2022 Student Delegates Invitational

Problems by
Corbin Diaz, Jacob Buchsbaum, Nicolas Fernandez Baigun

Statistics Individual

January 22, 2022

1. Lil Timmy Tim is trying to answer this question for Ms. Lawton, but he's too busy on Arrakis to do it. He can't say no, because no means no. Help him out by finding the variance of the following data set:

$$\{3, 1, 4, 1, 5, 2, 7, 1, 8, 2, 69\}$$

A. $\frac{43676}{121}$ B. $\frac{2\sqrt{10919}}{11}$ C. $\frac{21838}{55}$ D. $\frac{\sqrt{1201090}}{55}$ E. NOTA

2. Nico, on a quest to make statistics more popular, is trying to find out how much of the Heritage student body knows about statistics. He wants to take a sample of the student body to ask them if they like statistics or not, but he is not sure what the best way to do so is. Help him out by finding which of the following sampling methods is *most likely* to reflect the true feelings of the population.

A. Asking the first 50 students he sees in the morning.
B. Posting a form and asking students to voluntarily answer.
C. Randomly sampling 50 students from the entire population and asking them.
D. Separating the student body based on their highest level of statistics taken (None, Honors, AP) and randomly sampling within each group proportional to size.
E. NOTA

3. Using the empirical rule, what percentage of all statistics test takers placed below Rosa if she scores a 95 on this test, assuming the scores of this test are normally distributed and $\mu = 85$ and $\sigma = 5$?

A. 68% B. 97.7% C. 95% D. 97.5% E. NOTA

4. Srijan is using his calculator on an AP Calculus practice exam, but his Nspire isn't working correctly and only gives the correct answer 40% of the time. Assuming Srijan knows how to do all 55 questions on the test and any errors are due to the calculator, what is the probability he gets at least half the questions correct?

A. 0.11 B. 0.07 C. 0.06 D. 0.04 E. NOTA

5. Anagh is shooting some b-ball on the court while he waits for the award ceremony to start, with a probability 0.1 of making any given shot. It takes him 5 minutes to take a shot, but the awards are in 61 minutes and he refuses to leave until he makes a shot! Find the probability to 2 decimal places that he makes it to the awards ceremony (assume Anagh can make it from the court to the awards ceremony in less than a minute because of his extremely long legs).

A. 0.72 B. 0.28 C. 0.10 D. 0.03 E. NOTA

6. Susvik and Susvaa Selvan are very sussy and are comparing their wins as the imposter in Among Us. Susvik has more wins than 69% of the population with 50. Susvaa has 110 wins and only 4.2% of the population has more wins than him. If the distribution of wins is Normally distributed and z-scores are rounded to 2 decimal places, what is the sum of the mean and standard deviation of the distribution to 3 decimal places?

A. 74.551 B. 74.390 C. 36.547 D. 36.398 E. NOTA

Use the following information for questions 7-8:
In the 1968 trial *People vs. Collins*, the couple Janet and Malcolm Collins were accused of robbery through probabilistic evidence. The prosecutor claimed that the chances of the couple matching the witness description of "Caucasian female with a blond ponytail and black male with a beard and mustache fleeing in a yellow car" were so small they had to be guilty. To do so, he assumed the following probabilities:

1. White woman with blonde hair: 1 out of 3
2. Woman with a ponytail: 1 out of 10
3. Black man with a beard: 1 out of 10
4. Man with mustache: 1 out of 4
5. Interracial couple in car: 1 out of 1000
6. Yellow car : 1 out of 10

The prosecutor then multiplied the probabilities together to get the probability all six conditions were satisfied, calculating a chance of 1 in 12 million.

7. Assuming the estimated probabilities of each of the individual six characteristics are correct, is the prosecutor's final calculation of 1 in 12 million correct?

A. Yes. The product rule states that given independent events A and B, $P(A \cap B) = P(A)P(B)$. Since the six characteristics are reasonably independent from one another, the prosecutor's calculation is correct.
B. Yes. The product rule states that given independent events A and B, $P(A \cap B) = P(A)P(B)$. While the six individual characteristics are not reasonably independent from one another, the probabilities given already take into account the chances of the characteristics influencing one another, so the prosecutor's calculation is correct.
C. No. The product rule states that given independent events A and B, $P(A \cup B) = P(A)P(B)$. Since the six characteristics are not reasonably independent from one another, the prosecutor's calculation is incorrect.
D. No. The product rule states that given independent events A and B, $P(A \cap B) = P(A)P(B)$. Since the six characteristics are not reasonably independent from one another, the prosecutor's calculation is incorrect.
E. NOTA

8. Regardless of your answer to the previous question, assume the prosecutors final calculation of 1 in 12 million is correct. Interpret how this result can be used as evidence in the courtroom.

A. Given a couple matches the description, there is a 1 in 12 million chance the couple is innocent. This result is unusable since you need the probability a couple matches the description given they are innocent.
B. Given a couple matches the description, there is a 1 in 12 million chance the couple is innocent. Therefore, the accused couple is highly likely to be guilty.
C. Given a couple is innocent, there is a 1 in 12 million chance they match the description. This result is unusable since you need the probability a couple is innocent given they match the description.
D. Given a couple is innocent, there is a 1 in 12 million chance they match the description. Therefore, the accused couple is highly likely to be guilty.
E. NOTA

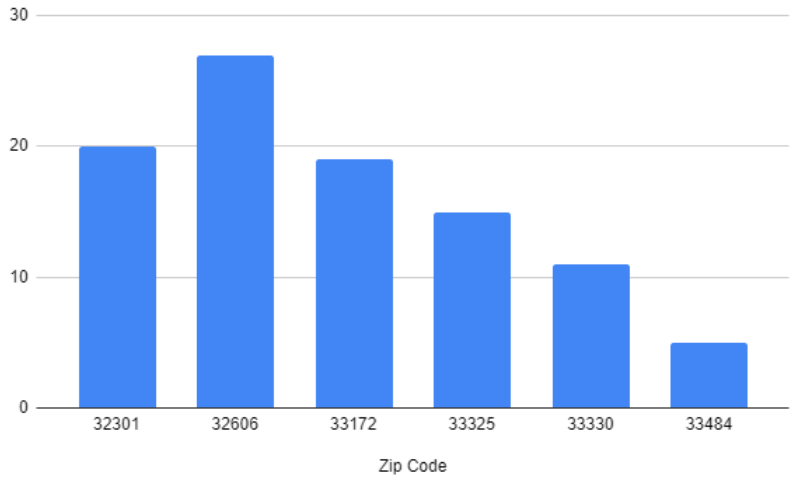
9. Carolina is going to the student center to meet up with Dr. Santos to apologize (she knows what she did). They will both arrive at the student center at some time between 1 and 2, but they do not know the exact time the other will arrive. Carolina will wait 10 minutes after arriving for Dr. Santos, but Dr. Santos has better things to do and will only wait 5 minutes. What is the probability the two will meet to 4 decimal places?

A. 0.2326 B. 0.4201 C. 0.3472 D. 0.7674 E. NOTA

10. Wild Monke disease is running wild through Gorillaland, and 20% of monkeys currently have it. One company, Orangutans Incorporated, has developed a test for the disease. As it was very rushed, the test has an 8% false positive rate (the probability a monkey tests positive given they do not have the disease) and a 14% false negative rate (the probability a monkey tests negative given they have the disease). Find the sum of the probabilities that a monkey has the disease given they tested positive and that a monkey has the disease given they tested negative. Round the final answer to four decimal places.

A. 0.7655 B. 1.2345 C. 1.6922 D. 0.3078 E. NOTA

11. Consider the following distribution. Based on shape, what can you say about the mean (μ) of the distribution and the median (M)?



A. $\mu > M$ B. $\mu < M$ C. $\mu = M$ D. Need more information E. NOTA

12. Let $f(x)$ be a valid probability density function over the domain $[0, 1]$. If the mean of the distribution is $\frac{3}{5}$ and the variance is $\frac{1}{10}$, find

$$\int_0^1 x^2(f(x))\,dx$$

A. $\dfrac{23}{50}$ B. $\dfrac{7}{10}$ C. $\dfrac{13}{50}$ D. $\dfrac{1}{2}$ E. NOTA

13. Let X be a randomly distributed variable and let $Var(X), Var(2X)$, and $Var(X+X)$ denote the variances of the distributions of $X, 2X$, and $X+X$ respectively. Given $Var(X) \neq 0$, find (rounded to four decimal places) the value of

$$\frac{Var(2X) - Var(X+X)}{Var(X)}$$

A. -0.5858 B. 0.5858 C. 2.0000 D. 0.0000 E. NOTA

Use the following information for questions 14-16:
In finance, the *expected rate of return* (E_R) is the average percentage change (or "return") on your investment. While one might invest in just one stock to get the most returns, it can lead to a large amount of *risk*, which can be measured by the *standard deviation of returns* (σ_R). Therefore, people typically invest in multiple stocks instead, which is called a stock portfolio.

In the Delegate Dow Jones, there are three main stocks: OdinCorp, Ritvik Retail, and Corbin Inc. The chart below, which is data for one day of trading, shows the probability of the stock market going up and down and each stock's expected percentage return for both ups and downs:

Stock Market		Expected Returns		
State	Probability	OdinCorp	Ritvik Retail	Corbin Inc.
Up	0.70	15%	5%	10%
Down	0.30	−20%	5%	1%

Clayton decides to casually invest \$100,000 that he has lying around into these three stocks. He builds his stock portfolio by putting 50% of his money into OdinCorp, 15% into Ritvik Retail, and 35% into Corbin Inc.

14. What is Clayton's expected rate of return E_R for one day of trading? (Give your answer as a percentage rounded to three decimal places)

A. 5.556% B. 5.555% C. 11.555% D. 11.556% E. NOTA

15. What is Clayton's standard deviation of returns σ_R for one day of trading? (Give your answer as a percentage rounded to three decimal places)

A. 0.298% B. 0.895% C. 4.732% D. 9.463% E. NOTA

16. Clayton is unhappy with how big of a risk his investment is, but doesn't want to decrease his expected returns too much. He creates what he calls the "Zipperian Ratio", which is given by E_R/σ_R, to better optimize his portfolio. Of the following, which investment option gives the maximum Zipperian Ratio? (Each ordered triple (x, y, z) means to invest x% in OdinCorp, y% in Ritvik Retial, and z% in Corbin Inc.)

A. $(60, 10, 30)$ B. $(20, 20, 60)$ C. $(20, 40, 40)$ D. $(50, 0, 50)$ E. NOTA

17. Define the *compactness* of a polynomial as the population standard deviation of its real roots. Find the compactness of the polynomial $P(x) = x^4 - 6x^3 + 8x^2 + 2x - 1$ (All of its roots are real).

A. $\sqrt{2.75}$ B. $\sqrt{4.75}$ C. $\sqrt{7.25}$ D. $\sqrt{9.25}$ E. NOTA

18. Continuing his quest to make statistics more popular, Nico has now collected a simple random sample of 50 students, 21 of whom say they like statistics. Nico wants to estimate the true proportion of Heritage students who would say they like statistics with 96% confidence. What is the interval that Nico constructs and can he conclude that a majority of students would say that they do not like statistics? Assume all conditions are met (Round endpoints to 2 decimal places).

A. (0.28, 0.56), Yes B. (0.28, 0.56), No C. (0.27,0.57), Yes D. (0.27,0.57), No E. NOTA

19. John believes that the true mean height of trophies from Cypress is greater than the true mean height of trophies from Buchholz, so he decides to run a hypothesis test. He takes a sample of 20 trophies from Cypress and finds a sample mean height of 1 foot with a standard deviation of 3 inches. He couldn't find as many Buchholz trophies so he only takes a sample of 10 trophies and finds a sample mean height of 9 inches with a standard deviation of 2 inch. Assuming all conditions are met, find (to 4 decimal places) and interpret the P-value for John's test.

A. The P-value 0.0006 represents the probability of observing a result greater than or equal to the one observed assumimg the null hypothesis is true.
B. The P-value 0.0006 represents the probability of observing a result greater than or equal to the one observed assumimg the alternate hypothesis is true.
C. The P-value 0.0016 represents the probability of observing a result greater than or equal to the one observed assumimg the null hypothesis is true.
D. The P-value 0.0016 represents the probability of observing a result greater than or equal to the one observed assumimg the alternate hypothesis is true.
E. NOTA

20. Steven loves pizza, so much so that he can't possibly imagine anyone eating more of it than him. To prove this, he runs a significance test to show that the true mean number of pizza slices people are willing to eat is less than his whopping 10 slices. The significance level is $\alpha = 0.05$ and he takes samples of size 50. If the true mean is actually 9 slices with a population standard deviation of 3 slices, what is the power of his test to 3 decimal places? Assume all conditions are met.

A. 0.238 B. 0.750 C. 0.754 D. 0.762 E. NOTA

21. P-hacking is the misuse of data analysis to find false positives in statistical tests through exhaustive searching. For example, a statistician engaging in p-hacking could continue taking samples from a population and running tests on them until they find a sample with a low enough P-value to reject the null hypothesis. Nico, continuing his quest to popularize statistics, has turned to the dark side and will run a right-tailed significance test on the proportion of Heritage students who would say they like statistics. Suppose the true proportion is 0.46 and that Nico knows this and will stop at nothing to reject this null value. If he takes sample sizes of 50, what is the minimum sample proportion he will need to reject this value at $\alpha = 0.01$, rounded to 3 decimal places? Assume all conditions are met.

A. 0.623 B. 0.624 C. 0.625 D. 0.626 E. NOTA

22. Just like there are one-sided and two-sided significance tests, there also one-sided confidence intervals. A lower one-sided t-interval states that the true population mean is less than or equal to a certain value $\bar{x} + t^* \frac{s}{\sqrt{n}}$. The critical value is calculated similarly to the two-sided version by bounding an area equal to the confidence level, but only on one side. If Alan takes a sample of size 30 that has mean $\bar{x} = 15$ and standard deviation $s = 1.5$, help him find the lower one-sided confidence interval with 95% confidence. Assume all conditions are met (Round the endpoint to 3 decimal places)

A. $(-\infty, 15.465)$ B. $(14.535, \infty)$ C. $(-\infty, 15.450)$ D. $(14.550, \infty)$ E. NOTA

23. Henry and Jonathan are playing an epic game of *Ultimate Frisbee*^{TM}. The game is divided into rounds, and a round ends once someone scores. In order to win, Henry must win two rounds in a row (assume Jonathan wins only if Henry gives up). Henry initially has a $\frac{1}{2}$ chance of winning a round. However, every time he wins a round, he gets overconfident and the probability becomes $\frac{1}{3}$. Every time he loses a round, he hypes himself up and his probability becomes $\frac{3}{5}$. Henry, fearful he may quit too early, wants to know many rounds he should reasonably stay in the game before giving up. What is the expected number of rounds until Henry wins?

A. $\frac{17}{2}$ B. $\frac{53}{6}$ C. $\frac{55}{6}$ D. $\frac{19}{2}$ E. NOTA

Use the following information for questions 24-25:
The *Federalist Papers* are a series of 85 essays written by Alexander Hamilton, James Madison and John Jay concerning the ratification of the U.S. Constitution. However, the papers were written anonymously, as free speech was not guaranteed at the time. While the authors of most of the papers were easily determined, 12 remained disputed for almost two centuries, mainly between Hamilton and Madison. In the 1960's, Frederick Mosteller and David Wallace analyzed the word counts of several essays in order to determine whether Hamilton or Madison wrote them. Consider the given word length counts for Federalist No. 84 and Federalist No. 55:

Word Length	1	2	3	4	5	6+
Federalist No. 84	85	880	759	538	403	1247
Federalist No. 55	52	368	317	243	203	555

24. Suppose Mosteller and Wallace run a χ^2-test to gather evidence on whether or not the two essays were written by the same author. What is the χ^2 contribution for the number of words of length 3 in Federalist No. 55?

A. $\frac{5759494}{9741435}$ B. $\frac{1561751361}{2641499300}$ C. $\frac{1561751361}{252985125}$ D. $\frac{5759494}{9537850}$ E. NOTA

25. Interpret the result of the appropriate χ^2-test proposed in the previous question using a significance level of 0.1 (Assume all conditions are met).

A. There is insufficient evidence to suggest that there is an association between the word length counts and essay number for Federalist No. 84 and 55.
B. There is convincing evidence to suggest Federalist No. 84 and 55 were written by the same author.
C. There is insufficient evidence to suggest Federalist No. 84 and 55 were written by the same author.
D. There is convincing evidence to suggest that the distribution of word length counts is the same for Federalist No. 84 and 55.
E. NOTA

26. The Fisher Test is a significance test that gives an exact P-value for two-way tables, as opposed to an approximation given by a χ^2-test. The Fisher Test is commonly used over the χ^2-test for small sample sizes and unequal distributions among cells in the table. For a 2x2 table, the P-value of the Fisher Test is equal to the product of the factorials of the marginal totals divided by the product of the factorials of the cell totals and table total. Using the Fisher Test, find the P-value for the following table to 3 decimal places (Assume all conditions are met).

	Uses Discord	Does Not Use Discord
Has a Significant Other	2	10
Does Not Have a Significant Other	15	5

A. 0.001 B. 0.002 C. 0.003 D. 0.004 E. NOTA

Use the following information for questions 27-28:
Spider-Man wants to see if there is a relationship between the number of antagonists in films he is in and the critical success of the film. Using the following data, he creates a least-squares regression line to try and predict the rating of a film based on the number of bad guys in it. Assume all conditions are met for both of the following questions.

Number of Villains	1	2	3	1	3	2	1	6	6
Rotten Tomatoes Score	90	93	63	72	51	92	90	93	97

27. Find the residual for *Spider-Man: No Way Home* (6 villains, 93% rotten tomatoes score) rounded to 4 decimal places.

A. 6.0035 B. -6.0035 C. 10.0035 D. -10.0035 E. NOTA

28. Spider-Man decides to run a linear regression t-test to see if there is a positive slope between the number of villains and the movie's score. What is the test statistic (rounded to three decimal places) and can he reject the null hypothesis at $\alpha = 0.05$?

A. 0.031; Yes B. 0.649; Yes C. 0.475; No D. 0.649; No E. NOTA

29. Jake is relating the number of board games one plays to the number of dates they go out on. Jake plots 20 data points and determines that the best equation to model this relation is in the form $y = ab^x$, where x is the number of board games played and y is the number of dates gone on. Let (x_i, y_i) denote the i^{th} data point, where $1 \leq i \leq 20$. Jake solves for this equation by building the appropriate **linear** least-squares regression line, and determines the value of b to be $\sqrt[3]{e}$. Given the following:

$$\sum_{i=1}^{20} x_i = 69, \qquad \prod_{i=1}^{20} x_i = 420, \qquad \sum_{i=1}^{20} y_i = e^{103}, \qquad \prod_{i=1}^{20} y_i = e^{123}$$

Find the value of a determined by Jake's least-squares regression line (Assume all conditions were met).

A. 5 B. e^5 C. 4 D. e^4 E. NOTA

30. Luke wants to find the beverage that helps you fall asleep the fastest, and runs an experiment to do so. To begin, he selects the beverages he wants to test: Water, Milk, Coffee, Tea, and Coke. He selects a random sample of five hundred people from everyone with a registered drivers license or learners permit in the state of Florida (meaning that his population is everyone with a registered drivers license or learners permit in the state of Florida). From there, he orders his sample in increasing age and assigns the treatment of Water to the first one hundred people, Milk to the second group of one hundred people, Coffee to the third group of one hundred people, Tea to the fourth group of one hundred people, and Coke to the last group of one hundred people. Each person is to drink sixteen ounces of their beverage at eleven PM, and immediately go to bed, and the time taken before falling into the sleep stage of rapid eye movement (REM) will be recorded by a sleep tracker. They are not allowed to use any electronic devices in this time period, and must all sleep on identical twin bed memory foam mattresses with only a single plume silk pillow, silk bed sheets, and a large comforter (to establish control). This is repeated on Mondays, Tuesdays, and Fridays for every week throughout the year (excluding holidays), for a total of one hundred forty-seven trial nights. Then, the mean time to fall asleep is calculated for each person in minutes, and recorded. Luke thinks he can establish a cause and effect relationship from this experiment, and his dog thinks he can generalize the results to the population. After conducting the experiment, Luke runs an Analysis of Variance (ANOVA), to find out which beverage results in the lowest mean time over all trial nights to fall into REM sleep. Analysis of Variance was developed by Ronald Fisher in 1918 and has been in use ever since. Put simply, ANOVA basically tells you if there are any statistical differences between the means of three or more independent groups. Like the t-test, ANOVA helps you find out whether the differences between groups of data are statistically significant. It works by analyzing the levels of variance within the groups through samples taken from each of them. If there is a lot of variance (spread of data away from the mean) within the data groups, then there is more chance that the mean of a sample selected from the data will be different due to chance. As well as looking at variance within the data groups, ANOVA takes into account sample size (the larger the sample, the less chance there will be of picking outliers for the sample by chance) and the differences between sample means (if the means of the samples are far apart, it's more likely that the means of the whole group will be too). All these elements are combined into a F value, which can then be analyzed to give a probability (p-value) of whether or not differences between your groups are statistically significant. There are also other types of ANOVA tests (which Luke uses to analyze the data), but you can read more about those in your spare time. Let the mean time before falling into REM sleep be denoted by $\bar{x}$ and the standard deviation be denoted by s_x. Here is Luke's data:

Beverage Type	$\bar{x}$	s_x
Water	90.0 min	2.0 min
Milk	89.0 min	1.5 min
Coffee	69.0 min	0.8 min
Tea	92.0 min	2.4 min
Coke	95.0 min	1.8 min

Surprisingly, Luke finds statistically significant evidence that Coffee results in the fastest time to fall into REM sleep. Who would've guessed!? Some possible reasons for this could include that coffee has relaxing effects or maybe that caffeine has an opposite effect on some people. Another possibility is that the act of drinking coffee before going to sleep might trick people into thinking that it'll take them longer to fall into REM sleep so they try and fall asleep faster to 'beat the system' (kind of how like some people try and stay awake after being put under general anesthesia - I can say this is true due to personal experience), or maybe that caffeine is a myth and doesn't actually work (who knows, maybe science is a lie? Maybe the Flat Earthers were right! NOOOOOOOOOO). However, Luke wouldn't be able to know the reason behind it, and would only know that there was statistically significant evidence towards Coffee being the beverage resulting in faster times to fall into REM. Between Luke and his dog, who is correct in this scenario?

A. Only Luke B. Only Luke's dog C. Both D. Neither E. NOTA