AP Exam Practice: Exploring Data

NAME_____

1. Two large corporations, A and B, hire many new college graduates as accountants at entry-level positions. In 2009 the starting salary for an entry-level accountant position was $36,000 a year at both corporations. At each corporation, data were collected from 30 employees who were hired in 2009 as entry-level accountants and were still employed at the corporation five years later. The yearly salaries of the 60 employees in 2014 are summarized in the boxplots below.

(a) Write a few sentences comparing the distributions of the yearly salaries at the two corporations.

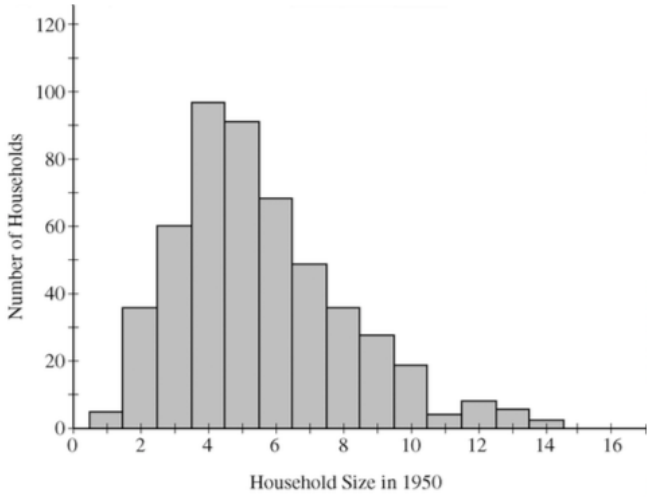
(b) Suppose both corporations offered you a job for $36,000 a year as an entry-level accountant.

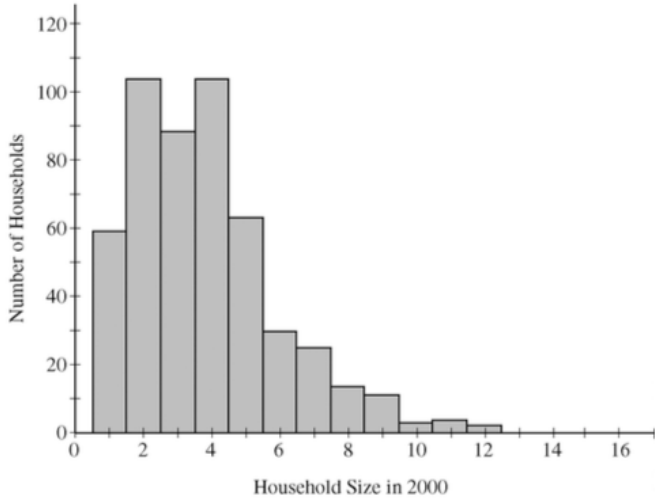
(i) Based on the boxplots, give one reason why you might choose to accept the job at corporation A.

(ii) Based on the boxplots, give one reason why you might choose to accept the job at corporation B.

2015 #1

2. Independent random samples of 500 households were taken from a large metropolitan area in the United States for the years 1950 and 2000. Histograms of household size (number of people in a household) for the years are shown below.

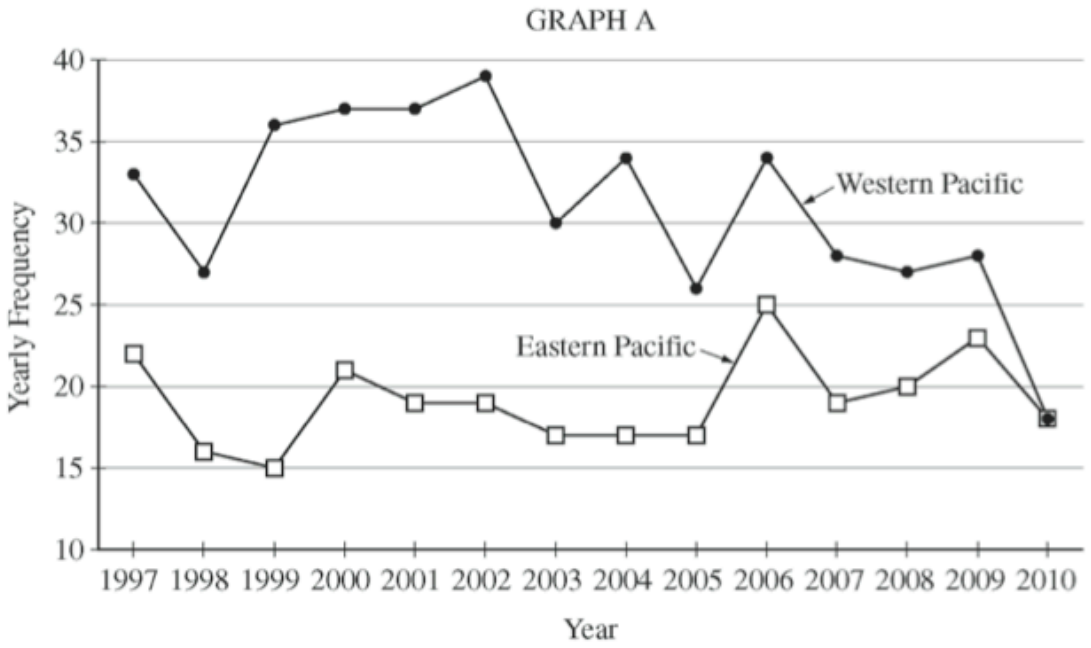




(a) Compare the distributions of household size in the metropolitan area for the years 1950 and 2000.

(b) A researcher wants to use these data to construct a confidence interval to estimate the change in mean household size in the metropolitan area from the year 1950 to the year 2000. State the conditions for using a two-sample t-procedure, and explain whether the conditions for inference are met.

3. Tropical storms in the Pacific Ocean with sustained winds that exceed 74 miles per hour are called typhoons. Graph A below displays the number of recorded typhoons in two regions of the Pacific Ocean—the Eastern Pacific and the Western Pacific—for the years from 1997 to 2010.



(a) Compare the distributions of yearly frequencies of typhoons for the two regions of the Pacific Ocean for the years from 1997 to 2010.

(b) For each region, describe how the yearly frequencies changed over the time period from 1997 to 2010.

2015 #1: Yearly Salaries

Student "F:"

(a) Write a few sentences comparing the distributions of the yearly salaries at the two corporations.

Corp A

Center at about \$50

Range (spread) of about \$42

(\$78 - \$36)

The distribution is shewed right

Corp B

- Center at about \$51
- Range (spread) of about \$17 (\$59-\$42)
- The distribution is roughly symmetrical

Student "E:"

(b) Suppose both corporations offered you a job for \$36,000 a year as an entry-level accountant.

(i) Based on the boxplots, give one reason why you might choose to accept the job at corporation A.

I would choose corporation A because the highest year salaries in A are bigger than the highest salary in B.

(ii) Based on the boxplots, give one reason why you might choose to accept the job at corporation B.

I would choose corporation B because the lowest year salary in B is higher than the lowest salary in A

F: No comparisons of center or spread, only "\$" for context. INCOMPLETE
E: No mention of "potential to earn more money" PARTIAL

2012 #3: Household Sizes

(a) Compare the distributions of household size in the metropolitan area for the years 1950 and 2000.

shape – Both distributions are skewed right; distribution of Year 2000 has a stronger skew

Center – The center has moved slightly (left) lower from 1950's to 2000's distribution.

spread – the spread of 1950's distribution is slightly greater than that of 2000's.

(a) Compare the distributions of household size in the metropolitan area for the years 1950 and 2000.

The distribution of household size in 1950 is slightly skewed to the right, but does not show any extreme outliers. It also includes households with sizes between 1 and 14.

The distribution of household size in 2000 is extremely skewed to the right and households with 10-12 people may be considered outliers. This distribution includes households with sizes between 1 and 12, which is different than the 1950's distribution.

2013 #6ab (Example B)

(a) Compare the distributions of yearly frequencies of typhoons for the two regions of the Pacific Ocean for the years from 1997 to 2010.

the distribution of the western pacific is slightly above the distribution of the Eastern pacific. Neither of the graphs have clusters. The strengths of the distributions are low since they don't follow a consecutive line. These scatterplots don't follow a shape.

(b)

western pacific:

From 1997 to 1988 the frequency decrease till it began to increase till 1999. From 1999 to 2002, it slighty increased but roughly stayed the same. It decreased to 2003, increased to 2004, decreased to 2005 increased to 2006 and decreased to 2008 with a slight increase but then a dramatic decrease to 2010 were the value meets up with the Eastern pacific.

Eastern pacific:

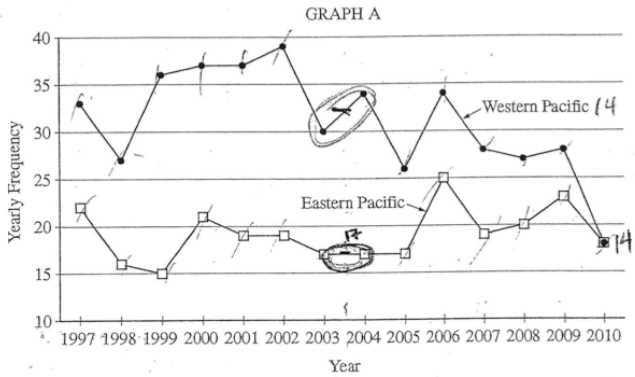
From 1997 to 1999 it decreases, increases to 2000. From 2000 to 2005 it moderately remains the same until it increases to 2006. 2006-2007 in decreases then increases to 2009. From 2009-2010 it decreases meeting the same value as western pacific.

GO ON TO THE NEXT PAGE

2013 #6ab (Example C)

(a) Compare the distributions of yearly frequencies of typhoons for the two regions of the Pacific Ocean for the years from 1997 to 2010.

The western pacific is much more variable and spread from 1997 to 2010 than the Eastern Pacific is. The center of the W.P. is also almost nearly twice the size of the E.P. The W.P. graph also appears to be more skewed than the E.P. graph.



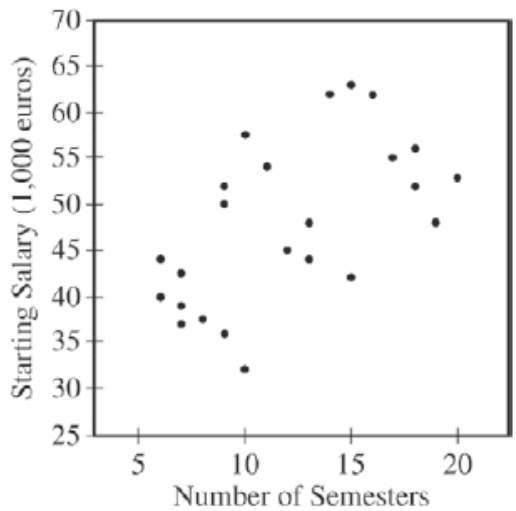
(b)

W.P. From 1997 to 1998, there was a drastic drop in the frequency. 1998 - 1999 also experienced a drastic change, only it was positive. From 1999 - 2002, there is steady rise and it reaches the peak frequency in 2002. After 2002 until 2007, there are drastic frequency changes both positively and negatively. 2007 to 2009, the frequency is on a steady decline. After 2009, the frequency takes another drastic plumet.

E.P. 1997 to 1999 experiences a drop in frequency. 1999 to 2000 there is a drastic positive increase. After 2000, there is a steady decline into a plateau in the years 2003 - 2005. An enormous spike happens between the years 2005 and 2006, and then a drop in 2006-0 From 2007 to 2009, the frequency slowly increases until a final drop yet again in 2009. The peak in frequency appears in 2006.

AP Exam Practice: Bivariate Data

A newspaper in Germany reported that the more semesters needed to complete an academic program at the university, the greater the starting salary in the first year of a job. The report was based on a study that used a random sample of 24 people who had recently completed an academic program. Information was collected on the number of semesters each person in the sample needed to complete the program and the starting salary, in thousands of euros, for the first year of a job. The data are shown in the scatterplot below.



(a) Does the scatterplot support the newspaper report about number of semesters and starting salary? Justify your answer.

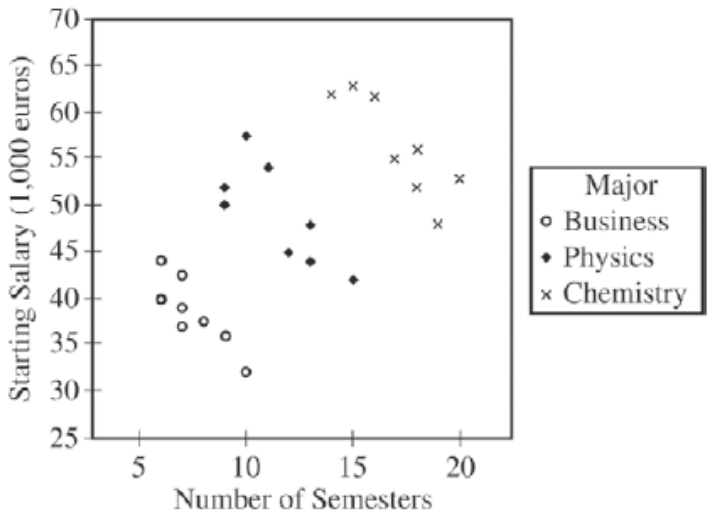
The table below shows computer output from a linear regression analysis on the data.

Predictor	Coef	SE Coef	T	P
Constant	34.018	4.455	7.64	0.000
Semesters	1.1594	0.3482	3.33	0.003
S = 7.37702	R-Sq = 33.5%		R-Sq(adj) = 30.5%	

(b) Identify the slope of the least-squares regression line, and interpret the slope in context.

(2016 #6)

An independent researcher received the data from the newspaper and conducted a new analysis by separating the data into three groups based on the major of each person. A revised scatterplot identifying the major of each person is shown below.

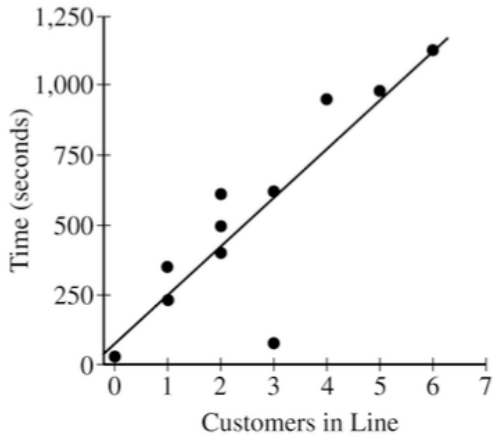


(c) Based on the people in the sample, describe the association between starting salary and number of semesters for the <u>business</u> majors.

(d) Based on the people in the sample, compare the median starting salaries for the three majors.

(e) Based on the analysis conducted by the independent researcher, how could the newspaper report be modified to give a better description of the relationship between the number of semesters and the starting salary for the people in the sample?

1. The manager of a grocery store selected a random sample of 11 customers to investigate the relationship between the number of customers in a checkout line and the time to finish checkout. As soon as the selected customer entered the end of a checkout line, data were collected on the number of customers in line who were in front of the selected customer and the time, in seconds, until the selected customer was finished with the checkout. The data are shown in the following scatterplot along with the corresponding least-squares regression line and computer output.



Predictor	Coef	SE Coef	T	P
Constant	72.95	110.36	0.66	0.525
Customers in line	174.40	35.06	4.97	0.001
S = 200.01		R-Sq = 73.33%	R-Sq (adj) = 70.37%	

(a) Identify and interpret in context the estimate of the intercept for the least-squares regression line.

(2018 #1)

(b) Identify and interpret in context the coefficient of determination, r^2 .

(c) One of the data points was determined to be an outlier. Circle the point on the scatterplot and explain why the point is considered an outlier.

ANSWERS to AP Exam Practice: Bivariate Data

2016 #6: Academic Semesters

Part (a):

The scatterplot supports the newspaper report about number of semesters needed to complete an academic program and starting salary because it shows a positive association between these two variables.

Part (b):

The slope is 1.1594. For each additional semester needed to complete an academic program, the predicted starting salary increases by €1,159.40.

Part (c):

For the business majors alone, there is a strong, negative, linear association between number of semesters and starting salary. Business majors who need a greater number of semesters to complete an academic program tend to have lower starting salaries.

Part (d):

Business majors have the lowest median starting salary at around €38,000, followed by physics majors at around €48,000, and then chemistry majors with the highest median starting salary at around €55,000.

Part (e):

The newspaper report should be modified to account for major. Overall, majors that take longer to complete tend to have higher starting salaries, with chemistry the highest, physics the next highest, and business the lowest. However, within a major, students who take a greater number of semesters tend to have lower starting salaries.

(a) Does the scatterplot support the newspaper report about number of semesters and starting salary? Justify your answer.

yes because the general direction of the spread on the scatterplot is positive, supporting the positive, linear correlation between number of semesters needed and starting salary. The newspaper states that as the number of semesters needed increases, the starting salary inc. which is not true of every point on the scatterplot, but is true of the overall positive trend.

(b) Identify the slope of the least-squares regression line, and interpret the slope in context.

$\hat{y} = ax + b$

(a) = slope = 1.1994

This slope means that for every 1 semester increase, the starting salary increases by 1.1994 thousands of euros.

Score: E P I Reason:

(c) Based on the people in the sample, describe the association between starting salary and number of semesters for the business majors.

There is a strong, negative, linear relationship between starting salary and number of semesters with the business majors. As the number of semesters increases there is a clear decrease in the starting salaries of the business majors.

(d) Based on the people in the sample, compare the median starting salaries for the three majors.

The median starting salary for business majors is the lowest of the three at about 38,000 euros. The Physics majors have the next highest median starting salary at about 47,000 euros. The chemistry majors have the highest median starting salary of about 55,000 euros.

Score: E P I Reason:

(e) Based on the analysis conducted by the independent researcher, how could the newspaper report be modified to give a better description of the relationship between the number of semesters and the starting salary for the people in the sample?

The newspaper ~~could~~ ~~seperate~~ could release a new article that retracts the statements of the old article and shows the new researchers findings. They could state that stratifying the data based off of major would yield more accurate findings. The newspaper report could show 3 seperate scatterplots with each major type and show the negative/inverse relationship between number of semesters and starting salary, as well as the median starting salary for each job.

Score: E P I Reason:

2018 #1 Notes:

Part (a) is scored as follows: Essentially correct (E) if the response satisfies the following three components:

1. Correctly identifies 72.95 as the intercept.
2. Communicates the concept of a y -intercept in a context that includes both time and zero customers.
3. Indicates that the value of the intercept is a prediction by using language such as **"predicted," "estimated," or "average"** value of y.

Part (c) is scored as follows: Essentially correct (E) if the response satisfies the following two components:

1. Correctly identifies the **outlier**.
2. Describes an unusual feature of the identified scatter plot point, **relative to the remaining data points**, that is sufficient to identify it as the outlier. Examples include:
   - The combination of x and y values is unusual compared to the other points.
   - The value of y is much lower than would be expected (or predicted), given the remaining data.
   - The residual for the point is unusually large relative to the other residuals. Partially correct (P) if the response satisfies component 1 but does not satisfy component 2. Incorrect (I) if the response does not meet the criteria for E or P.

Notes:

- In the absence of any point being circled on the graph, component 1 can still be satisfied by explicitly referring to the coordinates of the outlier. Valid coordinates for outlier identification must specify an x value of 3 and a y value that is strictly between 0 and 250.
- A response that does not make a comparison to the remaining data points, such as stating the outlier has a large residual or is nowhere near the regression line, does not satisfy component 2.
- A response that makes a comparison to the remaining data points based upon an unusual feature that is *insufficient* for outlier identification, such as stating the point is the only point with that particular y value, does not satisfy component 2.
- In the absence of explicit numerical calculation, a response that appeals to the influence that the outlier has on the regression coefficient estimates or on the sample correlation coefficient does not satisfy component 2.

2016 #6:
ab: P
(b lacks nondeterministic lang.)
cd: E
e:P
(retracting old is incorrect

AP Exam Practice: Sampling

1. At a certain university, students who live in the 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 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.

(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.

2. An environmental science teacher at a high school with a large population of students wanted to estimate the proportion of students at the school who regularly recycle plastic bottles. The teacher selected a random sample of students at the school to survey. Each selected student went into the teacher's office, one at a time, and was asked to respond yes or no to the following question.

Do you regularly recycle plastic bottles?

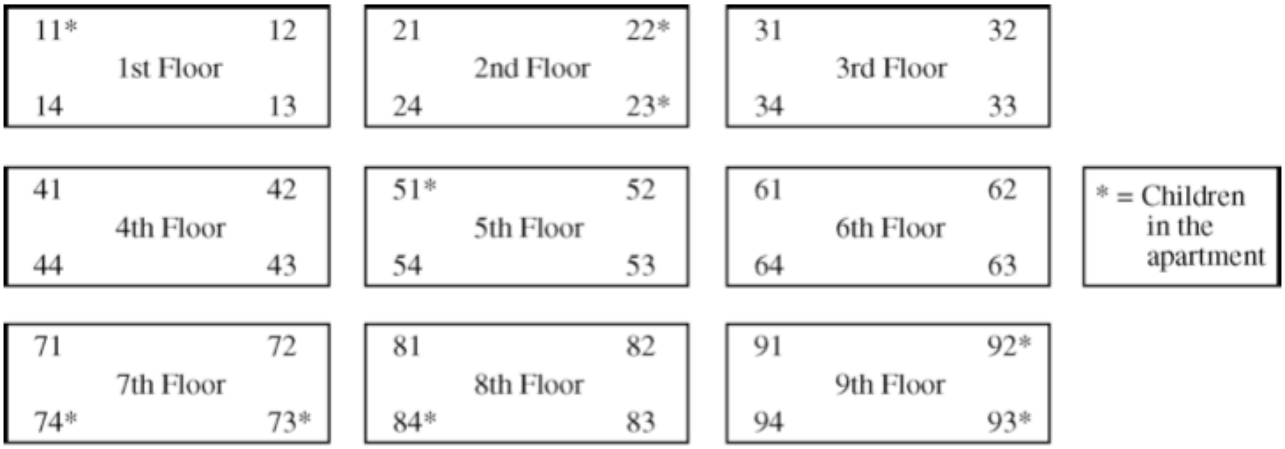
Based on the responses, a 95 percent confidence interval for the proportion of all students at the school who would respond yes to the question was calculated as (0.584, 0.816).

(a) How many students were in the sample selected by the environmental science teacher?

(b) Given the method used by the environmental science teacher to collect the responses, explain how bias might have been introduced and describe how the bias might affect the point estimate of the proportion of all students at the school who would respond yes to the question.

3. 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.



(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.

ANSWERS to AP Exam Practice: Sampling

1. 2004B #2--Dining Hall
   a) Students who arrive early may have different opinions about food quality than other students (late-diners), thus producing a biased sample.
      This bias can be avoided by taking a simple random sample of all dorm residents.
   b) The phrase about the dining hall food needing improvement may lead more students to respond that the food really does need improvement. Also, the phrase indicating increased cost may also influence students to answer insincerely.
      Better wording: "Do you feel the quality of food needs improvement?"

2. 2018 #2

Part (b)

Essentially correct (E) if the response includes the following three components:

1. Explains <u>why</u> the responses to the survey might differ from the truth about student recycling in this context (e.g., the survey was not anonymous, the question was asked by an authority figure)
2. Explains <u>how</u> the responses to the survey might differ from the truth about student recycling (e.g., "students might say yes when they actually don't recycle", "students lie and say yes", "students don't recycle but lie to the teacher")
3. Describes the effect of the bias on the point estimate (or the proportion, percentage, number of "yes" responses in the sample) and doesn't contradict the bias described.

Partially correct (P) if the response includes only two of the components required for E.

Incorrect (I) if the response includes at most one of the components required for E.

Notes:

- To satisfy component 1, the response must provide a reason that is based on a bias created by the teacher asking students in person. For example, a response that addresses the wording of the question, voluntary response, or sampling variability does not satisfy component 1.
- To satisfy component 2, the response needs to explicitly contrast what the students say with what they do. It isn't enough to contrast saying yes with saying no.
- Evidence used to address component 3 cannot be used to also address component 2. For example, a response that says "Students might lie, producing an estimate that is too high" addresses the effect of the bias on the point estimate, but should not be combined with the statement about students lying to infer that students do not actually recycle. However, a response that says "Students may lie and say yes, producing an estimate that is too high" satisfies both components 2 and 3.
- If the response is clearly about the population proportion and not about the point estimate, component 3 cannot be satisfied.
- Statements such as "the interval will be too high" do not satisfy component 3 because they don't specifically address the point estimate.

3. 2011 #3--Sampling Apartments
   a) Use a random number generator to select two <u>different</u> random integers from 1 to 9. The two numbers correspond to the two floors that will be sampled, for a total of eight apartments.
   b) Using the cluster sampling method described above, it would be possible to select NO apartments with children if, say, floors 3 and 6 were picked. Since it's reasonable to believe that carpets of apartments with children will have more wear than apartments without children, a cluster sample may not provide the owner with a good representative sample of carpet wear in this building.
      A stratified sample as described will allow the owner to see how well the carpet wears in both apartments with and without children.

2018 #2: Environmental Science Teacher (Student Samples)

(b) Given the method used by the environmental science teacher to collect the responses, explain how bias might have been introduced and describe how the bias might affect the point estimate of the proportion of all students at the school who would respond yes to the question.

There is bias in this sample because students do not want the teacher to know that they do not recycle. Therefore, the proportion of the sample of student who regularly recycle plastic bottles is higher than the true proportion.

(b) Given the method used by the environmental science teacher to collect the responses, explain how bias might have been introduced and describe how the bias might affect the point estimate of the proportion of all students at the school who would respond yes to the question.

There could be bias when the environmentc science teacher who's in favor of recycling askes the students if they recycle. The student may be embarassed to say no or pressured to say yes in fear of the teacher judging or disapproving of them. This is an example of response bias. This could cause an over prediction of the population proportion of students at that school who recycle.

Both earn a "P." Neither student states how the survey response would differ from the truth.

2011 #3: Apartments (Student Samples)

Student #1, Part (a):

(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.

The way this method is used is a floor is first randomly selected, then an apartment in the floor is randomly selected. And you keep on doing it until the person selected eight different apartments.

Part (b):

The advantage of this that you get to see how the carpet wears with an apartment with children and in an apartment without children. If you use the cluster sampling method is you run the risk of not selecting an apartment with children for example. Or maybe not maybe not enough to render it statistically significant. With the stratified random sample its a guarentee that you are going to get enough of both.

Student #2:

Part (a):

apartments using this method.

You could randomly select 2 floors to carpet instead of randomly selecting individual rooms to carpet.

Part (b):

An advantage to this would be to see if kids play a big role in the carpets wear because if you use a cluster sample on the floors you might not get any kids located on that floor.

1a: No cluster, no method (I); b: (E)
2a: No "how" (P); b: weak, but (E)

Experimental Design Problems

(from the AP Statistics Exam) NAME_____

2019 #2

Researchers are investigating the effectiveness of using a fungus to control the spread of an insect that destroys trees. The researchers will create four different concentrations of fungus mixtures: 0 milliliters per liter (ml/L), 1.25 ml/L, 2.5 ml/L, and 3.75 ml/L. An equal number of the insects will be placed into 20 individual containers. The group of insects in each container will be sprayed with one of the four mixtures, and the researchers will record the number of insects that are still alive in each container one week after spraying.

(a) Identify the treatments, experimental units, and response variable of the experiment.

Treatments:

Experimental units:

Response variable:

(b) Does the experiment have a control group? Explain your answer.

(c) Describe how the treatments can be randomly assigned to the experimental units so that each treatment has the same number of units.

2. As dogs age, diminished joint and hip health may lead to joint pain and thus reduce a dog's activity level. Such a reduction in activity can lead to other health concerns such as weight gain and lethargy due to lack of exercise. A study is to be conducted to see which of two dietary supplements, glucosamine or chondroitin, is more effective in promoting joint and hip health and reducing the onset of canine osteoarthritis. Researchers will randomly select a total of 300 dogs from ten different large veterinary practices around the country. All of the dogs are more than 6 years old, and their owners have given consent to participate in the study. Changes in joint and hip health will be evaluated after 6 months of treatment.

   (a) What would be an advantage to adding a control group in the design of this study?

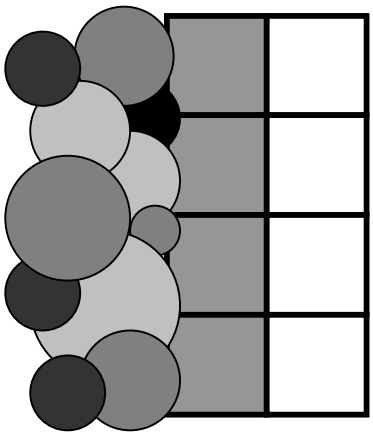
   (b) Assuming a control group is added to the other two groups in the study, explain how you would assign the 300 dogs to these three groups for a completely randomized design.

   (c) Rather than using a completely randomized design, one group of researchers proposes blocking on clinics, and another group of researchers proposes blocking on breed of dog. How would you decide which <u>one</u> of these two variables to use as a blocking variable?

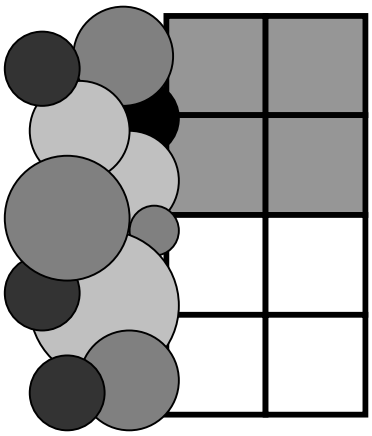
3. Students are designing an experiment to compare the productivity of two varieties of dwarf fruit trees. The site for the experiment is a field that is bordered by a densely forested area on the west side. The field has been divided into eight plots of approximately the same area. The students have decided that the test plots should be blocked. Four trees, each of two varieties, will be assigned at random to the four plots within each block, with one tree planted in each plot.

The two blocking schemes shown below are under consideration. For each scheme, one block is identified by the white region and the other block indicated by the grey region in the figures.

Blocking Scheme A



Blocking Scheme B



<u>Key</u>

Block 1

Block 2

a. Which of the blocking schemes , A or B, is better for this experiment? Explain.

b. Even though the students have decided to block, they must randomly assign the varieties of the trees to the plots within each block. What is the purpose of this randomization in the context of the experiment?

4. The dentists in a dental clinic would like to determine if there is a difference between the number of new cavities in people who eat an apple a day and in people who eat less than one apple a week. They are going to conduct a study with 50 people in each group.

Fifty clinics patients who report that they routinely eat an apple a day and 50 clinic patients who report that they eat less than one apple a week will be identified. The dentists will examine the patients and their records to determine the number of new cavities the patients have had over the past two years. They will then compare the number of new cavities in the two groups.

a. Why is this an observational study and not an experiment?

b. Explain the concept of confounding in the context of this study. Include an example of a possible confounding variable.

c. If the mean number of new cavities for those who ate an apple a day was statistically significantly smaller than the mean number of new cavities for those who ate less than one apple a week, could one conclude that the lower number of new cavities can be attributed to eating an apple a day? Explain.

5. A biologist is interested in studying the effect of growth-enhancing nutrients an differernt 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 <u>statistical</u> advantage to having only tiger shrimps in the experiment. Explain why this is an advantage.

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

6. When a tractor pulls a plow through an agricultural field, the energy needed to pull that plow is called the draft. The draft is affected by environmental conditions such as soil type, terrain, and moisture.

A study was conducted to determine whether a newly developed hitch would be able to reduce draft compared to the standard hitch. (A hitch is used to connect the plow to the tractor.) Two large plots of land were used in this study. It was randomly determined which plot was to be plowed using the standard hitch. As the tractor plowed that plot, a measurement device on the tractor automatically recorded the draft at 25 randomly selected points in the plot.

After the plot was plowed, the hitch was changed from the standard one to the new one, a process that takes a substantial amount of time. Then the second plot was plowed using the new hitch. Twenty-five measurements of draft were also recorded at randomly selected points in this plot.

a) What was the response variable in this study?

Identify the treatments.

What were the experimental units?

b) Given that the goal of the study is to determine whether a newly developed hitch reduces draft compared to the standard hitch, was randomization used properly in this study? Justify your answer.

c) Given that the goal of the study is to determine whether a newly developed hitch reduces draft compared to the standard hitch, was replication used properly in this study? Justify your answer.

d) Plot of land is a confounding variable in this experiment. Explain why.

SOLUTIONS to Experimental Design Problems

1. Fungus Spray (2019 #2)

Part (a) is scored as follows:

Essentially correct (E) if the response satisfies the following three components:

1. Identifies the 4 concentrations (or mixtures or sprays) as the treatments
2. Identifies the 20 containers as the experimental units
3. Identifies the number of insects that are still alive in each container as the response variable

(LOTS of notes about part (a)…)

Part (b) is scored as follows:

Essentially correct (E) if the response indicates that there is a control group and justifies this claim by identifying the control group or by explaining that there is a treatment which contains no fungus.

Part (c) is scored as follows:

Essentially correct (E) if the response satisfies the following three components:

1. Creates appropriate labels for the units/treatments (e.g., label the containers from 1–20, label 20 slips of paper with 5 for each treatment)
2. Describes how to correctly implement the random assignment process
3. The random assignment process results in an equal number of experimental units assigned to each treatment

2. Dogs' Hip Health (2007 #2)

Part (a):

A control group gives the researchers a comparison group to be used to evaluate the effectiveness of the treatments. The control group allows the impact of the normal aging process on joint and hip health to be measured with appropriate response variables. The effects of glucosamine and chondroitin can be assessed by comparing the responses for these two treatment groups with those for the control group.

Part (b):

Each dog will be assigned a unique random number, 001–300, using a random number generator on a calculator, statistical software, or a random number table. The numbers will be sorted from smallest to largest. The dogs assigned the first 100 numbers in the ordered list will receive glucosamine. The dogs with the next 100 numbers in the ordered list will be assigned to the control group. Finally, the dogs with the numbers 201–300 will receive chondroitin.

Part (c):

The key question is which variable has the strongest association with joint and hip health. The goal of blocking is to create groups of homogeneous experimental units. It is reasonable to assume that most clinics will see all kinds and breeds of dogs so there is no reason to suspect that joint and hip health will be strongly associated with a clinic. On the other hand, different breeds of dogs tend to come in different sizes. The size of a dog is associated with joint and hip health, so it would be better to form homogeneous groups of dogs by blocking on breed.

(a) What would be an advantage to adding a control group in the design of this study?

An advantage to adding a control group to this design would be that it gives the experiment something to compare its results to, to see how much of a difference the treatments make.

(b) Assuming a control group is added to the other two groups in the study, explain how you would assign the 300 dogs to these three groups for a completely randomized design.

For every dog that is chosen roll a die. IF the die is a 1 or 2 give the dog the glucosamine. IF the die is a 3 or 4 give the dog the chondroitin. IF the die is a 5 or 6 put the dog in the control group. This will completely randomize the design.

(c) Rather than using a completely randomized design, one group of researchers proposes blocking on clinics, and another group of researchers proposes blocking on breed of dog. How would you decide which <u>one</u> of these two variables to use as a blocking variable?

I would decide to use the blocking on breed of dog. The clinic the dog is in should not affect the medicine the dog is given. However, different breeds of dogs might respond to the medicines differently. Therefore, the blocking on breed of dog should be used.

(a) What would be an advantage to adding a control group in the design of this study?

The advantage to adding a control group in the design of this study would be to have something to compare the results to. This helps to reduce the effects of confounding variables. For example the weather which can affect joint pain.

(b) Assuming a control group is added to the other two groups in the study, explain how you would assign the 300 dogs to these three groups for a completely randomized design.

To obtain a compleatly randomized design I would number each dog 1 to 300 and then using a random number generator I would select 100 numbers ignoring repeats the 100 dogs corresponding to those 100 numbers will be placed in the first treatment group and will recieve glucosamine. I will repeat this process selecting 100 new numbers, these 100 dogs will be placed in the second treatment group and will recieve chondroitin and the remaining 100 dogs will be the control group and will recieve a placeb.

(c) Rather than using a completely randomized design, one group of researchers proposes blocking on clinics, and another group of researchers proposes blocking on breed of dog. How would you decide which one of these two variables to use as a blocking variable?

Whichever variable has more variation should be used as a block. I think breed of dog will cause more variation in the experiment because different kinds of dogs can respond differently to the treatment, but which clinic the dogs come from probably will have less effect on the experiment.

(a) What would be an advantage to adding a control group in the design of this study?

It would be an advantage to add a control group to this study because then after 6 months, you have a group to compare with the treated dogs in the study, to see if the treatments really had an impact in promoting joint and hip health and reducing the onset of canine osteoarthritis.

(b) Assuming a control group is added to the other two groups in the study, explain how you would assign the 300 dogs to these three groups for a completely randomized design.

For a completely randomized design, I would assign each of the 300 dogs a number, 1-300, and then put all the numbers into a hat. Then draw 100 numbers out of the hat and assign them to group 1, the control group. Then pull out 100 more numbers and assign them to group 2, the glucosamine treatment. Then with the 100 left over dogs assign them to group 3, for the chondroitin treatment. That way, you will have three groups for a completely randomized design.

(c) Rather than using a completely randomized design, one group of researchers proposes blocking on clinics, and another group of researchers proposes blocking on breed of dog. How would you decide which <u>one</u> of these two variables to use as a blocking variable?

Rather than using a completely randomized design, I would encorporate blocking on the specific breed of dog, because the different treatments could possibly have a different effect on the different types of dogs, therefore I would use the blocking on breed of dog so it would eliminate any variables that could change the actual data. I would rather block on breed of dog than clinic, because breeds of dogs seem to be much more different than different clinics, therefore I would block on breeds of dogs.

PEP = 2 (No context in (a); no comparison in (c))
EEE = 4 (Complete)
EEP = 3 ("eleminate any variables")

3. Fruit Trees (2001 #4)

Part (a):

Blocking scheme A is preferable because it creates homogeneous blocks with respect to forest exposure. That is, plots in the same block have similar exposure to the forest.

Part (b):

Randomization of varieties of trees to the plots within each block should reduce any possible bias due to confounding variables, such as fertility or moisture, on the productivity of the two types of dwarf trees.

OR

Randomization of varieties of trees to the plots within each block should even out (or equalize) the effect of other characteristics of the plots that might be related to the productivity of the trees.

4. Dentists and Apples (1999 #3)

a. The student can appeal to any of three reasons in judging this study not an experiment:

   1. there is no random assignment of subjects to treatments;
   2. there are no treatments imposed;
   3. existing data is being used.

b. Two variables are confounded if their effect on the number of new cavities cannot be distinguished from one another. The student must mention not only that the confounding variables may affect the outcome but that they have differential effects within the two groups. For instance: confounding would occur if patients who eat an apple a day differ from those who eat less than one apple a week **on some variable that is related to dental health**. In this example, diet or general level of health are examples of what might be confounding variables. For example, it is possible that people who eat an apple a day are more nutrition conscious and have a more healthy diet in general than those who eat one or fewer apples per week, and this might explain the observed difference in dental health.

Note:
There are many possible examples of confounding variables. Any reasonable example of a confounding variable is acceptable, as long as a good explanation is given and the connection between the confounding variable and group membership is clear. Lack of a definition here can be rectified by a response in (c) that demonstrates a clear understanding of the concept of confounding variable.

c. No, because it is not an experiment, and cause-and-effect conclusions cannot be drawn from an observational study.

   OR

   No, because there are possible confounding variables.

Notes:

1. In b), a good definition of confounding with a bad example should be regarded as temporarily weak. An example that does not mention group affiliation can be recovered in part c). To recover the definition of confounding in c) they must connect the term with the definition. To recover the group affiliation, they may do by example in c).

2. If the student, in attempting to discuss group differentiation, only mentions one of the groups, that is OK - we will consider the other implied. For example, it is counted correct if the student says, "The apple-eating group may be more health-conscious…" they need not explicitly deny health-consciousness to the one-apple-a-week group.

3. The constructions "Some people" and "A person in the apple-eating group may …" are NOT enough to establish group differentiation; this construction suggests only that some subset of the group may differ from the rest of the group. This is just natural variation. The construction "A person who is an apple-eater may …" can establish group differentiation if it is clear that this is describing a representative member of the group.

4. Mentioning initial non-equivalence of groups without tying that non-equivalence to the outcome is not correct. Mentioning concepts such as self-reporting bias, social desirability, etc. may constitute measurement error in the study but is not confounding.

5. In (c), appealing to the definition of confounding variables in (b) would get a "correct" for (c) if the definition in (b) is correct. If the definition in (b) is weak, that appeal alone would not get credit.

6. If the definition in (b) is "there are other variables that affect the outcome measure, such as age, health, etc," this is **not** regarded as a correct definition for purposes of appealing to the definition from (c). For purposes of reading part (b), this definition would be regarded as weak.

7. If in (c), they give an example which is the equivalent of confounding, and **refer to this as confounding**, they would get credit for (b).

5. Shrimp (2006 #5)

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.

Notes:

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

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.

6. 2006B, #5

Part (a):

The response variable was the amount of draft. The two treatments were the standard hitch and the new hitch. The experimental units were the two large plots of land.

Part (b):

Yes, the two hitches (treatments) were randomly assigned to the two plots (experimental units).

Part (c):

No, each treatment (type of hitch) was applied to only one experimental unit (plot of land). Replication is used to repeat the treatments on different experimental units so general patterns can be observed. There is no replication in this study.

Part (d):

Although 25 measurements were taken at different locations in the two plots, each hitch was used in one plot (experimental unit) only. Thus, if a difference in the draft is observed we will not know whether the difference is due to the hitch or the plot. In statistical language, the treatments (hitches) are confounded with the plots.

AP Exam Practice: Probability

4. A company manufactures model rockets that require igniters to launch. Once an igniter is used to launch a rocket, the igniter cannot be reused. Sometimes an igniter fails to operate correctly, and the rocket does not launch. The company estimates that the overall failure rate, defined as the percent of all igniters that fail to operate correctly, is 15 percent.

   A company engineer develops a new igniter, called the super igniter, with the intent of lowering the failure rate. To test the performance of the super igniters, the engineer uses the following process.

   Step 1: One super igniter is selected at random and used in a rocket.

   Step 2: If the rocket launches, another super igniter is selected at random and used in a rocket.

   Step 2 is repeated until the process stops. The process stops when a super igniter fails to operate correctly or 32 super igniters have successfully launched rockets, whichever comes first. Assume that super igniter failures are independent.

   (a) If the failure rate of the super igniters is 15 percent, what is the probability that the first 30 super igniters selected using the testing process successfully launch rockets?

   (b) Given that the first 30 super igniters successfully launch rockets, what is the probability that the first failure occurs on the thirty-first or the thirty-second super igniter tested if the failure rate of the super igniters is 15 percent?

   (c) Given that the first 30 super igniters successfully launch rockets, is it reasonable to believe that the failure rate of the super igniters is less than 15 percent? Explain.

3 . There are four runners on the New High School team. The team is planning to participate in a race in which each runner runs a mile. The team time is the sum of the individual times for the four runners. Assume that the individual times for the four runners are all independent of each other. The individual times, in minutes, of the runners in similar races are approximately normally distributed with the following means and standard deviations.

	Mean	Standard Deviation
Runner 1	4.9	0.15
Runner 2	4.7	0.16
Runner 3	4.5	0.14
Runner 4	4.8	0.15

a) Runner 3 thinks he can run a mile in less than 4.2 minutes in the next race. Is that likely to happen? Explain.

b) The distribution of possible team times is approximately normal. Find the mean and standard deviation of this distribution.

c) Suppose the team's best time to date is 18.4 minutes. What is the probability that the team will beat its own best time in the next race?

3. Each full carton of Grade A eggs consists of 1 randomly selected empty cardboard container and 12 randomly selected eggs. The weights of such full cartons are approximately normally distributed with a mean of 840 grams and a standard deviation of 7.9 grams.

(a) What is the probability that a randomly selected full carton of Grade A eggs will weigh more than 850 grams?

(b) The weights of the empty cardboard containers have a mean of 20 grams and a standard deviation of 1.7 grams. It is reasonable to assume independence between the weights of the empty cardboard containers and the weights of the eggs. It is also reasonable to assume independence among the weights of the 12 eggs that are randomly selected for a full carton.

Let the random variable X be the weight of a single randomly selected Grade A egg.

i) What is the mean of X?

ii) What is the standard deviation of X ?

3. A medical researcher surveyed a large group of men and women about whether they take medicine as prescribed. The responses were categorized as never, sometimes, or always. The relative frequency of each category is shown in the table.

	Never	Sometimes	Always	Total
Men	0.0564	0.2016	0.2120	0.4700
Women	0.0636	0.1384	0.3280	0.5300
Total	0.1200	0.3400	0.5400	1.0000

(a) One person from those surveyed will be selected at random.

(i) What is the probability that the person selected will be someone whose response is never and who is a woman?

(ii) What is the probability that the person selected will be someone whose response is never or who is a woman?

(iii) What is the probability that the person selected will be someone whose response is never given that the person is a woman?

(b) For the people surveyed, are the events of being a person whose response is never and being a woman independent? Justify your answer.

(c) Assume that, in a large population, the probability that a person will always take medicine as prescribed is 0.54. If 5 people are selected at random from the population, what is the probability that at least 4 of the people selected will always take medicine as prescribed? Support your answer.

ANSWERS: AP Exam Practice: Probability

2016 #4: Model Rockets

Part (a):

If the failure rate for the super igniters is 15 percent, then the probability that each igniter fails is 0.15, and the probability that it does not fail is 0.85. Therefore the probability that the first 30 igniters tested do not fail is $(0.85)^{30} \approx 0.0076$. The solution can also be written as $(1 - 0.15)^{30} \approx 0.0076$.

Part (b):

Given that there are no failures in the first 30 trials, the probability that the first failure occurs on the 31st trial is 0.15, and the probability that it does not occur on the 31st but occurs on the 32nd trial is $(0.85)(0.15) = 0.1275$. Therefore the probability that the first failure occurs on the 31st or 32nd super igniter tested is $0.15 + 0.1275 = 0.2775$.

Note that this is equivalent to asking for the probability that the first failure occurs on the first or second trial, which is $0.15 + (0.85)(0.15) = 0.2775$.

Part (c):

The result of the probability calculation in part (a) provides a reason to believe that the failure rate of the super igniters is less than 15 percent. The calculated probability of 0.0076 shows that there is less than a 1 percent chance that 30 or more igniters in a row would not fail if the failure rate was 15 percent. This probability is smaller than conventional significance levels such as $\alpha = 0.05$ or $\alpha = 0.01$, and thus is small enough to make it reasonable to believe that the failure rate of the super igniters is less than 15 percent.

2002 #3 (Four Runners)

Part (a) is *essentially correct* if:

the probability is calculated correctly, it is correctly assessed as unlikely, and the justification is acceptable

OR

the student does not compute the probability, but appeals to the fact that a time of 4.2 minutes is more than 2 standard deviations below the mean of a normal distribution and then uses this information to reach a conclusion with appropriate communication.

Part (a) is *partially correct* if:

the probability computed is not correct (for example, $P(z > -2.14)$ or $P(z < +2.14)$ might be computed), but the given probability is correctly assessed

OR

an argument is based on the number of standard deviations from the mean without invoking normality.

Part (b) is *essentially correct* if both the mean and the standard deviation of the team time distribution are correctly computed (except for purely arithmetic mistakes).

Part (b) is *partially correct* if only one of these is correctly computed (except for purely arithmetic mistakes).

CAUTION: A standard deviation of .3 (numerically correct) can arise from this incorrect calculation: $\frac{(.15+.16+.14+.15)}{\sqrt{4}} = 0.3$

Part (c) is *essentially correct* if the probability is correctly calculated using a mean which is either correct or carried from (b) as well as a standard deviation which is either correct or carried from (b).

Part (c) is *partially correct* if:

both the mean and standard deviation are correct or carried from (b), but the computed probability is incorrect
OR
the mean or standard deviation is incorrectly derived from (b) but the subsequent probability calculation is correct.

2013 #3: Grade A Eggs

Part (a):

Let W denote the weight of a randomly selected full carton of eggs. W has a normal distribution with mean 840 grams and standard deviation 7.9 grams.

The z-score for a weight of 850 grams is $z = \frac{850 - 840}{7.9} \approx 1.27$.

The standard normal probability table reveals that
$P(W > 850) = P(Z > 1.27) \approx 1 - 0.8980 = 0.1020$.

Part (b):

(i) Let W represent the weight of a randomly selected full carton of eggs, P the weight of the packaging, and X_i the weight of the *ith* egg, for $i = 1, 2, \ldots, 12$.

Note that $W = P + X_1 + X_2 + \ldots + X_{12}$.
Properties of expected values establish that $\mathrm{E}(W) = \mathrm{E}(P) + \mathrm{E}(X_1) + \ldots + \mathrm{E}(X_{12})$.
Because all 12 eggs have the same mean weight, this becomes $\mathrm{E}(W) = \mathrm{E}(P) + 12 \times \mathrm{E}(X_i)$.
We were told that $\mathrm{E}(W) = 840$ and $\mathrm{E}(P) = 20$, so we can solve
$840 = 20 + 12 \times \mathrm{E}(X_i)$ to find $\mathrm{E}(X_i) = \frac{840 - 20}{12} \approx 68.33$ grams.

(ii) Because of independence, properties of variance establish that
$$\mathrm{Var}(W) = \mathrm{Var}(P) + \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \ldots + \mathrm{Var}(X_{12}).$$

Because all 12 eggs have the same variance of their weights, this becomes
$$\mathrm{Var}(W) = \mathrm{Var}(P) + 12 \times \mathrm{Var}(X_i).$$

We were told that $\mathrm{SD}(W) = 7.9$ and $\mathrm{SD}(P) = 1.7$. Therefore, $\mathrm{Var}(W) = (7.9)^2 = 62.41$ and $\mathrm{Var}(P) = (1.7)^2 = 2.89$.
We can solve $62.41 = 2.89 + 12 \times \mathrm{Var}(X_i)$ to find $\mathrm{Var}(X_i) = \frac{62.41 - 2.89}{12} = 4.96$. Thus, $\mathrm{SD}(X_i) = \sqrt{(4.96)} \approx 2.23$ grams.

2019 #3: Medicine Survey

Part (a):

(i) $P(\text{never and woman}) = 0.0636$.

(ii)
$$\begin{aligned} P(\text{never or woman}) &= P(\text{never}) + P(\text{woman}) - P(\text{never and woman}) \\ &= 0.12 + 0.53 - 0.0636 \\ &= 0.5864 \end{aligned}$$

(iii) $P(\text{never} \mid \text{woman}) = \frac{P(\text{never and woman})}{P(\text{woman})} = \frac{0.0636}{0.53} = 0.12$

Part (b):

Yes, the event of being a person who responds never is independent of the event of being a woman because

$$P(\text{never} \mid \text{woman}) = P(\text{never}) = 0.12.$$

Part (c):

Define X as the number of people in a random sample of 5 people who always take their medicine as prescribed. Then X has a binomial distribution with $n = 5$ and $p = 0.54$, and

$$\mathrm{P}(X \geq 4) = \binom{5}{4}(0.54)^4(0.46)^1 + \binom{5}{5}(0.54)^5(0.46)^0 = 0.19557 + 0.04592 = 0.24149.$$

AP Exam Practice: Sampling Distributions

2. A local radio station plays 40 rock-and-roll songs during each 4-hour show. The program director at the station needs to know the total amount of airtime for the 40 songs so that time can also be programmed during the show for news and advertisements. The distribution of the lengths of rock-and-roll songs, in minutes, is roughly symmetric with a mean length of 3.9 minutes and a standard deviation of 1.1 minutes.

   (a) Describe the sampling distribution of the sample mean song lengths for random samples of 40 rock-and-roll songs.

   (b) If the program manager schedules 80 minutes of news and advertisements for the 4-hour (240-minute) show, only 160 minutes are available for music. Approximately what is the probability that the total amount of time needed to play 40 randomly selected rock-and-roll songs exceeds the available airtime?

3. Trains carry bauxite ore from a mine in Canada to an aluminum processing plant in northern New York state in hopper cars. Filling equipment is used to load ore into the hopper cars. When functioning properly, the actual weights of ore loaded into each car by the filling equipment at the mine are approximately normally distributed with a mean of 70 tons and a standard deviation of 0.9 ton. If the mean is greater than 70 tons, the loading mechanism is overfilling.

(a) If the filling equipment is functioning properly, what is the probability that the weight of the ore in a randomly selected car will be 70.7 tons or more? Show your work.

(b) Suppose that the weight of ore in a randomly selected car is 70.7 tons. Would that fact make you suspect that the loading mechanism is overfilling the cars? Justify your answer.

(c) If the filling equipment is functioning properly, what is the probability that a random sample of 10 cars will have a mean ore of 70.7 tons or more? Show your work.

(d) Based on your answer in part (c), if a random sample of 10 cars had a mean ore weight of 70.7 tons, would you suspect that the loading mechanism was overfilling the cars? Justify your answer.

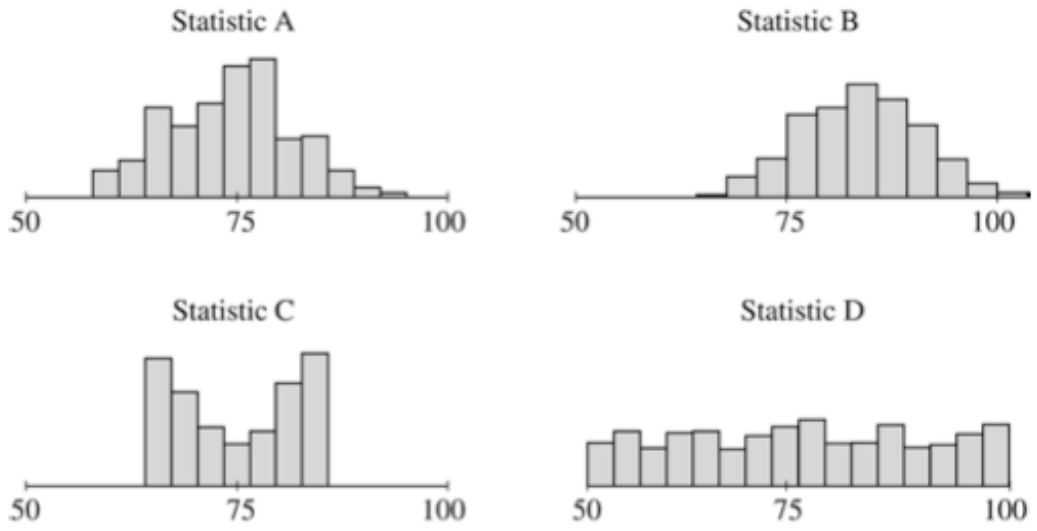
2. A tire manufacturer designed a new tread pattern for its all-weather tires. Repeated tests were conducted on cars of approximately the same weight traveling at 60 miles per hour. The tests showed that the new tread pattern enables the cars to stop completely in an average distance of 125 feet with a standard deviation of 6.5 feet and that the stopping distances are approximately normally distributed.

   (a) What is the 70th percentile of the distribution of stopping distances?

   (b) What is the probability that at least 2 cars out of 5 randomly selected cars in the study will stop in a distance that is greater than the distance calculated in part (a) ?

   (c) What is the probability that a randomly selected sample of 5 cars in the study will have a mean stopping distance of at least 130 feet?

2. Four different statistics have been proposed as estimators of a population parameter. To investigate the behavior of these estimators, 500 random samples are selected from a known population and each statistic is calculated for each sample. The true value of the population parameter is 75. The graphs below show the distribution of values for each statistic.



(a) Which of the statistics appear to be unbiased estimators of the population parameter? How can you tell?

(b) Which of statistics A or B would be a better estimator of the population parameter? Explain your choice.

(c) Which of statistics C or D would be a better estimator of the population parameter? Explain your choice.

ANSWERS: AP Statistics Exam Practice: Sampling Distributions

2010 #2 (Radio Station)

Part (a):

The sampling distribution of the sample mean song length has mean $\mu_{\bar{X}} = \mu = 3.9$ minutes and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1.1}{\sqrt{40}} \approx 0.174$ minutes. The central limit theorem (CLT) applies in this case because the sample size ($n = 40$) is fairly large, especially with the population of song lengths having a roughly symmetric distribution. Thus, the sampling distribution of the sample mean song length is approximately normal.

Part (b):

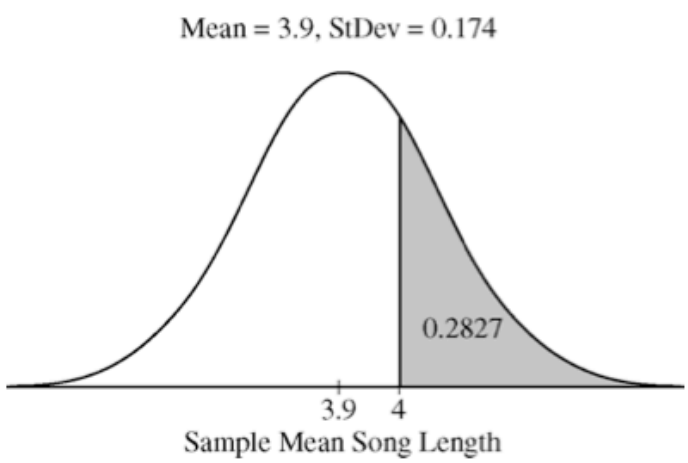
The probability that the total airtime of 40 randomly selected songs exceeds the available time (that is, the probability that the total airtime of 40 randomly selected songs is greater than 160 minutes) is equivalent to the probability that the sample mean length of the 40 songs is greater than $\frac{160}{40} = 4.0$ minutes.

According to part (a), the distribution of the sample mean length $\bar{X}$ is approximately normal. Therefore,

$$P(\bar{X} > 4.0) \approx P\left(Z > \frac{4.0 - 3.9}{0.174}\right) = P(Z > 0.57) = 1 - 0.7157 = 0.2843.$$

(The calculator gives the answer as 0.2827.)

The approximate sampling distribution of the sample mean song length and the desired probability are displayed below.



Part (b) (*alternative*):

An equivalent approach is to note that the sampling distribution of the total airtime, T, for the 40 songs is approximately normal, with mean $40(3.9) = 156$ minutes and standard deviation $\sqrt{40}(1.1) \approx 6.96$ minutes. The z-score for a total airtime of 160 minutes is then $z = \frac{160 - 156}{6.96} \approx 0.57$, and the calculation proceeds as above.

2004B #3 (Trains Carrying Bauxite)

Let X = weight of ore in a randomly selected car.

Part (a):

$$P(X > 70.7) = P\left(Z > \frac{70.7 - 70}{.9}\right) = P(Z > 0.78) = 0.2177$$

Part (b):

No. Approximately 22% of the cars will have ore weights of 70.7 or greater when the filling equipment is working properly, so a car that was filled with 70.7 tons of ore would not be an unusual occurrence.

Part (c):

$$P(\bar{X} > 70.7) = P\left(Z > \frac{70.7 - 70}{\frac{0.9}{\sqrt{10}}}\right) = P\left(Z > \frac{0.7}{0.285}\right) = P(Z > 2.46) = 0.0069$$

Part (d):

Yes, we would suspect that the filling mechanism is overfilling. If it is working properly, the probability that the mean weight of the ore in 10 randomly selected cars is 70.7 or greater is 0.0069, which is very small.

Note: To receive complete credit for part (a) or part (c) students must show how the probability is computed. Since part (a) and part (c) involve different normal distributions, it is important to identify which normal distribution is used in each part. As shown above, this could be done by displaying a probability statement containing the mean and standard deviation for the appropriate normal distribution. It could be done in other ways, such as listing the mean and standard deviation and displaying an appropriate graph.

2009 #2 (Tire Manufacturer)

Part (a):

Let X denote the stopping distance of a car with new tread tires where X is normally distributed with a mean of 125 feet and a standard deviation of 6.5 feet. The z-score corresponding to a cumulative probability of 70 percent is $z = 0.52$. Thus, the 70th percentile value can be computed as:

$$x = \mu_X + z\sigma_X = 125 + 0.52(6.5) = 128.4 \text{ feet.}$$

Part (b):

From part (a), it was found that a stopping distance of 128.4 feet has a cumulative probability of 0.70. Thus the probability of a stopping distance greater than 128.4 is $1 - 0.70 = 0.30$.

Let Y denote the number of cars with the new tread pattern out of five cars that stop in a distance greater than 128.4 feet. Y is a binomial random variable with $n = 5$ and $p = 0.30$.

$$P(Y \geq 2) = 1 - P(Y \leq 1) = 1 - \left[\binom{5}{0}(0.3)^0(0.7)^5 + \binom{5}{1}(0.3)^1(0.7)^4\right]$$

$$= 1 - 0.5282 = 0.4718.$$

Part (c):

Let $\bar{X}$ denote the mean of the stopping distances of five randomly selected cars. All tires have the new tread pattern. Because the stopping distance for each of the five cars has a normal distribution, the distribution of $\bar{X}$ is normal with a mean of 125 feet and a standard deviation of $\frac{6.5}{\sqrt{5}} = 2.91$ feet. Thus,

$$P(\bar{X} > 130) = P\left(Z > \frac{130 - 125}{6.5/\sqrt{5}}\right) \approx P(Z > 1.72) = 0.0427.$$

2008B #2 (Four Histograms)

a) A, C and D are unbiased since their means are all ≈ 75.
b) A since it is an unbiased estimator (its mean is ≈75). Statistic B is biased (its mean is ≈85)
c) C is better. Both C and D are unbiased (centered ≈75), but C has less variability than D.

AP Exam Practice: Inference

2006B #2

A large company has two shifts—a day shift and a night shift. Parts produced by the two shifts must meet the same specifications. The manager of the company believes that there is a difference in the proportions of parts produced within specifications by the two shifts. To investigate this belief, random samples of parts that were produced on each of these shifts were selected. For the day shift, 188 of its 200 selected parts met specifications. For the night shift, 180 of its 200 selected parts met specifications.

(a) Use a 96 percent confidence interval to estimate the difference in the proportions of parts produced within specifications by the two shifts.

(b) Based only on this confidence interval, do you think that the difference in the proportions of parts produced within specifications by the two shifts is significantly different from 0 ? Justify your answer.

2005B #6

6. Regulations require that product labels on containers of food that are available for sale to the public accurately state the amount of food in those containers. Specifically, if milk containers are labeled to have 128 fluid ounces and the mean number of fluid ounces of milk in the containers is at least 128, the milk processor is considered to be in compliance with the regulations. The filling machines can be set to the labeled amount. Variability in the filling process causes the actual contents of milk containers to be normally distributed. A random sample of 12 containers of milk was drawn from the milk processing line in a plant, and the amount of milk in each container was recorded.

   (a) The sample mean and standard deviation of this sample of 12 containers of milk were 127.2 ounces and 2.1 ounces, respectively. Is there sufficient evidence to conclude that the packaging plant is <u>not</u> in compliance with the regulations? Provide statistical justification for your answer.

2005 #4

Some boxes of a certain brand of breakfast cereal include a voucher for a free video rental inside the box. The company that makes the cereal claims that a voucher can be found in 20 percent of the boxes. However, based on their experiences eating this cereal at home, a group of students believes that the proportion of boxes with vouchers is less than 0.2. This group of students purchased 65 boxes of the cereal to investigate the company's claim. The students found a total of 11 vouchers for free video rentals in the 65 boxes.

Suppose it is reasonable to assume that the 65 boxes purchased by the students are a random sample of all boxes of this cereal. Based on this sample, is there support for the students' belief that the proportion of boxes with vouchers is less than 0.2 ? Provide statistical evidence to support your answer.

2004 #6:

6. A pharmaceutical company has developed a new drug to reduce cholesterol. A regulatory agency will recommend the new drug for use if there is convincing evidence that the mean reduction in cholesterol level after one month of use is more than 20 milligrams/deciliter (mg/dl), because a mean reduction of this magnitude would be greater than the mean reduction for the current most widely used drug.

   The pharmaceutical company collected data by giving the new drug to a random sample of 50 people from the population of people with high cholesterol. The reduction in cholesterol level after one month of use was recorded for each individual in the sample, resulting in a sample mean reduction and standard deviation of 24 mg/dl and 15 mg/dl, respectively.

   (a) The regulatory agency decides to use an interval estimate for the population mean reduction in cholesterol level for the new drug. Provide this 95 percent confidence interval. Be sure to interpret this interval.

Deeper question:

(b) Because the 95 percent confidence interval includes 20, the regulatory agency is not convinced that the new drug is better than the current best-seller. The pharmaceutical company tested the following hypotheses.

$$H_0: \mu = 20 \text{ versus } H_a: \mu > 20,$$

where μ represents the population mean reduction in cholesterol level for the new drug.

The test procedure resulted in a t-value of 1.89 and a p-value of 0.033. Because the p-value was less than 0.05, the company believes that there is convincing evidence that the mean reduction in cholesterol level for the new drug is more than 20. Explain why the confidence interval and the hypothesis test led to different conclusions.

2006 #4

Patients with heart-attack symptoms arrive at an emergency room either by ambulance or self-transportation provided by themselves, family, or friends. When a patient arrives at the emergency room, the time of arrival is recorded. The time when the patient's diagnostic treatment begins is also recorded.

An administrator of a large hospital wanted to determine whether the mean wait time (time between arrival and diagnostic treatment) for patients with heart-attack symptoms differs according to the mode of transportation. A random sample of 150 patients with heart-attack symptoms who had reported to the emergency room was selected. For each patient, the mode of transportation and wait time were recorded. Summary statistics for each mode of transportation are shown in the table below.

Mode of Transportation	Sample Size	Mean Wait Time (in minutes)	Standard Deviation of Wait Times (in minutes)
Ambulance	77	6.04	4.30
Self	73	8.30	5.16

(a) Use a 99 percent confidence interval to estimate the difference between the mean wait times for ambulance-transported patients and self-transported patients at this emergency room.

(b) Based only on this confidence interval, do you think the difference in the mean wait times is statistically significant? Justify your answer.

4. High cholesterol levels in people can be reduced by exercise, diet, and medication. Twenty middle-aged males with cholesterol readings between 220 and 240 milligrams per deciliter (mg/dL) of blood were randomly selected from the population of such male patients at a large local hospital. Ten of the 20 males were randomly assigned to group A, advised on appropriate exercise and diet, and also received a placebo. The other 10 males were assigned to group B, received the same advice on appropriate exercise and diet, but received a drug intended to reduce cholesterol instead of a placebo. After three months, posttreatment cholesterol readings were taken for all 20 males and compared to pretreatment cholesterol readings. The tables below give the reduction in cholesterol level (pretreatment reading minus posttreatment reading) for each male in the study.

Group A (placebo)

Reduction (in mg/dL)	2	19	8	4	12	8	17	7	24	1

Mean Reduction: 10.20 Standard Deviation of Reductions: 7.66

Group B (cholesterol drug)

Reduction (in mg/dL)	30	19	18	17	20	−4	23	10	9	22

Mean Reduction: 16.40 Standard Deviation of Reductions: 9.40

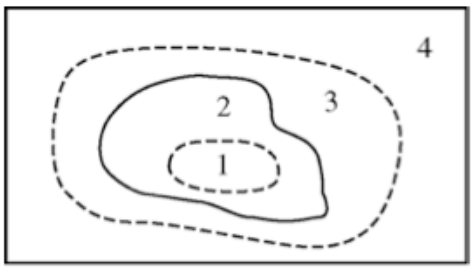
Do the data provide convincing evidence, at the $\alpha = 0.01$ level, that the cholesterol drug is effective in producing a reduction in mean cholesterol level beyond that produced by exercise and diet?

2008 #5

A study was conducted to determine where moose are found in a region containing a large burned area. A map of the study area was partitioned into the following four habitat types.

(1) Inside the burned area, not near the edge of the burned area,
(2) Inside the burned area, near the edge,
(3) Outside the burned area, near the edge, and
(4) Outside the burned area, not near the edge.

The figure below shows these four habitat types.



Note: Figure not drawn to scale.

The proportion of total acreage in each of the habitat types was determined for the study area. Using an aerial survey, moose locations were observed and classified into one of the four habitat types. The results are given in the table below.

Habitat Type	Proportion of Total Acreage	Number of Moose Observed
1	0.340	25
2	0.101	22
3	0.104	30
4	0.455	40
Total	1.000	117

(a) The researchers who are conducting the study expect the number of moose observed in a habitat type to be proportional to the amount of acreage of that type of habitat. Are the data consistent with this expectation? Conduct an appropriate statistical test to support your conclusion. Assume the conditions for inference are met.

(b) Relative to the proportion of total acreage, which habitat types did the moose seem to prefer? Explain.

2012 #5: Windmills

5. Windmills generate electricity by transferring energy from wind to a turbine. A study was conducted to examine the relationship between wind velocity in miles per hour (mph) and electricity production in amperes for one particular windmill. For the windmill, measurements were taken on twenty-five randomly selected days, and the computer output for the regression analysis for predicting electricity production based on wind velocity is given below. The regression model assumptions were checked and determined to be reasonable over the interval of wind speeds represented in the data, which were from 10 miles per hour to 40 miles per hour.

Predictor	Coef	SE Coef	T	P
Constant	0.137	0.126	1.09	0.289
Wind velocity	0.240	0.019	12.63	0.000

S = 0.237 R-Sq = 0.873 R-Sq (adj) = 0.868

(a) Use the computer output above to determine the equation of the least squares regression line. Identify all variables used in the equation.

(b) How much more electricity would the windmill be expected to produce on a day when the wind velocity is 25 mph than on a day when the wind velocity is 15 mph? Show how you arrived at your answer.

(c) What proportion of the variation in electricity production is explained by its linear relationship with wind velocity?

(d) Is there statistically convincing evidence that electricity production by the windmill is related to wind velocity? Explain.

2015 #2 Restaurant Discounts

2. To increase business, the owner of a restaurant is running a promotion in which a customer's bill can be randomly selected to receive a discount. When a customer's bill is printed, a program in the cash register randomly determines whether the customer will receive a discount on the bill. The program was written to generate a discount with a probability of 0.2, that is, giving 20 percent of the bills a discount in the long run. However, the owner is concerned that the program has a mistake that results in the program not generating the intended long-run proportion of 0.2.

The owner selected a random sample of bills and found that only 15 percent of them received discounts. A confidence interval for p, the proportion of bills that will receive a discount in the long run, is 0.15 ± 0.06. All conditions for inference were met.

(a) Consider the confidence interval 0.15 ± 0.06.

(i) Does the confidence interval provide convincing statistical evidence that the program is not working as intended? Justify your answer.

(ii) Does the confidence interval provide convincing statistical evidence that the program generates the discount with a probability of 0.2 ? Justify your answer.

A second random sample of bills was taken that was four times the size of the original sample. In the second sample 15 percent of the bills received the discount.

(b) Determine the value of the margin of error based on the second sample of bills that would be used to compute an interval for p with the same confidence level as that of the original interval.

(c) Based on the margin of error in part (b) that was obtained from the second sample, what do you conclude about whether the program is working as intended? Justify your answer.