

1.  A. The upper quartile of the data set is 15 and the lower quartile is 5, the difference between those two is 10.
2.  D. The different age groups are the different strata in this case.
3.  E. If the value 69 was plugged in for  $x$ , the resulting mean of the data set would be 70.3
4.  D.  $\text{invNorm}(0.66, 83, 3, \text{LEFT}) - \text{invNorm}(0.66, 66, 3, \text{LEFT})$
5.  B. The liquid used is being changed and controlled by Sina, the experimenter.
6.  B. This question is asking for the probability that students preferred Haikyuu!! The number of students who preferred Haikyuu!! can be found by looking at the count in the second cell of the "Total" row, 79, and dividing that by the sum of the counts of all of the cells, 279. This gives a final answer of  $\frac{79}{279}$ .
7.  B. This question is asking what is the probability that a student's favorite anime is My Hero Academia, given that the student is either a freshman or a sophomore. There are 60 freshmen and sophomores who's favorite anime is My Hero Academia, and there is a total of  $60 + 38 + 46 = 144$  freshmen and sophomores, giving an answer of  $\frac{60}{144}$ .
8.  B. This question is asking for the probability that a student's favorite anime is Black Clover or that they are a junior or senior.  $46+52+42+41 = 181$  students surveyed were Juniors and Seniors or preferred Black Clover, giving a final answer of  $\frac{181}{279}$ .
9.  B. This question is asking for the probability that a student's favorite anime is My Hero Academia and that they are a junior or senior. 42 students surveyed were either Juniors and Seniors AND preferred My Hero Academia, giving a final answer of  $\frac{42}{279}$ .
10.  B. If the p-value is above 0.5, the distribution is skewed left.
11.  A.  $\text{normalCDF}(450, 540, 400, 80)$
12.  E.  $0.226^4 = 0.0026$
13.  D.  $1 - \text{normalcdf}(450, 540, 400, \frac{80}{\sqrt{6}})$
14.  C. The four conditions for conducting a significance test on the slope of a least squares regression line are: I) the observations must be independent, II) for any fixed value of  $x$ , the response  $y$  varies normally about the true line, III) the standard deviation of the response  $y$  about the true line is the same everywhere.
15.  D. The described scenario can be represented with a negative binomial distribution. If we let the  $k$  be the number of successes,  $r$  be the number of failures, and  $p$  be the probability of success in each trials, then the probability  $P(k, r)$  of having  $r$  failures before  $k$  successes can be determined with the probability density function:  $P(k, r) = \binom{k+r-1}{r} * p^k * (1 - p)^r$ . In this problem,  $k = 10$ ,  $r = 5$ , and  $p = 0.6$ , giving us that  $P(10, 5) = \binom{14}{5} * (0.6)^{10} * (0.4)^5 = 0.124$  when rounded to the nearest thousandths digit.
16.  C. Increasing the effect size is one of the ways to increase the power of a statistical test.
17.  A. The probability for scenario A is equal to  $0.25 * (0.35 + 0.10) * (0.45 + 0.10 - 0.05) = 0.056$ . The probability for scenario B is equal to  $0.35 * (0.25 - 0.05) * (0.45 - 0.05 + 0.10) = 0.035$ . The probability for scenario C is equal to  $0.45 * (0.20 - 0.05) * (1 - (0.35 - 0.10 + 0.10)) = 0.044$ . The probability for scenario D is equal to  $0.25 * (0.45 + 0.10) * (0.35 + 0.10 - 0.10) = 0.048$ . Out of these four scenarios, scenario A has the highest probability, so A is the correct answer.
18.  B. I is true. II is false, because the t-distribution tends to have a lower mean and higher extremes. III is false because a bernoulli distribution is only a binomial distribution of trial size  $n = 1$ .

19. **C**. The scenario described is a geometric distribution, with  $\mu = \frac{1}{p} = 25$ , and  $\sigma = \sqrt{\frac{1-p}{p^2}} = \frac{\sqrt{0.96}}{0.04}$ . The margin of error of the interval is equal to  $z^* \frac{\sigma}{\sqrt{n}}$ . Using the invNorm function, we find that the critical z value for this confidence interval is 2.17. Plugging in the values we know, setting the MoE expression less than or equal to 6, and solving for n, we get that  $n = \left(\frac{z^* \sigma}{6}\right)^2 = \frac{(2.17 * \frac{\sqrt{0.96}}{0.04})^2}{36} = 78.48166\dots$ , which gives us the smallest possible sample size of 79, after rounding up to the nearest whole number.
20. **E**. The degrees of freedom for a Goodness of Fit test is just the number of levels minus 1, so in this case,  $6 - 1 = 5$ .
21. **A**. The appropriate test is a chi-square test of independence/association. We enter the data cells of the table in a matrix on the calculator, for the observed counts, and then select  $\chi^2$ -Test under stat→TESTS, using an empty matrix for the expected counts. After pressing calculate, we get that the p-value is 0.3336480965 and that  $df = 2$  (which is just the number of rows minus one times the number of columns minus 1), the product of which is 0.67 rounded to the nearest hundredth digit.
22. **E**. The probabilities of each outcome don't add up to 1, so we cannot calculate the expected value.
23. **D**. This relationship can be modeled by a function of the form  $y = 2^{bx}$ . This is an exponential function, so the appropriate transformation to achieve linearity would be to take the log of both x and y, and the base does not matter as long as it is the same on both sides.
24. **C**. The definition of covariance is the measure of the variance between two random variables.
25. **A**. We know that Tanusri watches at least 20 minutes per each episode. There is a 0.25 chance she re-watches any given episode, which can add 20 more minutes. As the chance doesn't change, we can represent the number of minutes she spends on each episode with an infinite geometric sequence with first term 20 and common ratio 0.25. We can simplify it by factoring out the 20, so the first term is 1. The sum of this series is  $20\left(\frac{1}{1-0.25}\right) = \frac{80}{3}$ . Multiplying this by 201 for every episode gives us an expected total of 5360 minutes.
26. **A**. There is one factor or explanatory variable, the type of music listened to. There are three levels of this factor, which is rap, alt country, and classical music.
27. **D**. The probability that Vishal and Vishnav stop drawing letters after their 4th letter is equal to  $\frac{1}{2} * \frac{1}{2} * \frac{1}{2} * \frac{1}{2} * \frac{1}{2} = \frac{1}{32}$ . The probability that Vishal draws the letters "v," "i," "s," and "h" in that order is equal to  $\frac{1}{6} * \frac{1}{5} * \frac{1}{4} * \frac{1}{3} = \frac{1}{360}$ , and the probability that Vishnav does the same is equal to  $\frac{1}{7} * \frac{1}{6} * \frac{1}{5} * \frac{1}{4} = \frac{1}{840}$ . Therefore, the combined probability of both of them drawing the letters of "vish" is  $\frac{1}{32} * \frac{1}{360} * \frac{1}{32} * \frac{1}{840} = 3.23 * 10^{-9}$ .
28. **C**. The standard deviation of the difference of two correlated random variables A and B is equal to  $\sqrt{\sigma_A^2 + \sigma_B^2 - 2r\sigma_A\sigma_B} = \sqrt{25 + 36 - 2(0.1)(5)(6)} = 7.42$ .
29. **B**. One way to find the answer is just plugging in all the answer choices and seeing which one works. The other way is to use the binomial formula:  $P(n, k) = \binom{n}{k} p^k (1-p)^{n-k}$ , where  $n$  is the number of trials,  $k$  is the number of successes, and  $p$  is the probability of success in each trial. Plugging in the values we are given, we get that  $0.225 = \binom{6}{4} p^4 (1-p)^2 \rightarrow 0.225 = 15p^6 - 30p^5 + 15p^4 - 0.225 = 0$ . Graphing this polynomial in the calculator, we can find the zeros, of which there is only one that is  $\approx 0.5$ , which is 0.490.
30. **A**. The appropriate test is a two-sample T test, with  $H_0: \mu_D$  (pop. mean of the sample differences) = 0 and  $H_A: \mu_D \neq 0$ . Using the information given, we calculate the T statistic for this test:  $\frac{40 - 38}{\sqrt{\frac{6^2}{50} + \frac{8^2}{50}}} = 1.414$  Now we can use the tcdf calculator function to find the p-value of this test. By setting the lower bound as 1.414 and the upper bound as 1E99, and the  $df = 0.25$ , we get a p-value of 0.848474788. However, since this is a two-tailed test, we must multiply this by two for the left tail, giving us a final p-value of 0.170.