2022 National Convention Prob and Stats Hustle Answers

1. 25

2. $\frac{8\sqrt{3}}{3}$

3. 28

4. cluster

5. -2

6. 10.7

7. 453,600

8. 2.5

9. 4

10. 5

11. $\hat{y} = \frac{3}{20}x + \frac{63}{5}$ or $\hat{y} = 0.15x + 12.6$. The "hat" is not necessary on the "y."

12. $\frac{57}{64}$

13. 0.83

14. $26 + \sqrt{19}$

15. 0.9

16. 2

17. 50

18. $\frac{1}{3}$

19. $\frac{1}{8}$

20. 15

21. 0.21

22. 9

23. III, I, II

24. 42

25. stratified

2022 National Convention Prob and Stats Hustle Solutions

1. There are 420 and since 15 are taking neither, that means the union of taking AP Calculus and AP Statistics must equal 405. Therefore, AP Calculus + AP Statistics – Both = 405. $250 + 150 – \text{Both} = 405$, and solving this equation gives us 25 students taking both.

2. $s_x = \sqrt{\dfrac{(4-10)^2+(6-10)^2+(6-10)^2+(8-10)^2+3(10-10)^2+(12-10)^2+(14-10)^2+(20-10)^2}{10-1}} = \sqrt{\dfrac{192}{9}} = \dfrac{8\sqrt{3}}{3}$

3. $\dfrac{36-A}{B} = 2$ and $\dfrac{24-A}{B} = 0.5$.
   $36 - A = 2B$
   $24 - A = 0.5B$. Subtracting the bottom equation from the top equation we get $12 = 1.5B$; and thus, $B = 8$. Substituting this value into one of first two equations and solving, we get $A = 20$.
   $$A + B = 28$$

4. Dividing the population into classrooms, randomly choosing classrooms, and then selecting EVERY student in the chosen classrooms to be in the sample is an example of <u>cluster</u> sampling

5. The two statistics that are resistant measures of center and spread are the median and IQR respectively. The mean is affected by extreme values as are the standard deviation, range, and variance. The two statistics for correlation are also greatly affected by outliers and influential points. Therefore the sum is median + IQR = -8 + 6 = -2

6. The mean of a discrete random variable is $\sum x_i p_i = 2(0.1) + 6(0.2) + 10(0.35) + 14(0.2) + 20(0.15) = 0.2 + 1.2 + 3.5 + 2.8 + 3 = 10.7$.

7. The number of distinct arrangements of the letters in the word "regression" is $\dfrac{10!}{2!2!2!} = \dfrac{10*9*8*7*6*5*4*3*2}{8} = 453{,}600$. The computations you need to do are made easier than they appear by using $6! = 720$.

8. The test statistic $t = \dfrac{b}{SE_b}$. So $SE_b = \dfrac{b}{t} = \dfrac{8.75}{3.5} = 2.5$

9. Degrees of freedom = (# of rows – 1)(# of columns – 1) = 2*2 = 4

10. Expected cell count is $\dfrac{(row\ total)(column\ total)}{grand\ total} = \dfrac{20*15}{60} = 5$

11. The slope of the line is $b = r\dfrac{s_y}{s_x} = (0.1)\dfrac{1.2}{0.8} = \dfrac{3}{20}$. The intercept is found using $\bar{y} = a + b\bar{x}$. Substituting in the appropriate values and solving for a we get $a = 15 - \dfrac{3}{20}(16) = \dfrac{63}{5}$. Thus, the equation is $\hat{y} = \dfrac{3}{20}x + \dfrac{63}{5}$ or $\hat{y} = 0.15x + 12.6$.

12. The probability of obtaining at least two heads is equal to $1 - P(\text{zero heads}) - P(\text{one head}) =$ $1 - \left(\frac{1}{64} + \frac{6}{64}\right) = \frac{57}{64}$. There are $2^6 = 64$ possible outcomes when flipping a coin six times. There is one way to get 6 tails TTTTTT and six ways to get 5 tails HTTTTT, THTTTT, TTHTTT, TTTHTT, TTTTHT, TTTTTH.

13. Power $= 1 -$ Type II error. Type II error is not rejecting the null hypothesis when it should be rejected. Therefore, in this problem, power $= 1 - 0.17 = 0.83$.

14. The mean of $X + Y$ is equal to $12 + 14 = 26$. The standard deviation of $X + Y = \sqrt{10 + 9} = \sqrt{19}$. Their sum is $26 + \sqrt{19}$.

15. Students will hopefully notice that the slope of the LSRL in this problem is negative; and thus, the correlation coefficient is the negative square root of 0.81. The correct answer is 0.9 which is obtained without much difficulty.

16. Outliers can be found using the formulas: Q1 – 1.5IQR to identify low outliers and Q3 + 1.5IQR to identify high outliers. -7 – 1.5(-1-(-7)) = -16 is the low outlier cutoff and -1 + 1.5(-1-(-7)) = 8 is the high cutoff. Therefore, there are two outliers, 9 and 19.

17. The two missing values are 13 and 12. There are 13 different card values that can be used for the three of a kind. That leaves 12 different card values available for the pair needed to complete the full house. Therefore, $2(13 + 12) = 50$.

18. $P(\text{Blueberry} \mid \text{Low-Fat}) = \dfrac{(0.28)\left(\frac{1}{2}\right)}{(0.28)\left(\frac{1}{2}\right)+(0.60)\left(\frac{2}{5}\right)+(0.12)\left(\frac{1}{3}\right)} = \dfrac{1}{3}$.

19. The probability must equal one so substituting x = 0,1,2,3 into the function we end up with the equation 8c =1 and thus $c = \dfrac{1}{8}$.

20. Writing the numbers in ascending order we get {1, 6, 7, 9, 11, 14, 17, 26, 28}. The median of these numbers is 11. The median of {1, 6, 7, 9} is 6.5 = Q1. The median of {14, 17, 26, 28} is Q3 = 21.5. Thus, the IQR = 15.

21. For the interval (0.22, 0.64), the point estimate is 0.43 and the margin of error is therefore 0.64 - 0.43 = 0.21.

22. The average miles per shuttle ride is found using $Mean\ Miles = \dfrac{Mean\ Fare - 3}{2}$. So, the mean length of the rides, in miles, is 6. The standard deviation is not affected by the flat \$3 fee, so it is simply 6/2 = 3. Final answer is 6 + 3 = 9.

23. Event I: P(exactly 3) = $(5/6)^2(1/6) = 25/216$

Event II: P(no heads) = $(1/2)^3 = 1/8$ which is 25/200 so we know P(I) < P(II)

Event III: P(3 or 11) = (2/36) + (2/36) = 4/36 = 1/9 which is 25/225 so we know that P(III) < P(I)

Final answer in ascending order is III, I , II

24. When Mahadev consumes 8 beers, $\hat{y}$ = -1 + 4(8) = 31. He has a positive residual value of 11 and residuals are calculated using y - $\hat{y}$. So, the observed number of IQ points Mahadev lost after his consumption of 8 beers = 31 + 11 = 42.

25. Separating data in homogeneous groups and then taking a simple random sample from each one is stratified random sampling.