**Investigating Statistical Studies**

I. The Effect of Rock Music on Memorization

What is the effect of listening to rock music on one’s ability to memorize? To investigate this, a class of 28 high school students was asked to participate in a study. Fourteen cards marked “M” and 14 cards marked “S” were placed in a paper bag. Each student in the class was asked to draw a card from the bag. Students drawing an “M” were placed in the music group and students drawing an “S” were placed in the silence group. Each student was given a list of 16 words to memorize. Students were allowed two minutes to study the word list, followed by a one-minute pause, and then given two minutes to write down as many words from the list as possible. Students in the music group were placed in a room with rock music playing throughout the 5 minutes. Students in the silence group were placed in a similar room, but with no music. The number of words that were remembered by each student was the response variable of interest. The table and graph below summarize the results. (See *The GAISE Report*, pgs. 54 – 55.)



 (The table above was retrieved from *The GAISE REPORT,* pg. 55.)

1. Based on the table and graph above, what conclusions about the difference in the distributions of the recall word score for these two groups can be drawn? Justify your answer.
2. The class wants to know if listening to music is related to a student’s retention of words. Which statement best describes the study?
3. This is an observational study with random assignment of treatments, and therefore, the class may conclude there is a cause-and-effect relationship between listening to music and memorization.
4. This is an observational study with random assignment of treatments, and therefore, the class will not be able to conclude there is a cause-and-effect relationship between listening to music and memorization.
5. This is an experiment with random assignment of treatments, and therefore, the class may conclude there is a cause-and-effect relationship between listening to music and memorization.
6. This is an experiment with random assignment of treatments, and therefore, the class will not be able to conclude there is a cause-and-effect relationship between listening to music and memorization.
7. Which of the following statements correctly summarizes to whom the findings of this study may be applied?
8. The findings apply to all people.
9. The findings apply to all high school students.
10. The findings apply only to the students in this particular high school.
11. The findings apply only to the 28 students in this class.
12. Which of the following statements provides the best support for your answer in question 3 above?
13. The treatments were applied randomly.
14. The sample of subjects for this study was not selected randomly.
15. A difference this big in the median scores for the two treatments may always be generalized back to the population of interest.
16. A cause-and-effect relationship may always be generalized back to the population of interest, but an association may not.

II. Forearm Length and Height

 The following example appears in *The GAISE Report*.

Relationships among various physical features, such as height versus arm span and neck size versus shoe size, can be the basis of many interesting questions for student investigation. If I were painting a picture of a person, how could I get the relative sizes of the body parts correct? This question prompted students to carry out an investigation of one of the possible relationships, that between forearm length (the distance from the elbow crease to the wrist) and height. The students responsible for the study sampled other students on which to make forearm and height measurements. Although the details of how the sample actually was selected are not clear, we will suppose that it is representative of students at the school and has the characteristics of a random sample… The data obtained by the students (in centimeters) are provided in Table 14. (The GAISE Report, pg. 81.)



1. Which of the statements below are true for this model? (Select all that apply.)
2. The residual plot is scattered with no discernable pattern, indicating that the linear model is a good fit to the data.
3. There is a pattern to the residual plot, indicating that the linear model is a good fit to the data.
4. There is a negative association between forearm length and height.
5. The correlation coefficient of 0.8 indicates that there is a strong, positive linear relationship between a person’s forearm length and height.
6. The model predicts a value close to the actual observed value for forearm lengths of 39.5 cm, 43.5 cm, and 46.5 cm.
7. Consider the value predicted by the model for a forearm length of 47 cm.
8. The model overpredicts the actual forearm length by less than 10 cm.
9. The model overpredicts the actual forearm length by more than 10 cm.
10. The model underpredicts the actual forearm length by less than 10 cm.
11. The model underpredicts the actual forearm length by more than 10 cm.
12. Was this an observational study or an experiment? Justify your answer.
13. The findings of this study can be generalized to what group? Justify your answer.

III. Music Preference (Adapted from *The GAISE Report*, pg. 71.)

What type of music do students at our high school prefer? To answer this question, Kiera and Karston created the following survey:

|  |
| --- |
| *Music Survey*1. *What kinds of music do you like?*

*Do you like country music? Yes or No**Do you like rap music? Yes or No**Do you like rock music? Yes or No*1. *Which of the following types of music do you like most? Select only one.*

*Country Rap/Hip Hop Rock* |

1. Which of the following is the best method for selecting a sample of 40 students if Kiera and Karston want to be able to generalize their findings to all students at their high school?
2. Randomly select a lunch period and select the first 50 people who enter the school cafeteria. These 50 students will be asked to take the survey.
3. Randomly select 2 classrooms and ask the first 25 students that enter each room to take the survey.
4. Get a list of all students at the high school. Write each student’s name on a slip of paper, place the slips in a paper bag, mix, and then draw out 50 slips. The students whose names are drawn are asked to take the survey.
5. Post the survey in the school newspaper and ask for student responses. The first 50 students to complete and turn in the survey will form the sample.
6. Kiera and Karston want to know if there is a relationship between students who like rap and students who like rock music. Circle the one phrase in each box that correctly completes the sentence.

$$\begin{matrix}\begin{matrix}an observational study\\ \end{matrix}\\an experiment\end{matrix}$$

Because the survey is , Kiera and

Karston conclude that there is a

$$\begin{matrix}\begin{matrix}will be able to\\ \end{matrix}\\will not be able to\end{matrix}$$

cause-and-effect relationship between liking rap and liking rock

music.

1. After surveying 50 students, Kiera and Karston recorded their results in the table shown below.



(The table above was retrieved from *The GAISE REPORT,* pg. 72.)

Let RAP = the event that a randomly selected student in this sample likes rap music.

Let ROCK = the event that a randomly selected student in this sample likes rock music.

 Which of the following statements correctly identifies if RAP and ROCK are independent or dependent events and provides the correct justification?

1. Liking rap music and liking rock music are dependent events because P(RAP $∩$ ROCK) ≠ P(RAP) · P(ROCK).
2. Liking rap music and liking rock music are dependent events because P(RAP | ROCK) ≠ P(ROCK | RAP).
3. Liking rap music is independent of liking rock music because P(RAP) = P(ROCK).
4. Liking rap music is independent of liking rock music because P(RAP | ROCK) = P(RAP).

IV. Radish Seedlings (Selected from *The GAISE Report*, pg. 75 - 79.)

Example 4: An Experiment on the Effect of Light on the Growth of Radish Seedlings

What is the effect of different durations of light and dark on the growth of radish seedlings? This question was posed to a class of biology students who then set about designing and carrying out an experiment to investigate the question. All possible relative durations of light to dark cannot possibly be investigated in one experiment, so the students decide to focus the question on three treatments: 24 hours of light, 12 hours of light and 12 hours of darkness, and 24 hours of darkness. This covers the extreme cases and one in the middle.

With the help of a teacher, the class decided to use plastic bags as growth chambers. The plastic bags would permit the students to observe and measure the germination of the seeds without disturbing them. Two layers of moist paper towel were put into a disposable plastic bag, with a line stapled about ½ of the way from the bottom of the bag (see Figure 30) to hold the paper towel in place and to provide a seam to hold the radish seeds.



Although three growth chambers would be sufficient to examine the three treatments, this class made four growth chambers, with one designated for the 24 hours of light treatment, one for the 12 hours of light and 12 hours of darkness treatment, and two for the 24 hours of darkness treatment. One hundred twenty seeds were available for the study. Thirty of the seeds were chosen at random and placed along the stapled seam of the 24 hours of light bag. Thirty seeds were then chosen at random from the remaining 90 seeds and placed in the 12 hours of light and 12 hours of darkness bag. Finally, 30 of the remaining 60 seeds were chosen at random and placed in one of the 24 hours of darkness bags. The final 30 seeds were placed in the other 24 hours of darkness bag. After three days, the lengths of radish seedlings for the germinating seeds were measured and recorded. These data are provided in Table 12; the measurements are in millimeters. Notice that not all the seeds in each group germinated. (*The GAISE Report*, pgs. 75 – 76.)



1. What is the statistical question being considered in this study?
2. What is the population of interest?
3. How was the sample for this study selected? Is it representative of the population?
4. Were the radish seedlings assigned randomly to a treatment? If so, what is the treatment(s) and how were the seedlings assigned to the treatment(s)?
5. Is this an observational study or an experiment? Explain.

Below are a graph and summary statistics of the radish seedling lengths by treatment. 

 Treatment 1: 24 hours of light

 Treatment 2: 12 hours of light and 12 hours of darkness

 Treatment 3: 24 hours of darkness

1. Based on the table and graph above, what conclusions about the difference in the distributions of the radish seedling lengths for these three treatment groups can be drawn? Justify your answer.
2. Based on the design of this statistical study, is it possible to conclude that there is a cause-and-effect relationship, an association, or neither between the amount of light and the amount of growth in radish seedlings? Justify your answer.
3. The findings for this study can be generalized to what group?

Is there evidence that the 12 hours of light and 12 hours of dark (Treatment 2) group has a significantly higher mean than the 24 hours of light (Treatment 1) group? Is there evidence that the 24 hours of dark (Treatment 3) group has a significantly higher mean than the 12 hours of light and 12 hours of dark (Treatment 2) group? Based on the boxplots and the summary statistics, it is clear that the sample means differ. *Are these differences large enough to rule out chance variation as a possible explanation for the observed difference?* The Treatment 2 mean is 6.2 mm larger than the Treatment 1 mean. If there is no real difference between the two treatments in terms of their effect on seedling growth, then the observed difference must be due to the random assignment of seeds to the bags; that is, one bag was simply lucky enough to get a preponderance of good and lively seeds. But, if a difference this large (6.2 mm) is likely to be the result of randomization alone, then we should see differences of this magnitude quite often if we repeatedly rerandomize the measurements and calculate a new difference in observed means…



This dotplot was produced by mixing the growth measurements from Treatments 1 and 2 together, randomly splitting them into two groups of 28 measurements, recording the difference in means for the two groups, and repeating the process 200 times. (*The GAISE Report*, pgs. 77 - 78.)

1. Based on the results of this simulation, is there evidence that there is a statistically significant difference between the means for Treatments 1 and 2?
2. Yes, because an outcome of 6.2 or higher is an unlikely outcome.
3. Yes, because most of the outcomes are below a difference of 6.2.
4. No, because the distribution is approximately normal.
5. No, because at least 10 % of the trials had an outcome of 6.2.
6. Suppose the difference in the two treatment means in the experiment had been 2 mm. Do the results of the simulation provide evidence that a difference of 2mm between the means for Treatments 1 and 2 are significant?
7. Yes, because 2 mm was the most frequent outcome.
8. Yes, because about 16% of the outcomes were 2mm or greater.
9. No, because the distribution is approximately normal.
10. No, because an outcome of 2mm or more is not unusual.

In a comparison of the means for Treatments 2 and 3, the same procedure is used, except that the combined measurements are split into groups of 28 and 58 each time. (*The GAISE Report*, pg. 78.)



1. What was the observed difference in the means for Treatments 2 and 3 in the actual radish seedling experiment?
2. The dotplot above shows the differences in the mean for 200 trials using the values for Treatments 2 and 3. Based on the results of this simulation, is there evidence that there is a statistically significant difference between the means for Treatments 2 and 3? Explain your reasoning.

Teacher Notes

The four tasks in this worksheet are all based on examples found in *The GAISE Report*. Each task focuses on at least one other statistics standard in addition to the primary content standards listed below. You may wish to use these tasks separately for class discussion, review, or assessment. Some of the questions in these tasks were modeled after questions found on the PARCC Practice Tests for Algebra II.

Primary Content Standards

|  |  |  |  |
| --- | --- | --- | --- |
| Supporting | S | IC.1 | Understand and evaluate random processes underlying statistical experiments. Understand statistics as a process for making inferences about population parameters based on a random sample from that population.\* |
| ◼Major | S | IC.3 | Make inferences and justify conclusions from sample surveys, experiments, and observational studies. Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.\* |
| ◼Major | S | IC.6 | Make inferences and justify conclusions from sample surveys, experiments, and observational studies. Evaluate reports based on data.\* |

Additional Content Standards

|  |  |  |  |
| --- | --- | --- | --- |
| Supporting | S | ID.3 | Summarize, represent, and interpret data on a single count or measurement variable. Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).\* |
| Supporting | S | ID.6b | Informally assess the fit of a function by plotting and analyzing residuals.\* |
| Supporting | S | ID.8 | Interpret linear models. Compute (using technology) and interpret the correlation coefficient of a linear fit.\* |
| ◼Major | S | IC.5 | Make inferences and justify conclusions from sample surveys, experiments, and observational studies. Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.\* |
| 🞇Additional | S | CP.2 - 4 | Understand independence and conditional probability and use them to interpret data. 2. Understand that two events A and B are independent if the probability of A and B occurring together is the product of their probabilities, and use this characterization to determine if they are independent.\*3. Understand the conditional probability of A given B as P(A and B)/P(B), and interpret independence of A and B as saying that the conditional probability of A given B is the same as the probability of A, and the conditional probability of B given A is the same as the probability of B.\*4. Construct and interpret two-way frequency tables of data when two categories are associated with each object being classified. Use the two-way table as a sample space to decide if events are independent and to approximate conditional probabilities. For example, collect data from a random sample of students in your school on their favorite subject among math, science, and English. Estimate the probability that a randomly selected student from your school will favor science given that the student is in tenth grade. Do the same for other subjects and compare the results.\* |

Primary Math Practices

MP 1 Make sense of problems and persevere in solving them.

MP 2 Reason abstractly and quantitatively.

MP 4 Model with mathematics.

MP 6 Attend to precision.

References

Franklin, C., Kader, G., Mewborn, J. M., Peck, R., Perry, M. & Schaeffer, R. (2007) *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework*. Alexandria, VA: American Statistical Association.

Key

I. The Effect of Rock Music on Memorization

In addition to focusing on sampling and statistical studies, this task also addresses S-ID.3. (Interpret differences in shape, center and spread.)

1. The group that completed the memory test in silence did better than the group that listened to rock music during the memory test. The silence group had a larger median number of words recalled (10 words) than the group that listened to music (7 words). The music group had a larger range of with scores as low as 3 words up to 15 words. It is also worth noting that more that 50% of the students in the silence group had scores above 75% of the music group. And 100% of the silence group had scores above the bottom 25% of the music group.

2. C – The conclusion is that listening to rock music results in (causes) lower retention of words for this group of students.

3. D

4. B

II. Forearm Length and Height

In addition to focusing on sampling and statistical studies, this task also addresses S-ID.6b (residuals) and S-ID.8 (interpreting the correlation coefficient).

1. A, D, E
2. C

3. This was an observational study. In an experiment, the researcher manipulates factors of interest by assigning subjects to treatments. In this statistical study, researchers did not manipulate any factors, but instead, simply recorded the lengths of forearms and height.

4. The introductory paragraph for this problem states that while we are not provided with information about whether the sample of participants was selected randomly, we may assume that the sample is representative of all students at this high school. We can generalize our findings (students with larger forearms tend to be taller) for this study to the population of all students at the high school.

III. Music Preference

In addition to focusing on sampling and statistical studies, this task also addresses S-CP.2 - 4 (determining independence of two events).

1. C
2. an observational study; will not be able to
3. A

IV. Radish Seedlings

In addition to focusing on sampling and statistical studies, this task also addresses S-ID.3 (interpret differences in shape, center and spread) and S-IC.5 (use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant).

1. What is the effect of different durations of light and dark on the growth of radish seedlings?
2. The population of interest is likely to be all radish seedlings.
3. The information shared for this study says, “One hundred twenty seeds were available for the study.” It is not clear how these 120 seeds were selected.
4. The radish seeds were assigned randomly to the treatment groups. First, 30 seeds were randomly selected from the sample of 120 seeds and assigned to the 24 hours of light treatment group. Another 30 seeds were then randomly selected from the remaining 90 seeds in the sample and assigned to the 12 hours of light and 12 hours of darkness treatment group. Next, 30 seeds were randomly selected and assigned to the first 24 hours of darkness treatment group. The remaining 30 seeds were assigned to a second 24 hours of darkness treatment group.
5. This is an experiment, because the students assigned the radish seeds to treatments.
6. There is a difference in the mean lengths of the radish seedlings for the three groups. The radish seeds placed in 24 hours of darkness (Treatment 3) had the greatest average growth of 21.86 mm. The seeds placed in 12 hours of light and 12 hours of darkness (Treatment 2) had the second greatest average growth of 15.82 mm. The radish seeds receiving 24 hours of light (Treatment 1) had an average growth of 9.64 mm. Treatment 3 group also had the greatest variability in growth. Based on the boxplots, it appears that approximately 50% of the radish seedlings in the Treatment 3 group grew more than 75% of the seedlings in the Treatment 2 group. Approximately 75% of the seedlings in the Treatment 2 group were longer than 75% of the seedlings in the Treatment 1 group.
7. Since this was a controlled, randomized experiment (factors such as containers and moisture were controlled and kept the same for all three treatments and treatments were assigned randomly), we can conclude there is a cause-and-effect relationship between the amount of light exposure and the amount of growth for these radish seedlings. Darkness caused the radish seedlings in this treatment group to grow to a greater length.
8. As it is unclear how the original sample of 120 radish seeds was selected, we cannot generalize the results of this study to a larger population. The conclusion mentioned in number 7 above only applies to these 120 radish seeds.
9. A
10. D
11. The observed difference in the means from Treatments 2 and 3 is 6.02 mm (or about 6 mm).
12. The observed difference of 6 mm was exceeded only one time out of 200 trials (see Figure 33), giving extremely strong evidence of a statistically significant difference between the means for Treatments 2 and 3.

In summary, the three treatment groups show statistically significant differences in mean growth that cannot reasonably be explained by the random assignment of seeds to the bags. This gives us convincing evidence of a treatment effect—the more hours of darkness, the greater the growth of the seedling, at least for these three periods of light versus darkness.

 Some Additional Information from *The GAISE Report* on this study:

Students should be encouraged to delve more deeply into the interpretation, relating it to what is known about the phenomenon or issue under study. Why do the seedlings grow faster in the dark? Here is an explanation from a biology teacher. It seems to be an adaptation of plants to get the seedlings from the dark (under ground) where they germinate into the light (above ground) as quickly as possible. Obviously, the seedling cannot photosynthesize in the dark and is using up the energy stored in the seed to power the growth. Once the seedling is exposed to light, it shifts its energy away from growing in length to producing chlorophyll and increasing the size of its leaves. These changes allow the plant to become self-sufficient and begin producing its own food. Even though the growth in length of the stem slows, the growth in diameter of the stem increases and the size of the leaves increases. Seedlings that continue to grow in the dark are spindly and yellow, with small yellow leaves. Seedlings grown in the light are a rich, green color with large, thick leaves and short stems. (*The GAISE Report*, pgs. 78 – 79.)