frogs, and the student with the higher random number will receive the instructional program that uses computer software to simulate the dissection of a frog.

Part (b):

By not randomizing and allowing the students to self-select, there is a potential for changes to occur in the differences between pretest and posttest scores for a particular group because of the characteristics of students who choose a particular instructional method, not because of the instructional method itself. For example, suppose frog-loving students already know a lot about frog anatomy; one would therefore expect these students to be less likely to show a large change between the pretest and posttest scores. Suppose the frog-loving students tend to select the computer simulation method (perhaps because they do not like the notion of dissecting the frogs they love). The possible low change between pretest and posttest scores for the computer simulation group might then be attributed to the students' already knowing a lot about frog anatomy beforehand, not to the instructional method itself. The frog dissection group might see a larger change in scores because the students entering this group are those with the lower pretest scores (less prior knowledge) and who are thus more likely to show greater improvement between pretest and posttest scores.

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if a proper method of randomization is described that (1) creates two groups of equal size; *AND* (2) assigns the named treatments to the groups in a manner that knowledgeable statistics users would employ to assign the students to the two instructional groups.

Partially correct (P) if only one of the two criteria above is met.

Incorrect (I) if neither criterion is met.

Notes:

- Coin tossing (or equivalent method) using a stopping rule to obtain equal sample sizes requires placing the students in the class in a random order. If this method does not include a random order, at best, part (a) is scored as partially correct.
- In using a random number table, if numbers are specified, the student must work with two-digit numbers. For example, if using the first 24 integers, the student must use 01–24, not 1–24. If the student uses numbers such as 1–24, a solution that would otherwise be essentially correct becomes partially correct, and a partially correct response becomes incorrect.

Part (a) *alternative* is scored as follows:

Essentially correct (E) if (1) blocks are formed based on students' having similar pretest scores; *AND* (2) the two students in each block are assigned to different treatments; *AND* (3) the method of randomization used to assign the students in each block to the treatments is correct and can be implemented after reading the student's response (in a manner that knowledgeable statistics users would employ to assign the students to the two instructional groups).

Partially correct (P) if two of the three components above are presented correctly.

Incorrect (I) if no more than one of the three components is presented correctly.

Part (b) is scored as follows:

Essentially correct (E) if (1) the example gives a reasonable characteristic of the self-selected students in the study; *AND* (2) explains how this characteristic could be associated with changes in the differences between the pretest and posttest scores.

Partially correct (P) if (1) the example gives a reasonable characteristic of the self-selected students in the study; *AND* (2) a weak explanation is provided of how this characteristic could be associated with changes in the differences between pretest and posttest scores.

Note: A *weak* explanation of how a characteristic could be associated with changes in the differences between pretest and posttest scores must at least mention test scores or state that one group will perform better than the other. (Simply mentioning a behavioral difference is not sufficient.)

Incorrect (I) if an incorrect or no explanation is provided of how a characteristic could be associated with changes in the differences between pretest and posttest scores

OR

the example does not give a reasonable characteristic of the self-selected students in the study

OR

a student says that there must be an equal number of students in the class assigned to each treatment.

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

No part essentially correct and only one part partially correct

Q4: 2011

Intent of Question

The primary goals of this question were to assess students' ability to (1) describe a process for implementing cluster sampling; (2) describe a statistical advantage of stratified sampling over cluster sampling in a particular situation.

Solution

Part (a):

The following two-step process can be used to select the eight apartments.

Step 1: Generate a random integer between 1 and 9, inclusive, using a calculator, a computer program, or a table of random digits. Select all four apartments on the floor corresponding to the selected integer.

Step 2: Generate another random integer between 1 and 9, inclusive. If the generated integer is the same as the integer generated in step 1, continue generating random integers between 1 and 9 until a different integer appears. Again select all four apartments on the floor corresponding to the second selected integer.

The cluster sample consists of the eight apartments on the two randomly selected floors.

Part (b):

Because the amount of wear on the carpets in apartments with children could be different from the wear on the carpets in apartments without children, it would be advantageous to have apartments with children represented in the sample. The cluster sampling procedure in part (a) could produce a sample with no children in the selected apartments; for example, a cluster sample of the apartments on the third and sixth floors would consist entirely of apartments with no children. Stratified random sampling, where the two strata are apartments with children and apartments without children, guarantees a sample that includes apartments with and without children, which, in turn, would yield sample data that are representative of both types of apartments.

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the response correctly addresses the following two components:

1. Indication that two floors are randomly selected, with all four apartments on each of the selected floors forming the sample (or that the entire floors should be carpeted).
2. Description of a valid random sampling procedure for selecting two floors that could be implemented after reading the response (so that two knowledgeable statistics users would use the same method to select the floors).

Partially correct (P) if the response includes exactly one of the two components listed above.

Incorrect (I) if the response includes neither of the two components listed above *OR* the response does not involve taking a random sample of two floors out of the nine.

Note: Some possible errors in component 2 include the following:

- Using 10 random digits rather than nine
- Failing to explicitly deal with the issue of potentially repeated random numbers

Part (b) is scored as follows:

Essentially correct (E) if the response indicates the following two components:

1. The amount of carpet wear could be different for apartments with and without children.
2. The stratified random sample ensures that some apartments with children will be selected.

Partially correct (P) if the response includes exactly one of the two components listed above.

Incorrect (I) if the response fails to meet the criteria for E or P.

Notes

- If the response in part (b) says that this stratified sampling method guarantees proportional representation of apartments with and without children, then the second component is satisfied.
- If the sampling procedure in part (a) divides the floors into two groups — those that have apartments with children and those that do not ("prestratification") — and then selects one floor from each group, score part (b) based on the degree to which a statistical advantage of the stratified sampling in part (b) is addressed.