This event will work in 2 parts.

The first part is similar to the old rules of the first part of experimental design. You will be given the materials list with images and you will only design the experiment.
At the end of the test, please remember to email scienceolympiadiscool@gmail.com the procedure diagrams within 15 minutes of the block ends. Please include your school name and team number in the email.

The second part is a test where you will be tested on data analysis as well as experimental design procedures and best practices.

Both parts will be given at the start and turned in together at the end of 50 minutes.

If you have any urgent questions during the exam you may contact the chat.

If you have any feedback for the test, you may email the same email.

Good luck!!!

Test Writers: Peter Zhu, Anson Pham

PART 1: Experimental Design

Topic Area: Fluid Flow

A fluid is a substance that has no fixed shape and yields easily to external pressure.

You may assume you are conducting the experiment outside and the water from the tank can just drain onto the floor.

Materials:

(item 1) **1** clear, circular cylinder metal tank with open top (no lid)

- Initially filled halfway with water open to the atmosphere on Earth
- Clear so that measuring the height of the water in the tank is easy
- No holes are punched initially
- Diameter: 3 feet
- Height of cylinder: 5 feet
- Starting height of water: 2.5 feet

(item 2) **1** metal hole puncher

- Can punch holes of diameter 3/32", 1/8", 5/32", 3/16", 7/32", 1/4", and 9/32"
- Can be operated by hand
- The holes can be punched anywhere on the metal tank (sides or bottom at any height)

(item 3) **5** rubber hole plugs

- Can be used for any size hole to stop the flow of water from a created hole with the hole puncher.

(item 4) **10** plastic bags filled with 1 cup of salt each

(item 5) an unlimited supply of water to fill the tank back up from a hose

Measurement Equipment:

1 timer

- Measures to the nearest 0.01 seconds

1 linear measuring device

- Measures to the nearest 0.1 inches

Design an experiment that addresses the topic area of fluid flow. You must use at least 2 materials not including the measurement equipment on the list above. Your design report will include a statement of the problem, hypothesis, independent variables, dependent variables, controlled variables, constants, experimental control, materials list of the materials you used, and procedure with diagrams.

Images:

Item 1 (shown with holes punched on the side and bottom - initially there is no hole punched)
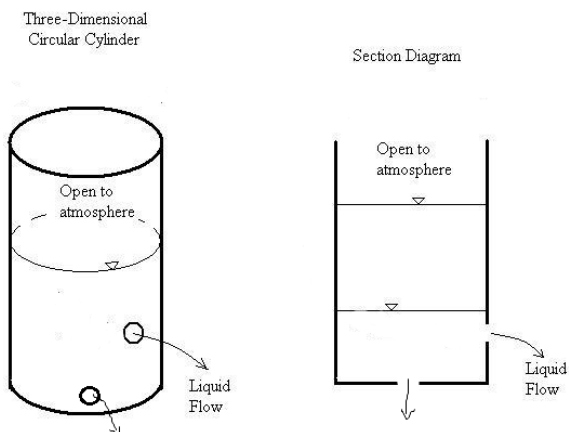
Item 2

Item 3

Item 4

shutterstock.com • 1424118269

Item 5

1. (2.00 pts)   A. Statement of Problem

Expected Answer:

2. (6.00 pts)   B. Hypothesis

Expected Answer:

3. (8.00 pts)   C. Independent Variable

Expected Answer:

4. (4.00 pts)   C. Dependent Variable

Expected Answer:

5. (4.00 pts)   C. Controlled Variables

Expected Answer:

6. (4.00 pts)   C. Constants

Expected Answer:

7. (4.00 pts)   D. Experimental Control (Standard of Comparison)

Expected Answer:

8. (4.00 pts)   E. Materials

Expected Answer:

9. (14.00 pts)   F. Procedure and Set-up Diagrams

Expected Answer:

PART 2: Data Analysis

You are taking the reading section of the SAT. Unfortunately, you forgot how to read so you have to guess on everything at random. Assume there are 100 questions and each question is a multiple choice question with 5 choices of which only 1 is correct. The grading scheme is as follows: you get 4 points for a correct answer, and you get -1 points for an incorrect answer. You guess on all the questions.

10. (0.50 pts)   What is the expected number of questions you answer correctly?

20

11. (0.50 pts)   What is your expected number of points you get on the test?

0

12. (1.00 pts)
What is the distribution for the number of questions you answer correct? For example, if you think this is a normal distribution with mean 0 and variance 1, write Normal ~(0,1)

Expected Answer: Binomial ~(100, 0.2)

13. (1.50 pts)   What is the standard deviation of the number of questions you answer correctly?

4

A student just performed an experiment and got the following data for repeated trials at some level of the experiment:

0.12 inches, 0.11 inches, 0.12 inches, 0.16 inches, 0.08 inches, 0.12 inches

For the answers below, do NOT include units in the box answer. If you think the answer is 32.25 inches, write 32.25.

14. (1.00 pts)   Assuming proper use of significant digits in measurement here, what is the nearest precision in inches for certain digits this device has?

0.1

15. (0.50 pts)   Calculate the mean. Round to the nearest 4 decimal places.

0.1183

16. (0.50 pts)   Calculate the median. Round to the nearest 2 decimal places.

0.12

17. (0.50 pts)   Calculate the mode. Round to the nearest 2 decimal places.

0.12

18. (1.00 pts)   Calculate the sample standard deviation. Round to the nearest 4 decimal places.

0.0256

19. (0.50 pts)   Calculate the range. Round to the nearest 2 decimal places.

0.08

20. (3.00 pts)   Does an outlier exist? Show your work.

Expected Answer: Yes by 1.5 IQR rules it does! 25th Percentile: 0.11 50th Percentile: 0.12 75th Percentile: 0.12 Interquartile Range: 0.01

Later on that day, the student realized that s/he accidentally added 0.05 inches to all the measurements. :'(

21. (0.50 pts)   This error changes the median value of the data.

◉ True   ○ False

22. (0.50 pts)   This error changes the standard deviation of the data.

○ True   ● False

Sig Figs!

23. (0.50 pts)   How many sig figs does the number 100 have?

1

24. (0.50 pts)   How many sig figs does the number 100.0 have?

4

25. (0.50 pts)   How many sig figs does the number 1010 have?

3

Sampling

The presidential election is coming up! You are working with RealClearPolitics to try and find the proportion of people that approve of a candidate in the state of California.

You are considering many ways to sample the population of eligible voters and you have proposed the following methods. Name 2 things wrong with each method below.

26. (1.00 pts)
You go to your local grocery store on a Monday and ask every tenth person that exits the store if they approve of the candidate. You get there in the morning and stay there until you collect 100 responses. If the person does not respond, you ask the next person that leaves the store and make gaps of every 10 people before asking another person.

Expected Answer: only considers local population, only shoppers, only in the morning, non response bias

27. (1.00 pts)
You randomly select 100 addresses in the state of California from a registered voter database and send them mail from USPS that asks if they would vote for the candidate. The respondents are supposed to mail you back, and based on the number of responses you get back you can send more surveys out until you reach 100 responses. You calculate the proportion of people that approve of the candidate from your responses you get.

Expected Answer: respondents have to mail you back (could choose not to respond), mail could get lost, approval rating does not neccesarily mean if they would vote for the candidate

28. (1.00 pts)
You want many responses. You make a post on Reddit that asks Reddit users to vote on if they approve of, are neutral, or disapprove of the candidate. Once you hit 10,000 votes, you close the poll.

Expected Answer: should not include a neutral option, non response bias, people can vote twice, doesnt have to be a voter, only active a certain part of the day which may exclude sertain populations

Precision, Accuracy, and Uncertainty

29. (0.50 pts)   It is possible for a measurement to be precise but not accurate.

◉ True   ○ False

30. (0.50 pts)   It is possible for a measurement to be accurate but not precise.

◉ True   ○ False

31. (0.50 pts)   If a measurement is accurate, it must always be precise.

○ True   ◉ False

32. (0.50 pts)   If a measurement is precise, it must be accurate.

○ True   ◉ False

33. (2.00 pts)
You are measuring the actual power consumption of your hairdryer whose rated power is known to be 2000 [W]. Given that a digital power meter you are going to use for the measurement is being affected by external electromagnetic interference (EMI), generating a precision error of 40 [W], and inherently produces an systematic error of 30 [W] due to bad calibration, what is the best estimate of the uncertainty of this measurement?

Expected Answer: 50 W is the answer, 30 + 40/2 = 50. Give 1 point for 70 W.

34. (0.50 pts)   Which term corresponds to a measure of the extent of variation among multiple measurements of the same quantity?

○ A) Accuracy

◉ B) Precision

○ C) Uncertainty

A university wants to study the type of students enrolled in its large classes, defined as classes with enrollments of 300 or more. There are 50 such classes. From each of these classes, one enrolled student is chosen uniformly at random to take part in a survey. You can assume that the selection from each class is performed independently of the selections in the other classes.

35. (1.00 pts)   There are students in the population of interest who are not in the sampling frame.

○ True   ● False

36. (1.00 pts)   There are students in the sampling frame who are not in the population of interest.

○ True   ● False

37. (1.00 pts)   The method of sampling produces a simple random sample of students enrolled in the big classes.

○ True   ● False

38. (1.00 pts)   Because a student is chosen from each class, all students in the big classes have the same chance of being selected.

○ True   ● False

Graphing things is cool!

39. (0.50 pts)   All graphs should have a title with labels and units for the x axis and y axis.

● True   ○ False

40. (0.50 pts)   The visual variable "size" can be encoded in many ways. Which of the following encodings makes it easiest for viewers to compare sizes in most cases?

○ A) Angle
○ B) Area
● C) Length
○ D) Volume

41. (1.00 pts)
You want to see the distribution of M&M colors in a bag - and you especially want to see the difference in distribution between different colors. What kind of graph/visualization should you use?

Expected Answer: 1 point for bar graph, 0.5 point for pie chart

42. (1.50 pts)   The independent variable goes on the x axis.

● True   ○ False

43. (2.00 pts)
Select all that apply - write your answers in the box below.

In a box plot, you can see which of the following: mean, median, mode, standard deviation, variance, 25th percentile, 75th percentile, 95th percentile, 100th percentile

Expected Answer: +0.5 correct median, 25, 75 100 -0.5 incorrect everything else

Bob the Builder is working as a data scientist studying newspapers. He decides to collect data about different articles on the front page of various newspapers. He records the date, newspaper source, headline font size, article length in words, topic area, and author.

For each of the following questions below, choose the visualization method that most makes sense:

44. (1.00 pts) Is there an association between headline font size and article length?

- ○ A) Histogram
- ○ B) Line Graph
- ○ C) Bar Chart
- ● D) Scatter Plot

45. (1.00 pts) Are there more long articles or short articles? How long do they tend to be?

- ● A) Histogram
- ○ B) Line Graph
- ○ C) Bar Chart
- ○ D) Scatter Plot

46. (1.00 pts) How does the average length of articles written on each day change over time?

- ○ A) Histogram
- ● B) Line Graph
- ○ C) Bar Chart
- ○ D) Scatter Plot

47. (1.00 pts) What is the distribution of article length between the New York Times and the Washington Post?

- ● A) Histogram
- ○ B) Line Graph
- ○ C) Bar Chart
- ○ D) Scatter Plot

Errors!!!

48. (2.00 pts)
Peter, a silly Wright Stuff builder, is testing a plane in the gym. He tries to tune the plane to increase the time aloft in the air. He decides to increase the angle of attack as well as put more winds on the rubber since in his last trial, the plane did not go high enough. He does the same procedures as he had before with the same winding techniques and propellor. He also noticed from last time that the plane got too close to the wall, so he moved a little farther away from the wall at launch. The conditions in the gym remain constant throughout his trials. After every flight, he took detailed notes of the process he followed to get the exact time he got. Was there any experimental design flaw that Peter made? If so, what was it?

Expected Answer: Changing more than one variable at a time.

49. (1.00 pts)   If a chemist consistently reads a graduated cylinder from the top (instead of directly looking at the meniscus), is this a random or systematic error?

○ A) Random Error
● B) Systematic Error

50. (1.00 pts)   If there are uncompensated vibrations on a table while trying to measure the speed of a rolling ball on that table, would this be a random or systematic error?

● A) Random Error
○ B) Systematic Error

51. (2.00 pts)   Explain why it is important to have multiple trials in an experiment.

Expected Answer: reduction in standard error, more certainty of measurements over many measurements, Law of Large Numbers

52. (1.00 pts)   The second to last digit in uncertainty determines the number of significant figures you should report.

○ True   ● False

53. (1.00 pts)   Estimates of uncertainty are always rounded up.

● True   ○ False

54. (1.00 pts)   When doing error propagation analysis, we always assume that each measurement is independent of one another.

● True   ○ False

The following two graphs (A and B) show monthly temperature and rainfall trends on the planet Naboo, where summer is from December to February. You can assume that both graphs represent data from a typical year on Naboo.
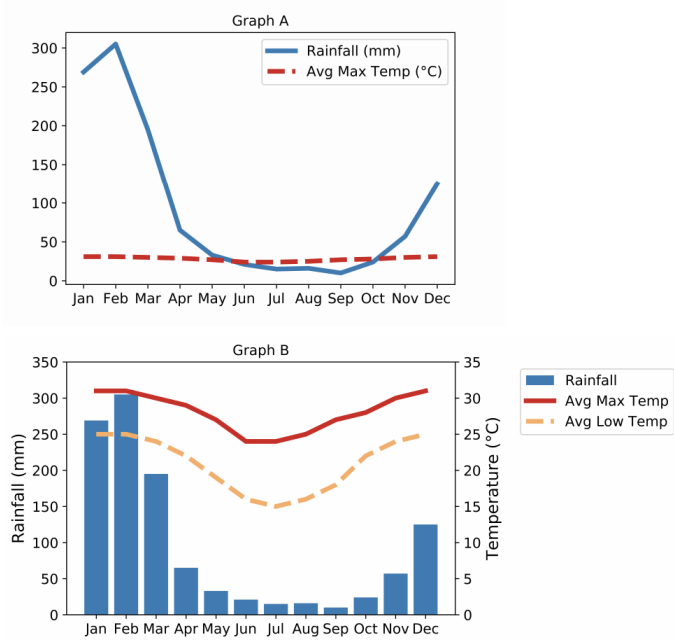
Graph A

Graph B

55. (2.00 pts)   Describe the trend you see in the data for the temperature on Naboo over the year. Be as specific as possible.

Expected Answer: Includes numbers and direction of trend.

56. (2.00 pts)   In Graph A, the average max temperature is almost constant across the months. In Graph B, it has a clear dip during the winter months. Why is this the case?

Expected Answer: Graph B is OK but Graph A has a problem because it uses the same vertical scale for both variables.

57. (1.00 pts)   We can conclude that the temperature on Naboo is always above 10°C.

○ True   ● False

58. (1.00 pts)   Select the reason for using a bar chart to represent rainfall and lines to represent temperature in Graph B.

○ A) Rainfall is measured in length (mm) and so it has to be encoded by length, but many different encodings work for temperature.

● B) Using a bar chart for rainfall and line plots for temperature is easier to read than a double bar chart for temperature with a line plot for rainfall.

○ C) It was just the choice of the creator; it would have been equally good to use a double bar chart for the mean minimum and maximum temperatures and a line plot for rainfall.

Random Assortment of Problems Below

59. (0.50 pts)   If you have an independent variable at 3 levels and you are performing 10 trials at each level. How many entries do you need to put in the data table?

```
30
```

60. (1.50 pts)

A statistician notices that the crime rate in January in the Bay Area is the lowest and every month after January the crime rate slightly increases until December and falls back down again the next year. This pattern has continued for the last 20 years. Would it be reasonable to say that the month of the year causes differences in the crime rate? Why?

Expected Answer: No there was no experiment done this is only correlation.

61. (2.00 pts)   Which one is likely to have the higher sample standard deviation after many trials?

The average value of rolling a dice 10 times, or the average value of rolling a dice 100 times. Explain your answer.

Expected Answer:

62. (1.00 pts)   A Yelp rating is a number of stars for a restaurant between 1 and 5. It is defined as a qualitative ordinal value rather than a quantitative discrete variable. Why is this?

Expected Answer: The difference between 1 and 2 stars is the not the same as the difference between 3 and 4 stars.

63. (2.00 pts)   What type of noise has a fairly flat power spectrum?

(Mark **ALL** correct answers)
- ☐ A) Red noise
- ☐ B) Grey noise
- ☑ C) White noise
- ☐ D) Black noise
- ☐ E) None of the above

I love linear regression!

64. (1.00 pts)   Suppose you are trying to estimate the value of a numerical variable y based on a linear function of a related numerical variable x.

The best among all horizontal lines (based on squared loss) has the equation estimate of y = _____

- ○ A) Average of x
- ● B) Average of y
- ○ C) Standard Deviation of x
- ○ D) Standard Deviation of y
- ○ E) Variance of x
- ○ F) Variance of y

65. (2.00 pts)   For the above part, the average squared loss is equal to the ____

- ○ A) Average of x
- ○ B) Average of y
- ○ C) Standard Deviation of x
- ○ D) Standard Deviation of y
- ○ E) Variance of x
- ● F) Variance of y

A data scientist studying a population of newborn babies is trying to estimate birth weight (measured in ounces) based on the number of gestational days (the number of days the mother was pregnant). The equation of the regression line is:

estimated birth weight = 0.5 * (gestational days) - 10.5

66. (1.00 pts)   The unit of measurement of the y intercept is ounces.

● True   ○ False

67. (1.00 pts)   Longer pregnancies are associated with higher birth weights.

● True   ○ False

68. (1.00 pts)   If a mother's pregnancy is extended by one day, her baby is estimated to gain an additional 0.5 ounces in birth weight.

○ True   ● False

69. (1.00 pts)   If Mother A's pregnancy lasted 10 more days than that of Mother B, then Mother A's newborn is estimated to weigh 5 ounces more than Mother B's newborn.

● True   ○ False

Joint Events!

Cov = Covariance

Var = Variance

Let X and Y be two different events. For example, rolling a dice or drawing a card out of a deck without replacement. X and Y could be independent events or dependent events.

70. (1.00 pts)   If X and Y are two independent events, then Cov (X, Y) = 0

● True   ○ False

71. (1.00 pts)   If Cov (X, Y) = 0, then X and Y are two independent events.

○ True   ● False

72. (0.50 pts)   Cov(X, X) = Var(X)

● True   ○ False

73. (0.50 pts)   Cov(X, Y) = Cov(Y, X)

● True   ○ False

74. (0.50 pts)   Var(3X) = 3Var(X)

○ True   ● False

75. (0.50 pts)   If the covariance of X and Y increases, then the correlation also increases.

● True   ○ False

Distributions

76. (2.00 pts)
Peter Hung really wants the Superman toy from his cereal box. There are millions of boxes (assume infinite) of cereal and he knows Superman appears in the box with a 30% chance. What is the expected number of boxes Peter Hung will need to buy to ensure he gets Superman? <u>Answer in below in the most simplified fraction form.</u>

| 10/3 |

77. (1.00 pts)
Males arrive at a bus stop at a rate distributed with Poisson(X). Males get onto the bus with probability 0.2. Females arrive at a bus stop at a rate distributed with Poisson(Y). Females get onto the bus with a probability of 0.8. What is the distribution of people that get onto the bus?

Expected Answer: Poisson(0.2X+0.8Y)

78. (1.50 pts)   Given an unfair coin (ie. the coin does not have a probabilty of landing head of 1/2), how can we simulate a fair coin?

Expected Answer: HT = true heads, TH = true tails pr run a simulation of you filp 10 and I filp 10 times and whoever has more heads is one way

79. (1.50 pts)   Imagine I randomly select one person on the street. What is the expected number of people I need to sample before finding someone taller?

Hint: This is a tricky problem!!!

Expected Answer: +1.5 for infinite, +0.5 for 2 or referencing geometric distribution

Testing Stuff

80. (4.00 pts)

A scientist has developed a drug but needs your help! S/he selected a random sample of 100 sick patients and gave them the vaccine and found 39 were cured instantly! Construct AND interpret a 95 percent confidence interval for the proportion of all sick patients who would be cured with this vaccine.

Expected Answer: similar to question 2: https://secure-media.collegeboard.org/ap/pdf/ap17-sg-statistics.pdf

81. (1.00 pts)   For a student's t distribution, when the degrees of freedom approaches infinity, what distribution does the t distribution tend towards?

Expected Answer: normal / standard normal / gaussian

82. (1.00 pts)   If the p value is 0.05, what is the probability of a type 1 error if the sample size is 500? Answer in decimal form

0.05

You ask a random sample of 500 Experimental Design students how long they spent preparing for this event. The median response was 15.8 hours. You decide to create a 90% confidence interval by bootstrapping the 500 students in the sample. The interval you obtain is [12.3 hours, 19.2 hours].

83. (2.00 pts)   Explain the bootstrap process and how that 90% confidence interval was created.

Expected Answer:

84. (1.00 pts)   Describe one way to decrease the width of the confidence interval.

Expected Answer:

85. (1.00 pts)

Suppose next year there are 10000 Experimental Design students nationwide. Each one of those students repeats the bootstrapping process and obtains their own 90% confidence interval. Approximately how many of the students would you expect to NOT contain the population's median time spent studying for this event.

1000

Bonus for fun thinking! (not worth any points, not a tiebreaker)

86. (0.00 pts)

Imagine you have a stick of some length. You uniformly at random pick 2 spots on the stick to cut it into 3 pieces. What is the probability that these pieces form a triangle?

Expected Answer: