

Answers: BACDD DAABB CEECB CABAA AAACD BBDCD

1) Andrew wants one group for each drug at each dosage quantity for each drug; so, he will end up having 9 groups: No Drugs, 1 Pill Naproxen, 2 Pills Naproxen, 3 Pills Naproxen, 4 Pills Naproxen, 1 Pill Aspirin, 2 Pills Aspirin, 3 Pills Aspirin, and 4 Pills Aspirin. There are not 10 treatment groups because there is only one “no pill” group. Each group needs 15 participants, so he needs 135 people total, B.

2) The smallest value is 2, and the largest value is 98. Range = $98 - 2 = 96$, A.

3) Q1 is 3 and Q3 is 8, so the IQR is 5. $1.5 * IQR$ is 7.5, so the range of data values that are not outliers is $[-4.5, 15.5]$. The values 16, 22, and 98 lie outside this range, so there are 3 outliers, C.

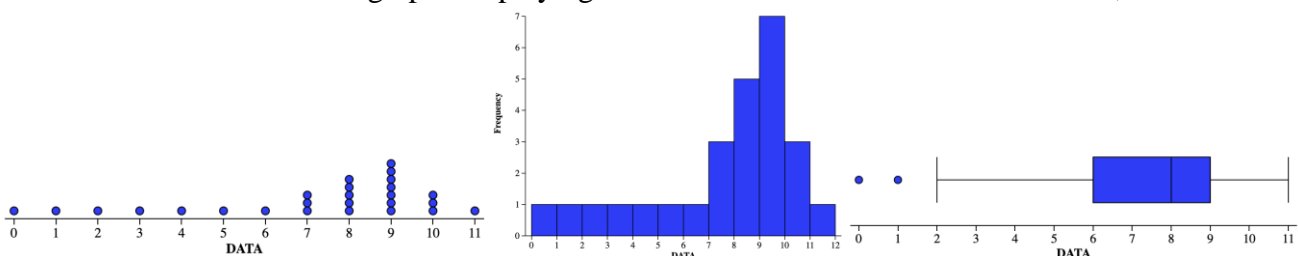
4) Using the first and second pieces of information, we can determine that the probability of a patient having Purpleitis is 10%. From this, we can find that the probability of a patient having Purpleitis and not Razz Apple Syndrome is 8%. Adding this to 81% gives us 89% for the probability of not having Razz Apple Syndrome. We can then find $\frac{8}{89} = 8.9888\%$ for $P(\text{Purpleitis} | \text{Not RazzApple})$. Rounding to the nearest percent yields 9%, D.

5) This can be solved via a system of equations. First, we calculate $invNorm(.80)$ and $invNorm(.90)$ to be 0.841621 and 1.28155, respectively. We can then make equations using the values and the Z-score formula. Then we can solve the linear system of equations between $0.841621 * \sigma = 6 - \mu$ and $1.28155 * \sigma = 7 - \mu$, which results in $\mu = 4.08692$ and $\sigma = 2.27309$, so $\mu + \sigma = 6.36001$, rounded to 6.36, D.

6) The first 20 Fibonacci numbers are: {1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, 1597, 2584, 4181, 6765}. Using a calculator, the population standard deviation is 1708.32, which is D.

7) First, we convert the provided range to Z-scores. $\frac{-1-1}{2} = -1$ and $\frac{5-1}{2} = 2$, so we want the interval between -1 and 2 standard deviations. The empirical rule states that 68% of data lies within one standard deviation, and 95% within 2. We want the sum of half of these values because we are within 2 deviations on one side of the mean, but only one on the other. $0.475 + 0.34 = 0.815$, A.

8) The distribution of the dataset is best described as left-skewed because there is a long tail to the left as can be seen in each of the graphs displaying the distribution of the data below. Thus, A.



9) According to page 79 of the current [AP Statistics Course and Exam Description](#), “When a sample is comprised entirely of volunteers or people who choose to participate, the sample will typically not be representative of the population (voluntary response bias).” This is not the case here since Slick Rick simply asks everyone he sees in his classes one day at school. What Slick Rick uses is clearly a convenience sample as defined on page 251 of the 6th edition of *The Practice of Statistics* since he “selects individuals from the population who are easy to reach.” This clearly results in undercoverage bias since, according to page 79 of the current [AP Statistics Course and Exam Description](#), “When part of the population has a reduced chance of being included in the sample, the sample will typically not be representative of the population (undercoverage bias).” Clearly, anyone who is not in one of Slick Rick’s classes has no chance (and hence, clearly a “reduced chance”) of being selected for his sample. There is no clear “non-response” since, presumably, everyone Slick Rick polls responds; and hence, there is no evident non-response bias. Next, no one with a any vested interest in the results of the study seems to have paid for the anything, so there is no funding bias. Finally, this is not sampling error (also known as sampling variability) which “refers to the fact that different random samples of the same size from the same population produce different estimates” according to page 298 of the 6th edition of *The Practice of Statistics*. All of this makes the answer B.

10) The situation described is modeled by a binomial distribution with $n = 14$ and $p = 0.08$. The formula for variance of a binomial distribution is npq . Substitution yields $14 * 0.08 * 0.92 = 1.0304$. Rounding results in 1.03, B.

11) The data John wants to display is that of a single categorical variable. The only suitable chart listed for this data type is a bar chart. It is not a box-and-whisker plot or histogram because they require quantitative data and ZIP codes represent locations on a map, and not quantitative. A scatterplot is not appropriate because they require two quantitative variables. Thus, C.

12) First, we can find the mean of the values of n , which is 10.5. Then, we can find the mean of a_n by multiplying 10.5 by 7 and then adding 6, to get 79.5. Doubling this, because doubling each element doubles the mean, yields 159, E.

13) Using the mean we computed for a_n from the solution for #12 (79.5), we can add 10 to this directly because adding to all elements is the same as adding to the mean, so 89.5, E.

14) This is a 3-circle Venn diagram scenario. First, we assign variables for the four unknown values, $w = \text{sleeping and jumping}$, $x = \text{sleeping only}$, $y = \text{jumping only}$, and $z = \text{bug catching only}$. Using the information given, we can form a 4-variable system of equations, with $x + y + z + w + 89 + 64 - 2 * 58 = 180$, $x + 89 + w - 58 = 120$, $y + 64 + w - 58 = 132$, and $z + 64 + 89 - 58 = 110$. Solving all of these yields $w = 87$, $x = 2$, $y = 39$, $z = 15$. We are looking for sleeping and jumping only, which is $w - 58$, so 29, C.

- 15) This is found by using the *I-Var Stats* program on the calculator, but make sure to use the sample standard deviation because Alex has collected a sample of students at Buchholz, not the entire population. This results in 0.833809, rounded to 0.834, B.
- 16) This is also found by using the *I-Var Stats* program on the calculator by entering {14,15,16,17,18} as the data list and {356, 478, 701, 456, 301} as the frequency list. This time, we want the population standard deviation because Kira has conducted a census. The result is $\sigma = 1.2453$, rounded to 1.245, C.
- 17) Kasra is trying to test if the actual percentage is higher than claimed; so, the test is one-sided. Additionally, he is only testing one percentage value; so, the test is for 1 proportion. Finally, the test is a Z-test. So, the answer is A.
- 18) Since the viewers are not selected by Jack and have to voluntarily call one of the numbers, this describes a Voluntary Response Sample, B.
- 19) This can be calculated using the *I-Var Stats* program on a calculator. Enter {1,2,3} as the data list, and {0.2, 0.55, 0.25} for the frequency list. This results in a mean of 2.05 and population standard deviation of 0.668954. Adding and rounding results in 2.72, A.
- 20) In order to calculate this, we can consider the vowels as a single “super-letter.” Thus, we need to permute six letters, ‘AIIOU’, ‘M’, ‘P’, ‘H’, ‘B’, and ‘S.’ This is simply 6! Then, we need to permute the 5 vowels in the “super-letter,” which is $5!/2!$ because the letter “I” is duplicated. The total number of permutations is $10!/2!$ Thus, the probability is $(6!*5!/2!)/(10!/2!)$, which rounds to 0.0238, A.
- 21) John collects data from every member of the population of interest for his survey, so he has taken a census, A.
- 22) This situation follows a geometric distribution, as Anna is continuing until she has a success. The formula for standard deviation of a geometric distribution is $\sigma = \frac{\sqrt{q}}{p}$. Substituting yields $\frac{\sqrt{1 - 0.015}}{0.015} = 66.1648$. Rounding yields 66, A.
- 23) The situation described follows a negative binomial distribution, which is the extension of the geometric distribution to multiple successes. The formula for the mean of this distribution is given by (Number of Successes Needed)/(Probability of Success). Substituting yields $2/0.06 = 33.33$, which rounds to 33, A.
- 24) In this experiment, Helena is changing the amount of gamma radiation to see if it has an effect on the lizard calculus ability, specifically how many questions can be solved. Thus, the IV is Radiation Level, and the DV is Number of Solved Problems, C.
- 25) In the experiment, the things Helena is performing the experiment on are the lizards, D.

26) Simpson's Paradox is when the comparison between parts of two variables is different from the comparison of their summaries. In the terms of this question, Simpson's Paradox is shown here because Hot Air is better when compared on averages, but Big Air is better when compared individually, B.

27) First, we calculate the test standard deviation via $s = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{.75(.25)}{40}}$. Then, we can find the test statistic: $z = \frac{p - p_0}{s} = \frac{.7 - .75}{.068465} = -0.730297$. We can then find the p-value by using $normalcdf(-\infty, -0.730297, 0, 1)$ to get 0.232604, B.

28) First, we can find the critical value using $invNorm$, which is $z^* = 1.95996$. Rounding to 3 decimal places, as instructed, yields $z^* = 1.960$. We can then find the confidence interval limits by computing $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$, which, when substituting values in, yields $8 \pm 1.960 \frac{16}{\sqrt{200}}$, which makes the interval (5.782513, 10.217487), rounded to six places, D.

29) A confidence interval describes a range of values that is likely to contain the population value with a certain degree of confidence. 95% is the level of confidence, not the probability. Thus, C.

30) Because the distributions are independent, the two means can be added to get the mean of $X + Y$, yielding 7. For the standard deviation, those cannot be combined directly, but variances can. Adding the two variances and square rooting yields $\sqrt{3^2 + 4^2} = 5$. $5 + 7 = 12$, D.